CIT. Journal of Computing and Information Technology, Vol. 30, No. 4, December 2022, 205–222
doi: 10.20532/cit.2022.1005688

Design of Electronic Voltage Transformer Error Pattern Recognition and Classification Algorithm Based on Data Mining

Deqian Kou and Yan Su

Inner Mongolia Ehv Power Supply Company, Hohhot, Inner Mongolia, China

With the development of smart grids, electronic voltage transformer (EVT) has gradually entered the stage of large-scale applications. Accurately identifying errors in electronic voltage transformers is crucial for the stability of power systems. Strengthening the measurement accuracy of EVT is of great significance for the operation of power systems and measurement and protection devices. However, due to the limitations of traditional verification methods, there are still challenges. To better improve the accuracy of transformer identification, a data-driven method for enhancing transformer error evaluation and prediction was developed. Based on the low accuracy of traditional EVT error verification and the difficulty of monitoring, data mining technology is proposed for EVT error analysis and evaluation. Recursive principal component analysis is used to separate errors from EVT measurement data, and feature statistics are used to monitor its operating status. Then, regression analysis under support vector machines is added to predict errors for active error correction and better evaluation of its status. The evaluation of the transformer monitoring dataset shows that the classification accuracy of error detection of the proposed method exceeds 93%, and the deviation between the predicted error value and the actual error value is less than 0.05%. Compared with methods such as artificial neural networks and ARMA, the average error rate has been reduced by more than 18%. The accuracy and average accuracy of the algorithm proposed in the study exceeded 80%, with values of 96.23% and 85.12%, respectively. The average error of the ratio difference feature of the EVT is only 0.023, and the average error of the angle difference is less than 0.01, which is much smaller than the algorithm used for comparison. The application response time is less than 0.1 s and the evaluation threshold can better identify data anomalies, with high application accuracy. This method can effectively provide real-time evaluation tools for the operational status of electronic voltage transformers and more accurately and proactively identify transformer errors from conventional data. This study provides an important data-driven solution for improving power grid reliability.

ACM CCS (2012) Classification: Computing methodologies → Machine learning → Machine learning algorithms

Keywords: data mining, electronic mutual sensor, PCA, SVM, *Q*-statistic, ratio difference characteristic

1. Introduction

The continuous development of sensor communication technology has gradually demonstrated the characteristics of intensification and informatization in power grid construction. Electronic voltage transformer (EVT), as a key measurement device in intelligent substations, can convert high voltage signals on the primary side into low voltage signals on the secondary side. At the same time, the EVT voltage measurement device can better adapt to the requirements of digital development in intelligent substations, with advantages such as simple insulation structure and wide dynamic range. The device is connected to high-voltage power grids, measurement and control devices, protection devices, metering devices, and other equipment, which can effectively monitor and measure the status of the power grid and electrical energy [1]. The accuracy and reliability of EVT measurement are closely related to the normal operation of measurement and control protection and metering devices and are important aspects that directly affect the stable operation of the power grid. Compared to electromagnetic transformers, EVT is prone to exhibit unstable error due to its complex structure and is prone to error exceeding anomalies in on-site operating environments [2]. Transformer error refers to the occurrence of data deviation when measuring or transmitting current and voltage due to factors such as electromagnetic heating, coil winding method, core magnetic flux density, average magnetic circuit length, *etc.* Strengthening the detection and correction of these errors can effectively improve the reliability and application performance of the power system [3]. However, the current measurement method for EVT error is through regular equipment testing or online verification, which is not conducive to the evaluation of EVT status and online monitoring due to short detection time and difficulty in batch testing of calibration equipment [4]. A number of scholars have proposed research ideas for the error diagnosis of power equipment. Among them, Rao *et al.* found through experimental evaluation that the integrated algorithm has good fault prediction performance in assisting in dissolved gas analysis transformer diagnosis [5]. Singh *et al.* proposed a modified lion algorithm for error prediction analysis of power transformers, and the results show that this method has significantly better root mean square errors than other algorithms [6]. Huang *et al.* proposed a combination of grey wolf optimization algorithm, differential evolution mechanism, and support vector machine to achieve transformer fault diagnosis. The results show that this method has high generalization ability, and its diagnostic accuracy is superior to genetic algorithms and others [7]. Stability and safety of the power system play important roles in its operation. Martin *et al.* proposed using neural markers to achieve keyword recognition for transformers and completed data labeling by designing model preprocessing [8]. Seyedshenava *et al.* analyzed voltage stability using finite element method and introduced evaluation metrics for data processing and comparison based on considering different short-circuit categories. The results show that the average error of this method is less than 5%, and it can effectively solve the nonlinear problem of voltage [9]. Liao *et al.* use graph convolutional neural networks for transformer fault diagnosis, and complete feature extraction training by designing adjacency matrices and backpropagation. This method exhibits good diagnostic accuracy and application performance [10]. In the research of electronic mutual sensors, traditional methods usually rely on physical sensors to collect and measure target physical quantities, which inevitably affects the sensor by environmental noise, interference, and faults, thereby limiting the reliability and stability of its network. At the same time, traditional measurement of EVT error requires nodes to be deployed in specific spatial positions, which limits the coverage and adaptability of sensor networks. Moreover, the use of equipment periodic testing or online verification analysis methods results in short detection time, increases communication and computing costs, and reduces the real-time performance of the system. The inspection equipment of different batches and manufacturers cannot guarantee the detection accuracy. Therefore, actively exploring the performance and reliability of electronic mutual sensor systems is an important research topic. Wireless communication, energy efficiency optimization algorithms, adaptive deployment strategies and others are used to improve the performance and reliability of electronic mutual sensor systems. Undermaintained EVT status can cause interference and impact on the normal operation of intelligent substations. Additionally, previous studies have shown that intelligent data mining algorithms can effectively analyze and predict the error situation of transformers. Data mining technology, as an important tool for data analysis, can achieve useful potential information extraction and automated classification on a large number of datasets. Utilizing data mining can lead to effective identification of hidden patterns and correlations of transformer errors, thereby improving the accuracy and efficiency of error identification.

There are many related works in the application of data mining in EVTs, mainly including data preprocessing, feature extraction, modeling, and classification. Feature extraction is based on the measurement data of EVTs and is used to extract meaningful features that can represent their performance and error patterns. Statistical features, frequency domain features, time-frequency analysis, and other methods can be used to extract the features. Additionally, various machine learning algorithms and data mining techniques can be used to construct models and classify data. Common methods include support vector machines, neural networks, decision trees, *etc.* The current application of data mining in EVTs still has certain shortcomings, such as the effectiveness of data mining being limited by the amount and quality of available data. If the measurement data of EVTs is limited or there are noise and outliers, it may affect the accuracy and stability of data mining. Moreover, the performance and error mode of EVTs are very complex, so feature selection and extraction may be a challenging task. Choosing appropriate features and extraction methods is crucial for accurate modeling and classification, but this may require the knowledge and experience of domain experts. The performance and error of EVTs may vary with time and environmental conditions. Therefore, the established model may need to be regularly updated and adjusted to adapt to new data and scenarios. Data mining can provide reference solutions for the optimization design of transformers and reduce the risk of power loss. Therefore, this study proposes to use data-driven methods to evaluate the measurement error status of EVTs and predict and analyze their error classification to better improve the error detection accuracy. At the same time, to avoid the impact of improper selection of data mining algorithms on the results, the research is based on transformer error extraction and time-domain feature analysis, using recursive principal component analysis for feature extraction. Subsequently, support vector machines under regression methods were introduced for error prediction, and dynamic time analysis was performed on the error characteristics to achieve better classification and prediction of results.

2. Design of Transformer Error Identification and Classification Based on Data Mining Methods

2.1. Error Separation of Electronic Voltage Transformer Based on Improved Principal Component Analysis

As an important part of connecting physical networks and control systems, EVT's data information status reflects the physical state of the power system and also reflects the correlation between measurement data and physical networks. The signal in an EVT flows through the primary sensor and converter, followed by the transmission system and secondary converter to achieve the digital information process. Factors such as temperature, humidity, and electromagnetic fields can cause EVT to exhibit complex error variations during the actual operation that can be mainly divided into two categories: systematic error and random error [11]. Random errors are mainly related to system noise, while system errors are related to the structural performance of mutual sensors. System errors are often caused by fixed and unchanging factors, including manufacturing errors, measurement errors, *etc.* Therefore, they can be further classified into two forms: abrupt and gradual errors. In EVT measurement, the deviation between the data measured on its side and the voltage cannot be avoided. But it only needs to meet the accuracy requirement of 0.2 level during the operation. The physical network model of an EVT that follows the normal distribution of variance can be expressed as:

$$u_s = k' B_s + v_s + s_x + N\left(0, \sigma^2\right). \tag{1}$$

In equation (1), u_s represents the EVT data measured at a specific time point, B_s represents the true value of the voltage signal, k' is the inductance coefficient, v_s is free noise, s_x is the error of the system itself, and σ represents variance. The generation of random errors may be caused by the influence of the magnetic field on the EVT sensing signal and conversion circuit. The system abnormal error of EVT is mostly reflected in the variation of transmission coefficient, and its mathematical model can be expressed as:

$$f_s = n_s B_s. \tag{2}$$

In equation (2), n_s represents the time function. According to the physical model on the EVT side, its measurement data is related to parameters such as voltage information, measurement error, system impulse situation, and power grid equivalent coefficient. The physical network correlation can be represented as in Figure 1.

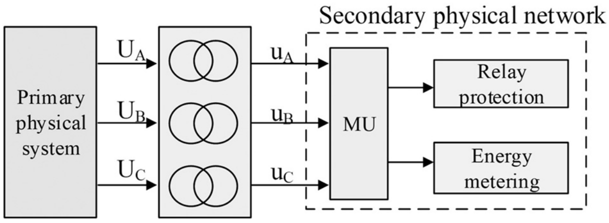


Figure 1. Schematic diagram of physical network correlation.

In Figure 1, the incomplete symmetry of the three-phase power system results in fluctuations in the actual value of its node voltage. The voltage fluctuation satisfies Gaussian independent distribution, and the variance result of three-phase voltage imbalance meets the accuracy requirements of the transformer. It can be explained with the help of equation (3).

$$M = \frac{\max\left[\left|U_A - U_{avg}\right|, \left|U_B - U_{avg}\right|, \left|U_C - U_{avg}\right|\right]}{U_{avg}} \tag{3}$$

In equation (3), U_A, U_B, U_C represent the effective values of the three-phase voltage. U_{avg} is the average of the effective values, and M is the voltage imbalance. When analyzing measurement data, distinguishing between the measurement error and measurement deviation of the transformer can reduce the physical fluctuations of the system to a lesser extent. In order to better analyze the data of EVT error, a study proposes an error state evaluation based on principal component analysis (PCA). PCA can represent two variables with discreteness in a linear combination, thereby achieving their coordinate transformation, that is, feature dimensionality reduction while preserving the original information of the data [12]. At the same time, considering that the original PCA model is difficult to detect the varying signal amplitude and phase position when analyzing nodes in power transmission and transformation systems, it is not suitable for three-phase time-varying linear systems [13]. Therefore, this study proposes Recursive Principal Component Analysis (RPCA) for analysis, which completes state assessment by continuously updating data. If the original data matrix is represented as X_1^0, the formula for standardizing the matrix is given in equation (4).

$$X_1 = \left(X_1^0 - I_{n1} b_1^T\right)\sum\nolimits_1^{-1} \tag{4}$$

In equation (4), n represents the normal operating state, I represents the identity matrix, b_1 represents the mean of the variable, and T is the matrix transpose symbol. In the RPCA process, the updating of sample points mainly relies on the rank correction method to update the scores of principal components and load vectors. The covariance of sample point updates can be expressed as:

$$R_{k+1} = \frac{k-1}{k} R_k + \sum_{k+1}^{-1} \Delta b_{k+1} \Delta b_{k+1}^T \sum\nolimits_{k+1}^{-1} + \frac{1}{k} x_{k+1}^T x_{k+1}. \tag{5}$$

In equation (5), k represents the number of data blocks, x_{k+1} is the standardized matrix of $k + 1$ data blocks, b_k represents the mean of the data block, and R_k is the covariance matrix of the data block. Since the matrix in the equation is a matrix with rank 1, equation (5) needs to be modified twice to obtain the eigenvalues of the covariance matrix after the two modifications:

$$R_{k+1}' = P_{k+1} D_{k2} P^T_{k+1}. \tag{6}$$

In equation (6), D_{k2} represents the diagonal matrix after rank correction, and P_{k+1} is the updated feature vector. On the transformer model, it is necessary to establish statistics for hypothesis testing in the subspace to better detect abnormal situations. The study constructs Hotelling T2 statistics and Q statistics for the principal component subspace and residual subspace, respectively [14–15]. The T2 statistic represents the standard sum of squares of the score vector, which conforms to the degree of freedom distribution. Its statistical control limit can be expressed as:

$$T_{UCL}^2 = \frac{k(n+1)}{n-k} \cdot F_\alpha\left(k, n-k\right). \tag{7}$$

In equation (7), k represents the degree of freedom, α is the confidence level, and n is the number of training data. F represents the percentile value of degrees of freedom in the central distribution. The mathematical expression of the control limit of the Q-statistic is:

$$Q_{UCL} = \frac{v_g}{2m_g} X^2_{2m_g^2/v_g,\alpha}. \tag{8}$$

In equation (8), v_g, m_g represents the mean and the variance of Q values, respectively. The evaluation limit of Q-statistic can be represented using Figure 2.

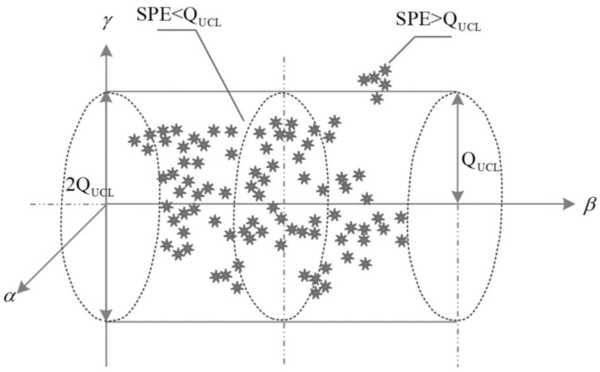


Figure 2. Evaluation limits for the Q-statistic.

In Figure 2, each solid point represents a data point and each point is related to β. The distance from the axis represents the square prediction error of the subspace. Data that is not within the range is abnormal data. The error in EVTs mainly consists of online and offline aspects, and the evaluation process is shown in Figure 3.

Offline learning mainly involves constructing an initial model for training samples, including calculating the mean vector and covariance matrix of the data, followed by recursive updates of the principal component model based on the relationship between the Q-statistic and the control thresholds. The update of the model and covariance matrix during the online process mainly involves replacing data at different sampling times [14]. The update of the principal component model is used to compare the data control limit and the squared prediction error (SPE) of the Q-statistic. If the statistic is greater than the limit, a fault prompt will be given, otherwise, the update will continue.

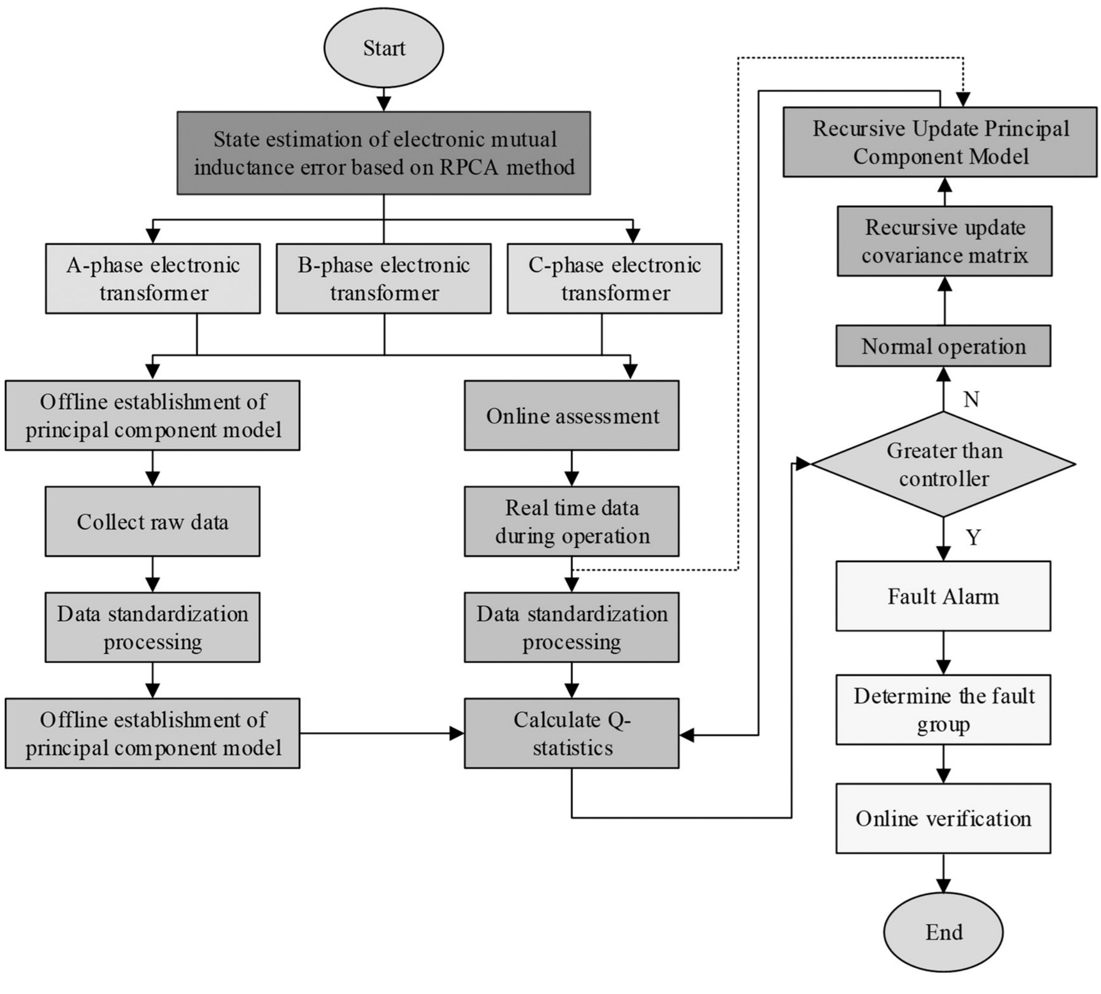


Figure 3. Flow chart for error measurement evaluation of electronic transformer under ERCA.

2.2. Predictive Analysis of Support Vector Regression in EVT Error State

By predicting and analyzing the error status of EVTs, possible errors or anomalies in the transformers can be identified and corrected in a timely manner. Furthermore, the analysis of error status can ensure EVT's reliability and accuracy under various working conditions, reducing potential measurement deviations and further errors [16]. The analysis of EVT error state prediction problems is common, and traditional prediction methods such as regression analysis and trend extrapolation are difficult to evaluate and to adjust the parameters of models. There is also a certain error in prediction accuracy [17]. With the development of network data, artificial intelligence algorithms have gradually become a hot research topic and have been involved and applied in fields such as electronic technology, power systems, and medicine. Therefore, this study selects intelligent prediction algorithms for EVT error state estimation. As a learning algorithm, the support vector machine has good generalization ability and fitting performance and can achieve the minimum effect of the system under a limited number of samples [18–19]. As a data mining algorithm, SVM can achieve nonlinear mapping of input vectors in high-dimensional space and construct the optimal classification plane. SVM completes the prediction generalization problem by utilizing a set of nonlinear functions during sample data processing. The mapping principle is to map the low dimensional sample space to the high dimensional sample space, and finally perform the optimal solution to achieve classification. The distance of the sample data on the hyperplane can be expressed as:

$$d = \frac{|\omega \cdot x + b|}{\|\omega\|}. \tag{9}$$

In equation (9), ω represents the generalized parameter, b represents deviation, and x is the sample point. In support vector machines, the mapping relationship between high-dimensional and low-dimensional data is achieved through the classical form of Lagrangian functions. Its kernel function can replace the inner product transformation in the case of linear separability, achieving the classical Lagrange function change under optimal classification, as shown in equation (10).

$$f(x) = \operatorname{sgn}\left\{\sum_{i=1}^{n} a_i^* y_i \left(x_{i \cdot x}\right) + b^*\right\} \tag{10}$$

In equation (10), (x_i, y_i) is the inner product transformation, a_i is the Lagrange multiplier, where x_i is the column vector of the sample output and y_i is the corresponding output value, and a_i^* represents a support vector. The structural risk minimization faced in SVM theory learning is to find suitable functions in the sample dataset to ensure that the algorithm has good fitting performance. Figure 4 is a schematic diagram of minimizing the structural risk.

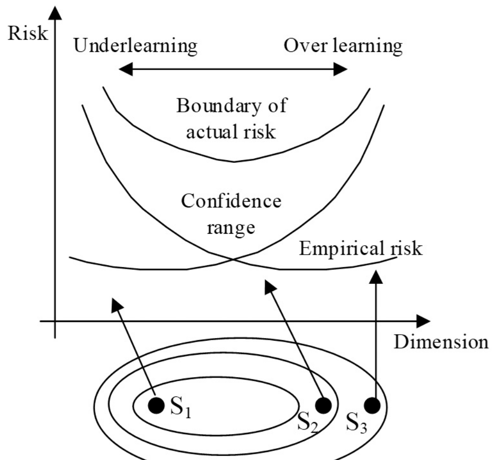


Figure 4. Schematic diagram of the minimal structural risk.

The structural risk function can be used to replace the distribution function, and then obtained by solving the minimum value of the function:

$$\min_{Sh}\left\{\operatorname{Re}(f) + \sqrt{\frac{h(\ln 2n/h + 1) - \ln \delta/4}{n}}\right\} \tag{11}$$

In equation (11), Sh represents the dimensional spatial structure of the hyperplane composed of data samples, h represents the dimension, δ is the reliability parameter, and $\operatorname{Re}(f)$ represents the structural risk function. This study incorporates regression analysis ideas based on SVM. The regression under SVM needs to ensure that the error between the actual and predicted values obtained from the sample data during training is less than or equal to the initial error value, and to minimize the expected risk function solution in this case. If the number of samples in the training is set to L, the linear regression function in the high-dimensional feature space can be expressed as:

$$f(x)' = w\,\phi(x) + b. \tag{12}$$

In equation (12), $\phi(x)$ represents a nonlinear mapping function. The insensitive loss function is:

$$L\big(f(x),\, y,\, \varepsilon\big) = \begin{cases} 0, & |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon, & |y - f(x)| > \varepsilon \end{cases}. \tag{13}$$

In equation (13), $f(x)$ represents the predicted value of the regression function, y is the true value, ε is the loss function parameter, and the value of ε is proportional to the error of the regression function. At the same time, to better evaluate and predict the EVT error results, the study proposes to use sliding time windows to collect feature parameter data at different times. By setting the time window t_i and step length t_s, the output voltage data of EVT can be obtained:

$$Z(t_i) = [z(t_i - t_s + 1), z(t_i - t_s + 2), ..., z(t_i)]. \tag{14}$$

By solving the characteristic parameter sequence related to in-phase voltage and output voltage at a certain moment, a stationary sequence can be obtained under different differential intervals. When conducting error evaluation, it is necessary to pay attention to the selection of step length. Neither too long nor too short can effectively describe the adaptive change process of the error state, and there is a possibility of misjudgment [20]. Therefore, the study utilizes the evaluation results of EVT to achieve real-time update of step length, and its mathematical expression is given as:

$$t_s\left(t_{i+1}\right) = t_s^{\max} + \left(t_s^{\max} - t_s^{\min}\right) * \min\left[\frac{\min|\Delta D|}{D(t_i) - D_{\lim it}(t_i)},\, 1\right]. \tag{15}$$

In equation (15), $\min|\Delta D|$ represents the minimum change in the reference evaluation index, $t_s^{\max}$ is the maximum step length for error state evaluation, and $t_s^{\min}$ is the minimum step length. $D(t_i)$, $D_{\lim it}(t_i)$ represent the evaluation indicators and their thresholds at different times. Equation (15) can effectively achieve real-time tracking and evaluation of EVT errors. The evaluation of the error state of EVT is mainly based on the evaluation threshold obtained from its reference data. The data will change with the evaluation results and the real-time error status with EVT. It should also be noted that when the EVT error shows an increase in positive polarity, the corresponding evaluation index results should also increase.

Based on EVT error state analysis, this study designed an overall system framework to achieve the collection and application analysis of data information and collected data from the front-end through the data layer, application layer, and display layer. The data collection section mainly includes the collection subunit, input/output subunit, data cache subunit, *etc.*, and is responsible for parsing messages, unit data processing, and data caching. The data layer includes two aspects: static application data of the device and real-time data of the device, in which the three-phase voltage of the EVT is dynamically captured [21]. The application layer and display layer are used to achieve visual monitoring, analysis, and result display of data, thereby enabling the evaluation of error systems. Figure 5 shows the architecture of the EVT error online evaluation system.

In Figure 5, the measurement data processing structure includes service interfaces, management configurations, data services, and other contents. The data integration module mainly extracts offline and real-time data, and then uploads the real-time data through the substation port. The data cleaning, storage center, and computing center are used for data identification, storage, and analysis of EVTs.

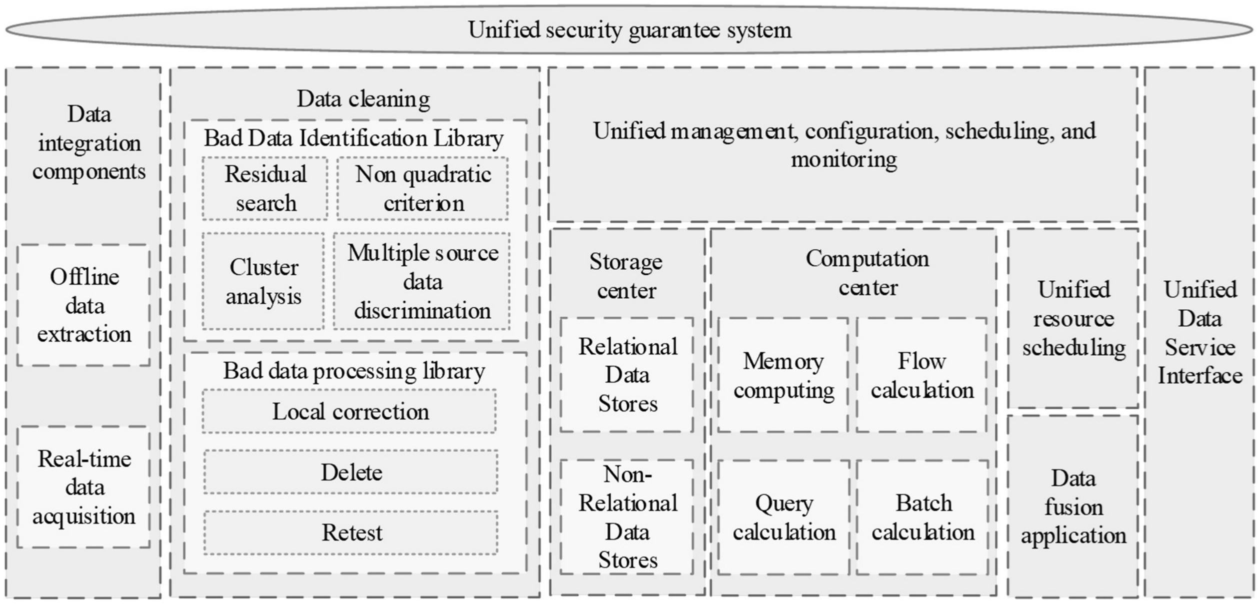


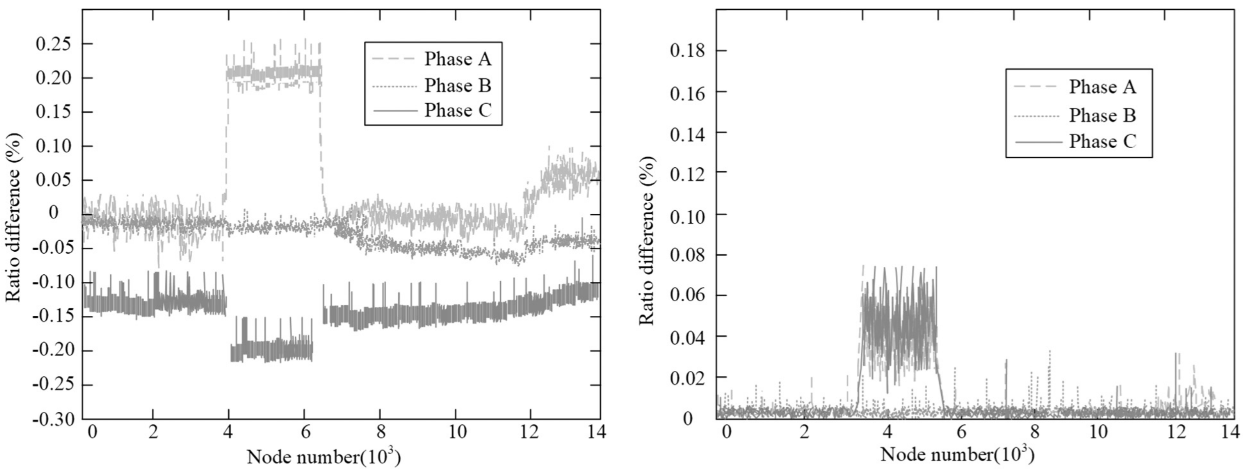
Figure 5. EVT error online evaluation system architecture.

Recursive principal component analysis can adaptively decompose input signals into multiple intrinsic mode functions based on their characteristics, thereby obtaining better signal representation and adapting to different types of error analysis. It can also extract the characteristics of signals at different time and frequency scales, thereby capturing the frequency variation characteristics of different types of errors in EVT measurement data, which helps to analyze and classify errors more finely. Support vector machine regression can model nonlinear relationships by using kernel functions to map data from the original space to high-dimensional feature spaces, taking into account the real-time dynamism of EVT measurement data. This can provide explanations for classification and recognition of different error types based on the relationship between input features and target errors. By integrating the two methods, a multi-stage error analysis and classification system can be constructed, improving the accuracy and stability of error analysis and classification of EVT measurement data, and enabling better understanding of the error feature results caused by data information differences.

3. Analysis of Transformer Error Application Results

Electronic voltage transformers are actually systems that sense, transform, transmit, and process signals in high-voltage systems. During the simulation process, they will inevitably be affected by signal interference and their performance will be affected. This study analyzes the error of the transformer based on the proposed recognition and prediction model algorithm, in order to provide better reference for the performance improvement of the equipment. Taking three-phase EVT as an example, the study simulates the sensing part of the EVT signal using a voltage divider resistor and sets the proportional parameter relationship between the resistors to 178K/6.81K. The secondary output of the transformer collected from the substation is reproduced using a three-phase programmable power source, with a proportional relationship of 110kV/100V. Simultaneously, a data acquisition system is used to record the sensor signal output and simulate the secondary conversion unit. The IEC61850-9-2 protocol completes online verification of data sent to the error status evaluation platform and the calibration system. The sampling frequency is set to 4 kHz, the number of samples per second is 4000, and the sampling period is 1 second. The amplitude, phase, and other data of EVTs is collected and compared in a real-time environment. Calibration analysis is performed with a standard transformer with an accuracy level of 0.05 to better validate the effectiveness of the research method. The voltage signal measurement data of three-phase EVT is around 64,000V~69,000V. The error situation of transformers is analyzed using different methods, and the results are shown in Figure 6.

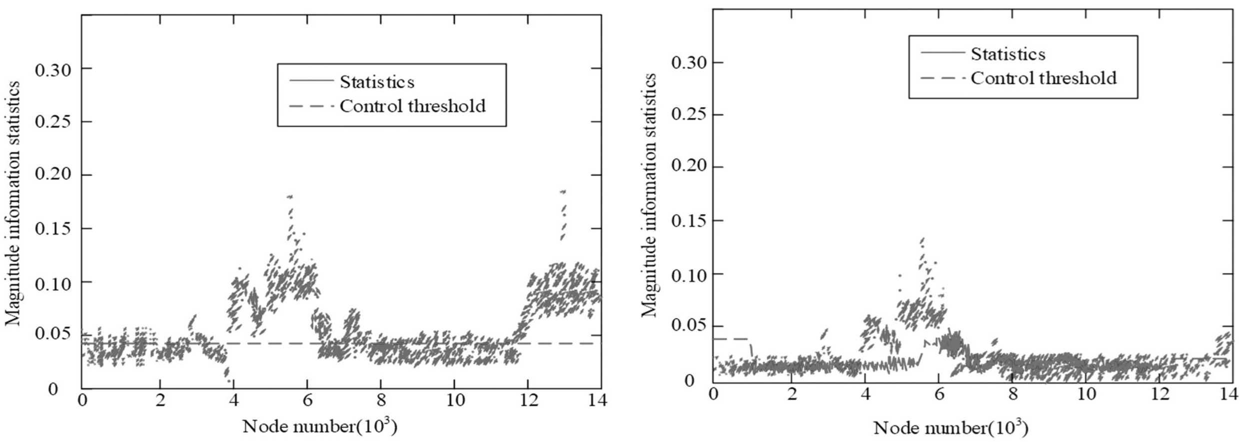
Figure 6 shows that the error results of the data sampling value under the standard transformer vary significantly. Specifically, phase A and phase C exhibit data fluctuations when the number of sampling points ranges from 4,000 to 6,100, resulting in a significant "fault" of decline. The measurement error variation of both reached 0.25% and -0.07%. From the overall trend, the measurement errors of phases A, B, and C reached 0.09%, 0.06%, and 0.03% when the sample size was greater than 11,000. This indicates that it is difficult to effectively identify and detect abnormal data. The error results of the three-phase transformer under the improved method are relatively small, and the overall curve change is relatively stable. The error curves of phase A and phase C have consistency in the sample size range of 4,000~6,200, indicating that they can effectively identify abnormal error situations. Subsequently, the square prediction error results at the sampling points were analyzed, and the results are shown in Figure 7.



(a) EVT measurement error

(b) Contribution of the SPE statistic for EVT

Figure 6. Error situation of a standard transformer.



(a) SPE statistical monitoring results under traditional principal element analysis

(b) Statistical monitoring results of SPE under improved principal element analysis

Figure 7. Square prediction error results of the sampling points.

In Figure 7, the RPCA method shows good applicability in detecting abnormal data in electronic transformers. The changes in its statistical threshold will exhibit fluctuations with the changes in statistics. However, the error situation of the transformer analyzed with the PCA method does not have good adaptability, and the statistical fluctuation value range of the sample points is from 0 to 0.02. To better test the effectiveness of the proposed prediction algorithm, the EVT error was analyzed with Artificial Neural Networks (ANN) [22], Auto-Regressive Moving Average Model (ARMA) [23], and Radial Basis Function Neural Network (RBFNN) [24]. The output data of an electromagnetic transformer was collected at a sampling frequency of 30 minutes, and the original data were predicted and analyzed. Comparison of the ratio error results under different algorithms is shown in Figure 8.

Figure 8 shows that, overall, the error ranking algorithm between the predicted results and the actual results is: The research algorithm > RBFNN > ANN > ARMA. Specifically, the average deviation between the proposed prediction algorithm and the true value during sample prediction is less than 0.05%, and the curve trend is basically consistent. The RBFNN, ANN, and ARMA models exhibit average equal error deviations of 0.12%, 0.25%, and 0.73%, respectively, and are more susceptible to the influence of the number of sampling points. Further testing was conducted on the methods' predictive effect with fitting accuracy, as shown in Figure 9.

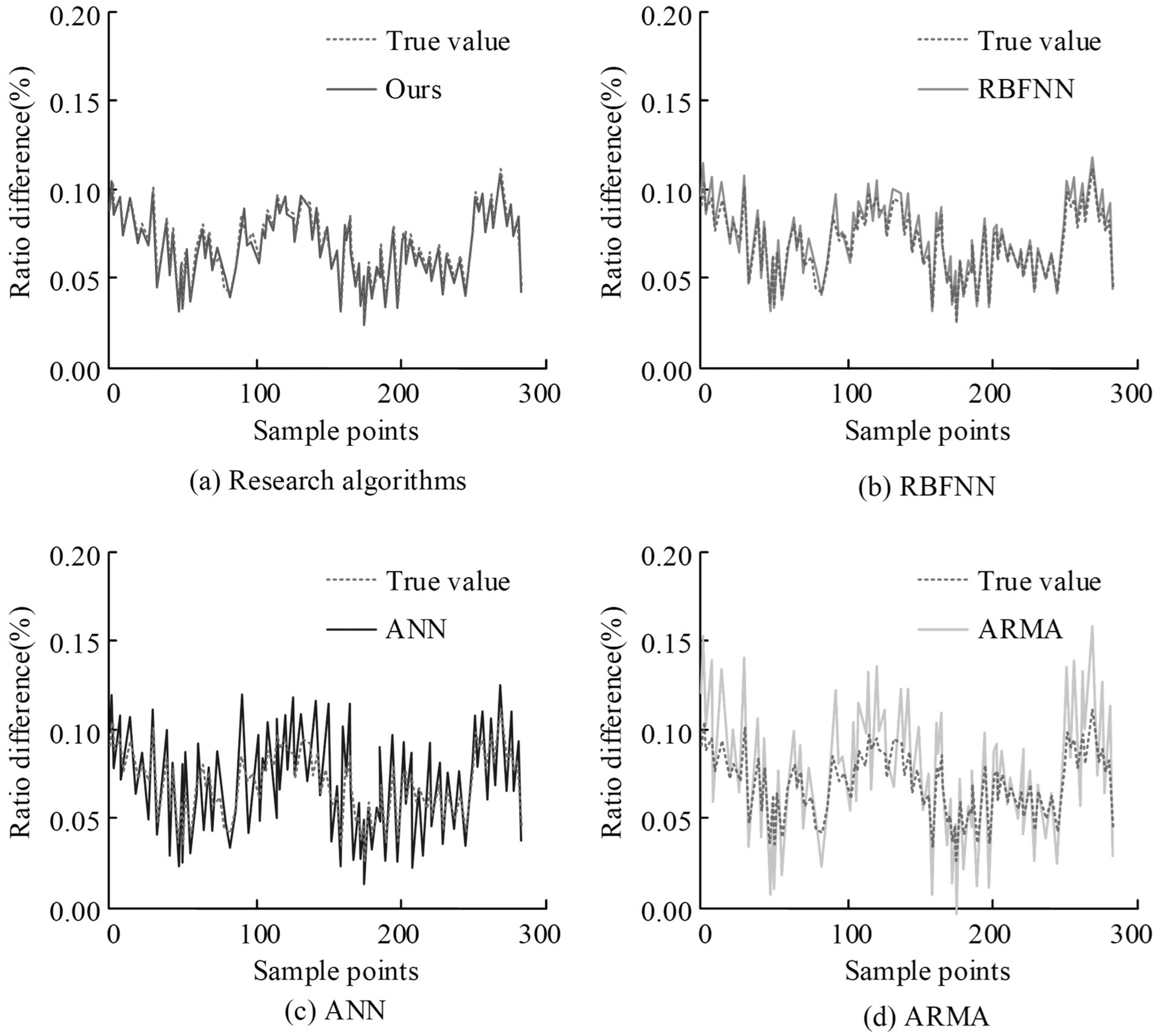


Figure 8. Error prediction results of different algorithms.

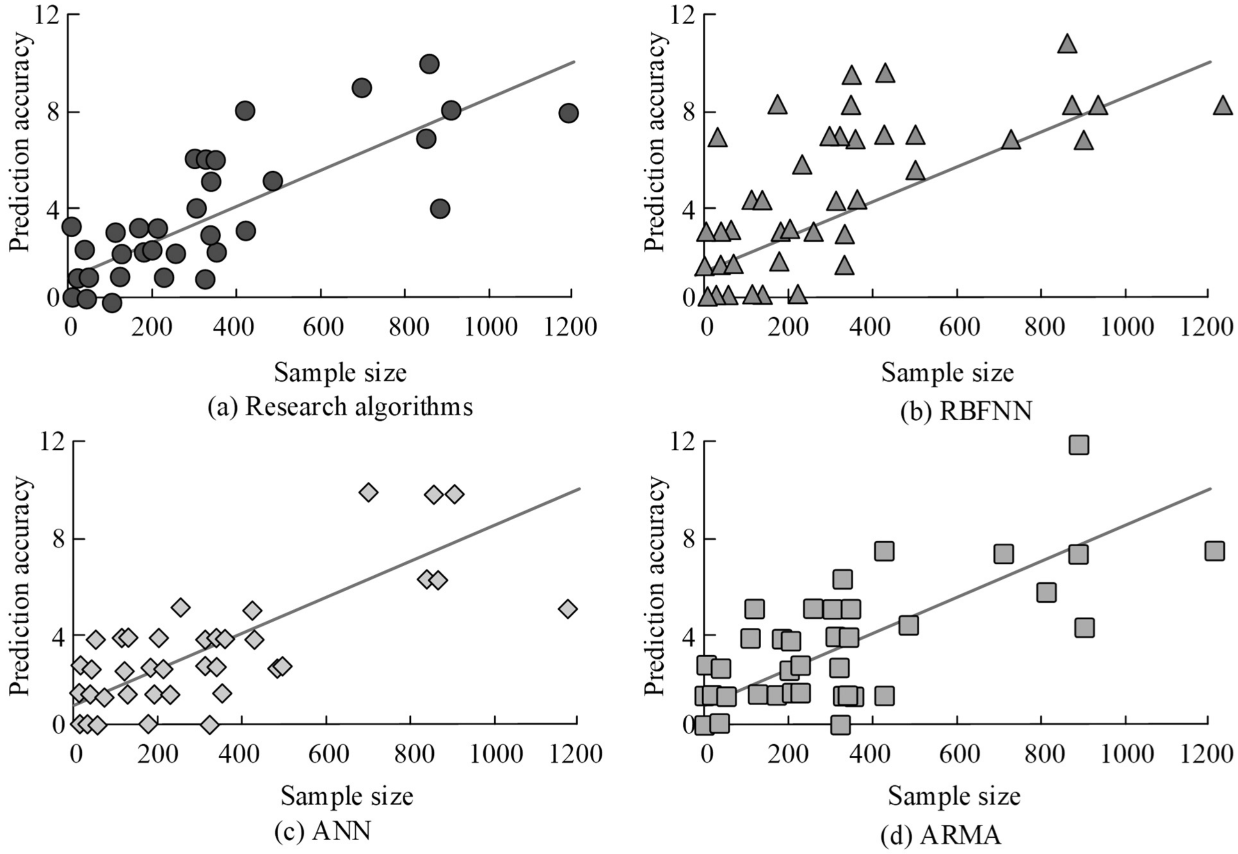


Figure 9. Prediction accuracy fitting results of the algorithms.

The fitting values of the research algorithm, RBFNN, ANN, and ARMA are 0.97, 0.89, 0.82, and 0.77, respectively. From this, it can be seen that the proposed error prediction algorithm has shown good accuracy in evaluating EVT. Subsequently, feature evaluation and test result analysis were conducted with the proposed hybrid combination error algorithm, and the results are shown in Figure 10.

In Figure 10, the hybrid (mixed) model (RPCA+SVM regression) can better predict the error state of the transformer. The average error of the ratio difference characteristic results was significantly less than that of the single model (SVM regression) ($0.023 < 0.046$). Additionally, the difference of the phase information between the mixed model and the single model is also different. When the number of sampling points exceeds 5,000, the fluctuation of the single model is obvious, and the maximum characteristic difference reaches 0.073 [25]. The mixed model exhibits a feature average error of less than 0.01. Li *et al.* proposed a method based on genetic algorithm and support vector machine (GA-SVM) to improve the classification of power transformer faults from two aspects: type and location [26]. They first extract unique features using the mathematical index of the induced current at the head and end of the transformer winding, and then feed this feature into a support vector machine for training. They use genetic algorithms to optimize the parameters of the SVM model. The results show that the model can effectively identify different fault types and determine their positions in the transformer winding, with diagnostic rates of 100% and 90% for fault types and fault locations, respectively.

Wu *et al.* proposed a combination of a genetic algorithm and an XGBoost for power transformer fault identification, which differs from traditional transformer fault diagnosis methods based on dissolved gas analysis and exhibits high fault identification accuracy [27]. The research approach of the above results is mainly based on the classification of fault problems and the optimization of parameter results. The results are similar to the proposed method of combining recursive principal component analysis with support vector machine.

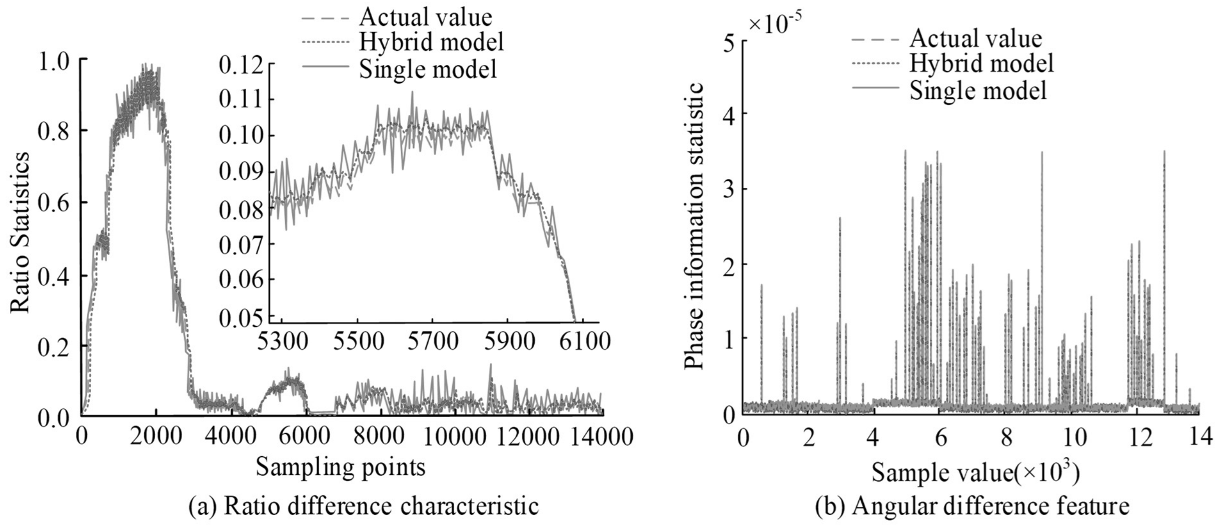
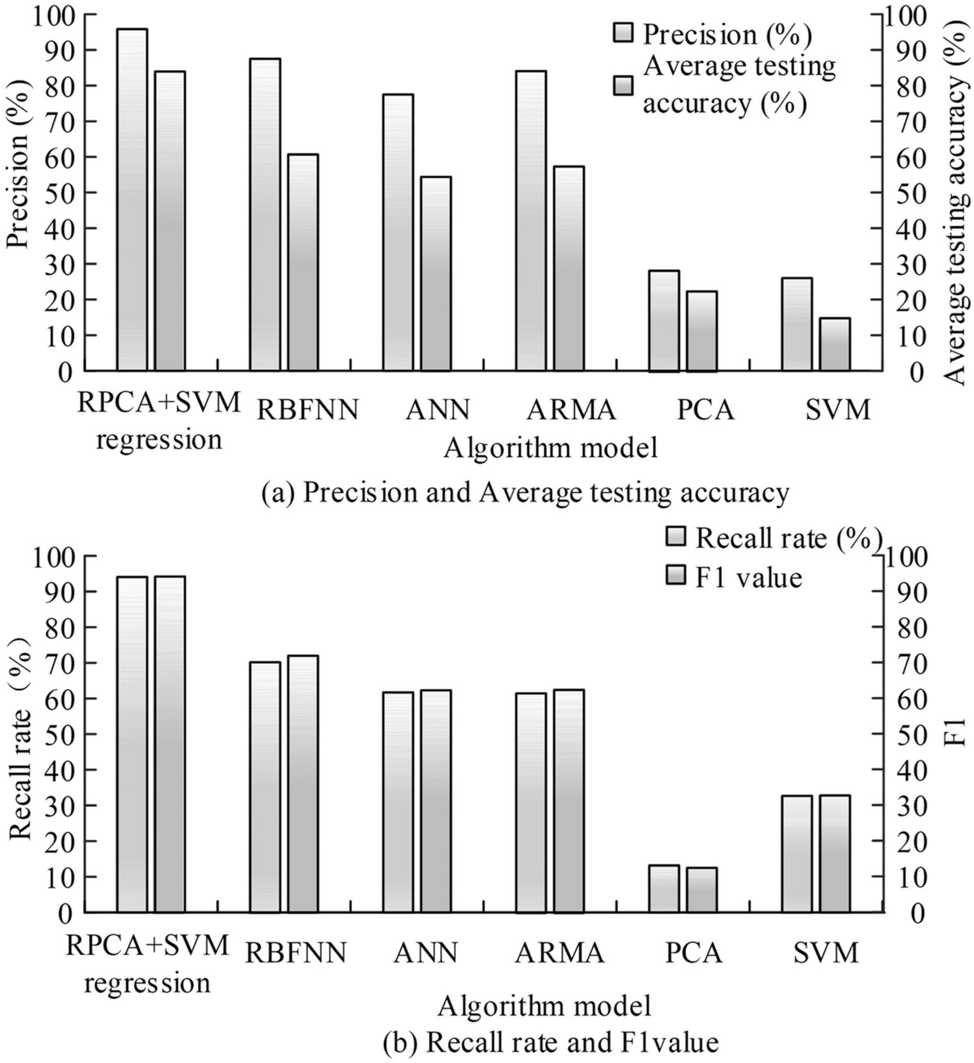


Figure 10. Comparison and angle difference results of the prediction model.

The proposed research method is also based on the combination of feature classification and predictive analysis to address the error problem of transformers, and it has been verified that it can effectively identify the error feature results of the model. Comparing the results of the research method, it can be seen that the classification algorithm can effectively classify feature or fault problems, which has significant value for problem solving. Subsequently, the recognition accuracy of the proposed hybrid model was analyzed, and the results are shown in Figure 11.

The results in Figure 11 indicate that the accuracy, recall, and F1 values of the algorithm proposed in the study are higher than those of other comparative algorithms. Specifically, the accuracy and average accuracy of the algorithm proposed in the study exceeded 80%, with values of 96.23% and 85.12%, respectively. The accuracy rates of RBFNN, ANN, ARMA, PCA, and SVM models were 87.12%, 78.56%, 82.11%, 25.39%, and 24.87%, respectively. In terms of recall rate and F1 value, the algorithms under RPCA and support vector machine regression exhibit a value of 95, with a difference of at least 20 compared to other algorithms. The worst performing PCA and SVM models have a recall rate and F1 value of no more than 40, which is the most significant difference compared to the mixed model proposed in the study.

By analyzing the above results, it can be seen that the proposed method has good advantages and performance in error identification and classification of EVTs. The reason is that the recursive principal component analysis is a data separation method that can decompose complex measurement data into independent components, thereby helping to identify and separate the error components of transformers. RPCA analysis can effectively handle the aliasing and correlation between multiple signals and has advantages for complex data separation in transformer error analysis. Secondly, support vector machine regression is a machine learning algorithm that can establish a nonlinear relationship model between input features and output values. Moreover, incorporating regression analysis and real-time update of step length on support vector machines can ensure that the actual error results are small, and thus grasp the relationship between the characteristics of measurement data and the actual error, achieving real-time prediction and evaluation of transformer errors. RPCA and SVR are both flexible and scalable algorithms. They can adapt to datasets of different sizes and complexities, provide flexible modeling and analysis capabilities, take into account the distribution characteristics of the data, and improve the robustness and generalization ability of the model. Compared to radi-



(a) Precision and Average testing accuracy
(b) Recall rate and F1value

Figure 11. Accuracy, recall, and F1 values of different algorithms.

al basis function neural networks and artificial neural networks, the combination of recursive principal component analysis and support vector machine regression is more explainable and understandable, as they provide clear feature separation and modeling processes. Compared with autoregressive moving average and principal component analysis, RPCA and SVR methods are more suitable for modeling and analyzing complex and nonlinear data and can provide more accurate results. Subsequently, stability analysis was conducted on the proposed transformer error detection algorithm. The stability analysis result is the average value after repeated experiments, and the selected time period is long and stable, with the results shown in Figure12.

In Figure 12, the variance results of the ratio difference and angle difference show that the amplitude changes of the two feature quantities are relatively stable in different test days, which can better reflect the changes in the time series. The ratio values are generally less than 1.8 and 1.6, and the response time of both is basically less than 0.1 seconds, demonstrating good application effects. Subsequently, an error evaluation was conducted, and the evaluation results are shown in Figure 13.

The evaluation results under the mixed model in Figure 13 indicate that, in most cases, the research method results in an evaluation threshold greater than the evaluation difference. Moreover, when the testing time range of three-phase EVT is greater than 1.5 days, the error results reflected by the ratio exceeding the tolerance are also within the accuracy range. A. The error evaluation range shown by phases B and C does not exceed 0.1%, indicating that this method has shown good accuracy in transformer analysis. This result has certain similarities with the research content proposed by Shahbazi *et al.* [28]. They used signal processing technology and artificial intelligence technology to identify and classify relay faults, as well as software simulation. Their results revealed that the signal features under time transformation exhibit good performance of the classifier, and the transformer under intelligent combination method can better identify disturbance and fault problems [28].

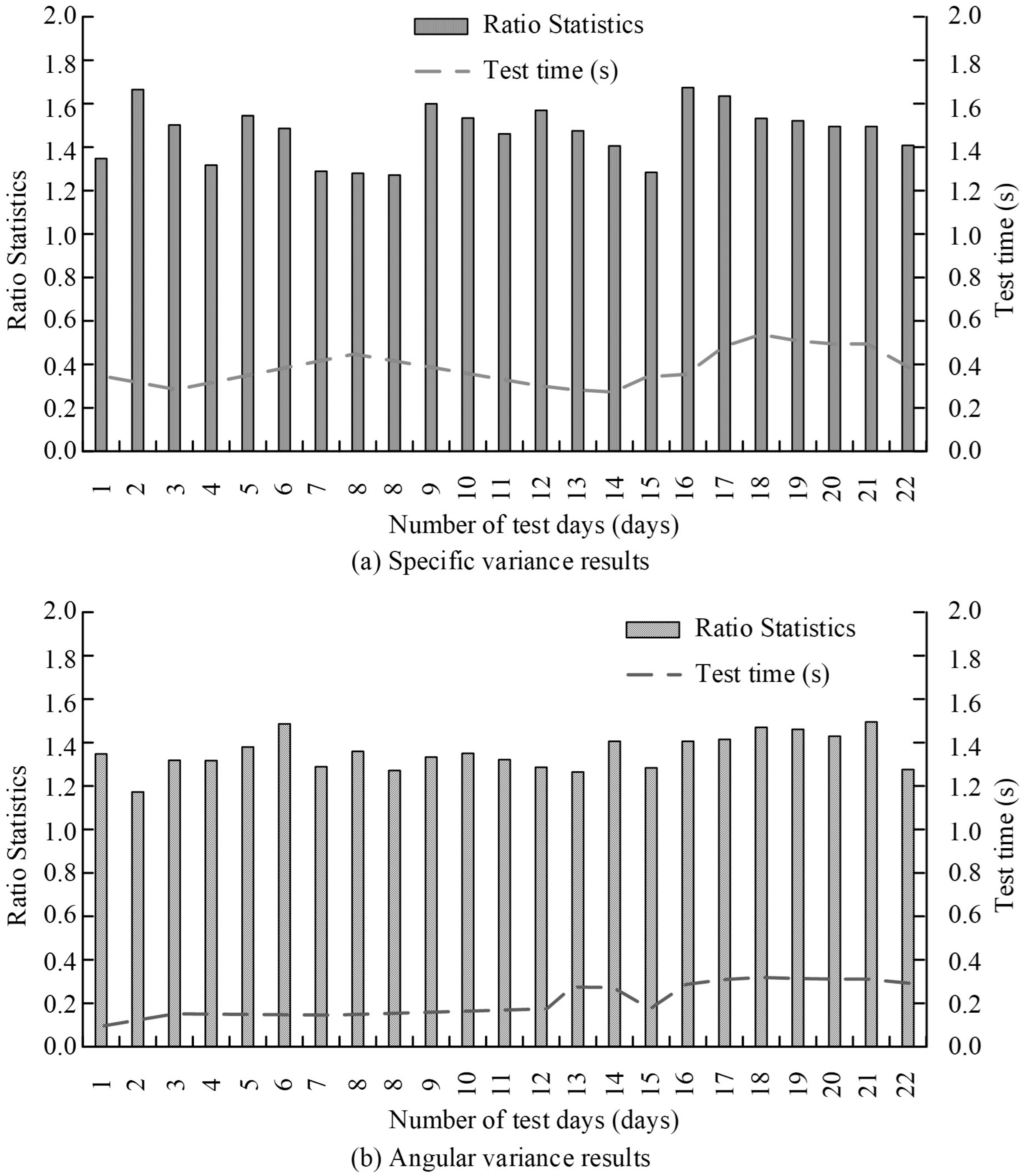


Figure 12. Stability analysis of the data.

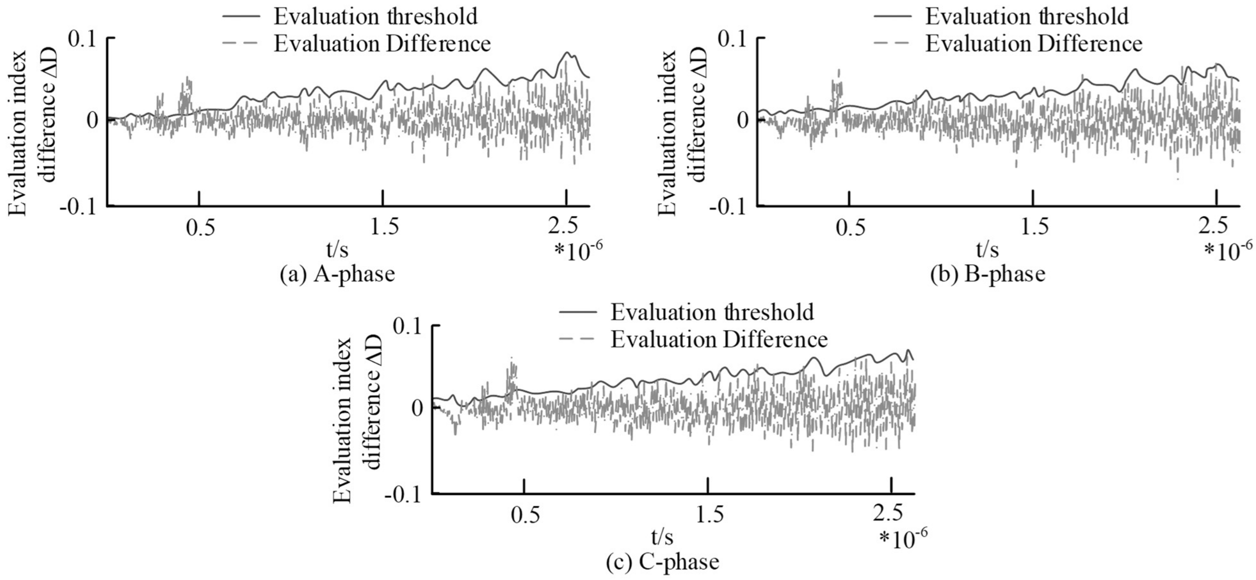


Figure 13. Evaluation results under the mixed model.

The advantages of the combination method reflected in this result for relays and the performance of the proposed combination model for transformer error analysis demonstrate that only by accurately identifying problem features can the performance of the algorithm be effectively improved, and the application effect can be improved [29].

4. Conclusion

The study proposes a data mining-based framework to address key challenges in electronic voltage transformer error diagnosis. The designed error analysis algorithm and model were tested, and the results showed that the error results of the data sampling value under the standard transformer showed significant transformation. The measurement error variation of phase A and phase C reached 0.25% and -0.07%. The measurement errors of phases A, B, and C basically reached 0.09%, 0.06%, and 0.03% when the sample size was greater than 11,000, making it difficult to effectively identify and detect abnormal data. The error results of the three-phase transformer under the improved method are relatively small, and the overall curve change is relatively stable. The error curves of phase A and phase C have consistency in the sample size range of 4,000~6,200, indicating that they can effectively identify abnormal error situations. In the analysis of error results, the algorithms' ranking based on model predictive performance is: The research algorithm > RBFNN > ANN > ARMA. The average deviation of the research algorithm is less than 0.05%, and the fitting value is 0.97. The predicted curve trend has high consistency with the actual curve trend. The RBFNN, ANN, and ARMA models showed average equal error deviations of 2.12%, 3.05%, and 4.23%, with fitting values of 0.89, 0.82, and 0.77, respectively. At the same time, the hybrid model can better predict the error state of the transformer. The average error of the ratio difference characteristic results is significantly less than that of the single model (0.023 < 0.046) and it has better stability and accuracy. The main achievement of the research is to use the design error analysis model to analyze EVTs. The average error of the ratio difference and angle difference characteristics displayed by the model is significantly smaller than that of other comparison algorithms, and the amplitude change of the research model is relatively stable, which can better reflect the changes in the time series. The ratios are generally less than 1.8 and 1.6. Moreover, the overall response time is relatively short, and the response time of the ratio difference and angle difference of the transformer is generally less than 0.1 seconds. The error evaluation range of transformer A, B, and

C phases does not exceed 0.1%, indicating that this method has good accuracy and high evaluation results in transformer analysis.

The ratio difference feature refers to the difference between the measurement results of the transformer and the true value, which can provide important information about the accuracy of the transformer measurement. A lower ratio difference characteristic value indicates that the measurement result is closer to the true value, while a higher ratio difference characteristic value indicates that there is a significant error in the measurement result. In error analysis, the ratio difference feature can help determine the size and distribution of errors and provide a basis for calibration and correction. The angular difference feature refers to the phase difference in the measurement results of the transformer. The phase difference of the transformer is crucial for the accuracy and stability of the measurement results. A smaller angular difference feature value indicates that the phase of the measurement result is very close to the true value, while a larger angular difference feature value indicates a significant phase difference, which may lead to inaccuracy in the measurement result. In the analysis of transformer error, the angular difference feature can provide information about phase error, helping to determine the operating status and error mode of the transformer. Analyzing the ratio difference and angle difference features can comprehensively evaluate the error mode and performance of the transformer, thereby optimizing the calibration and correction strategies of the transformer. The ratio difference feature and angle difference feature are key indicators for evaluating the accuracy and stability of transformers.

From a practical perspective, this study provides a new method for analyzing and evaluating the error of transformers in smart grids. The proposed method can be used for the classification and identification of EVT errors in power systems, real-time monitoring of transformer accuracy and health status, and timely detection and identification of possible error patterns. This method exhibits high accuracy and stable performance in better identifying different types of error patterns. Additionally, its real-time monitoring and evaluation ability in classifying and identifying error problems helps to take corresponding measures for correction and calibration to ensure the stable operation of the power system. The method is applied to power systems of different scales and properties to meet the needs of various practical application scenarios.

By improving the accuracy and monitoring effectiveness of transformer error assessment, the operational safety of the power system and measurement and control protection devices can be better guaranteed. This is of great significance for the power industry, as it can improve the stability and reliability of the power grid, reducing potential faults and losses. In future research, in order to further optimize and improve data mining-based transformer error analysis algorithms, it is necessary to explore more feature statistics and pattern recognition methods, expand the sample size range for transformer error data selection, and enhance the monitoring ability of transformer operation status. At the same time, the evaluation of the relationship between transformer error and power system performance needs to be strengthened and strategies for the operation of the power system need to be provided. The methods for solving Q statistic feature quantities should be enriched or deep learning and neural networks methods should be introduced to further improve the prediction and recognition capabilities of transformer errors. With the increasing integration of electronic voltage transformers, the developed error analysis technology provides crucial benefits in preventing catastrophic faults and power outages through timely diagnosis and prediction.

Acknowledgement

Inner Mongolia Power (Group) Co., Ltd. Technology Project: Research on Error State Evaluation Technology for 500 kV Transformer under Operating Condition. (2023-5-26)

References

[1] M. Grbić *et al.*, "Analysis of Influence of Measuring Voltage Transformer Ratio Error on Single-circuit Overhead Power Line Electric Field Calculation Results", *Electric Power Systems Research*, vol. 166, pp. 232–240, 2019. http://dx.doi.org/10.1016/j.epsr.2018.10.001

[2] S. G. Moshira *et al.*, "Image Retrieval Based on Deep Learning", *Journal of System and Management Sciences*, vol. 12, no. 2, pp. 477–496, 2022. http://dx.doi.org/10.33168/JSMS.2022.0226

[3] A. S. Dobakhshari *et al.*, "Online Non-Iterative Estimation of Transmission Line and Transformer Parameters by SCADA Data", *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2632–2641, 2021. http://dx.doi.org/10.1109/TPWRS.2020.3037997

[4] C. He *et al.*, "Evolutionary Large-Scale Multiobjective Optimization for Ratio Error Estimation of Voltage Transformers", *IEEE Transactions on Evolutionary Computation*, vol. 24, no. 5, pp. 868–881, 2020. http://dx.doi.org/10.1109/TEVC.2020.2967501

[5] U. M. Rao *et al.*, "Identification and Application of Machine Learning Algorithms for Transformer Dissolved Gas Analysis", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 28, no. 5, pp. 1828–1835, 2021. http://dx.doi.org/10.1109/TDEI.2021.009770

[6] H. D. Singh and J. Singh, "Enhanced Optimal Trained Hybrid Classifiers for Aging Assessment of Power Transformer Insulation Oil", *World Journal of Engineering*, vol.17 no. 3, pp. 407–426, 2020. http://dx.doi.org/10.1108/WJE-11-2019-0339

[7] X. Huang *et al.*, "Fault Diagnosis of Transformer Based on Modified Grey Wolf Optimization Algorithm and Support Vector Machine", *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 15, no. 3, pp. 409–417, 2020. http://dx.doi.org/10.1002/tee.23069

[8] M. Martinc *et al.*, "TNT-KID: Transformer-based Neural Tagger for Keyword Identification", *Natural Language Engineering*, vol. 28, no. 4, pp. 409–448, 2022. http://dx.doi.org/10.1017/S1351324921000127

[9] S. Seyedshenava and A. Ahmadpour, "Finite Element Method for Optimal Transformer Connection Based on Induction Motor Characteristics Analysis", *Shams Engineering Journal*, vol. 12, no. 2, pp. 1943–1957, 2021. http://dx.doi.org/10.1016/j.asej.2020.12.008

[10] W. Liao *et al.*, "Fault Diagnosis of Power Transformers using Graph Convolutional Network", *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 241–249, 2021. http://dx.doi.org/10.17775/CSEEJPES.2020.04120

[11] A. I. Khalyasmaa *et al.*, "Analysis of the State of High-Voltage Current Transformers Based on Gradient Boosting on Decision Trees", *IEEE Transactions on Power Delivery*, vol. 36, no. 4, pp. 2154–2163, 2021. http://dx.doi.org/10.1109/TPWRD.2020.3021702

[12] X. Yang *et al.*, "A Hybrid Machine-Learning Method for Oil-Immersed Power Transformer Fault Diagnosis", *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 15, no. 4, pp. 501–507, 2020. http://dx.doi.org/10.1002/tee.23081

[13] H. Kim and S. Madhavi, "A Reinforcement Learning Model for Quantum Network Data Aggregation and Analysis", *Journal of System and Management Sciences*, vol. 12, no. 1, pp. 283–293, 2022. http://dx.doi.org/10.33168/JSMS.2022.0120

[14] J. Jiang *et al.*, "Dynamic Fault Prediction of Power Transformers Based on Lasso Regression and Change Point Detection by Dissolved Gas Analysis", IEEE Transactions on Dielectrics and Electrical Insulation, vol. 27, no. 6, pp. 2130–2137, 2020. http://dx.doi.org/10.1109/TDEI.2020.008984

[15] A. G. C. Menezes *et al.*, "Induction of Decision Trees to Diagnose Incipient Faults in Power Transformers", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 29, no. 1, pp. 279–286, 2022. http://dx.doi.org/10.1109/TDEI.2022.3148453

[16] J. Meng *et al.*, "Online Monitoring Technology of Power Transformer based on Vibration Analysis", *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 554–563, 2021. http://dx.doi.org/10.1515/jisys-2020-0112

[17] J. Zhang *et al.*, "Applications of Artificial Neural Networks in Microorganism Image Analysis: A Comprehensive Review from Conventional Multilayer Perceptron to Popular Convolutional Neural Network and Potential Visual Transformer", *Artificial Intelligence Review*, vol. 56, no. 2, pp. 1013–1070, 2023. http://dx.doi.org/10.1007/s10462-022-10192-7

[18] Y. Zhang *et al.*, "Power Transformer Fault Diagnosis Considering Data Imbalance and Data Set Fusion", *High Voltage*, vol. 6, no. 3, pp. 543–554, 2021. http://dx.doi.org/10.1049/hve2.12059

[19] L. Wang *et al.*, "Gaussian Process Multi-Class Classification for Transformer Fault Diagnosis Using Dissolved Gas Analysis", *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 28, no. 5, pp. 1703–1712, 2021. http://dx.doi.org/10.1109/TDEI.2021.009470

[20] A. Doolgindachbaporn *et al.*, "Data Driven Transformer Thermal Model for Condition Monitoring", *IEEE Transactions on Power Delivery*, vol. 37, no. 4, pp. 3133–3141, 2022. http://dx.doi.org/10.1109/TPWRD.2021.3123957

[21] Q. Zhou *et al.*, "Semantic Communication With Adaptive Universal Transformer", *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 453–457, 2022. http://dx.doi.org/10.1109/LWC.2021.3132067

[22] Y. Guo *et al.*, "Spam Detection Using Bidirectional Transformers and Machine Learning Classifier Algorithms", *Journal of Computational and Cognitive Engineering*, vol. 2, no. 1, pp. 5–9, 2022. http://dx.doi.org/10.47852/bonviewJCCE2202192

[23] R. Dastres and M. Soori, "Artificial Neural Network Systems", *International Journal of Imaging and Robotics (IJIR)*, vol. 21, no. 2, pp. 13–25, 2021. [Online]. Available: https://hal.science/hal-03349542

[24] J. Contreras *et al.*, "ARIMA Models to Predict Next-Day Electricity Prices", *IEEE Power Engineering Review*, vol. 22, no. 9, pp. 57–57, 2002. http://dx.doi.org/10.1109/MPER.2002.4312577

[25] T. Baklacioglu, "Predicting the Fuel Flow Rate of Commercial Aircraft via Multilayer Perceptron, Radial Basis Function and ANFIS Artificial Neural Networks", *The Aeronautical Journal*, vol. 125, no. 1285, pp. 453–471, 2021. http://dx.doi.org/10.1017/aer.2020.119

[26] Z. Li *et al.*, "Classification and Location of Transformer Winding Deformations using Genetic Algorithm and Support Vector Machine", *Recent advances in electrical & electronic engineering*, vol. 14, no. 8, pp. 837–845(9), 2021. http://dx.doi.org/10.2174/2352096514666211026142216

[27] Z. Wu *et al.*, "Improved Genetic Algorithm and XGBoost Classifier for Power Transformer Fault Diagnosis", *Frontiers in Energy Research*, vol. 9, p. 745744, 2021. http://dx.doi.org/10.3389/fenrg.2021.745744

[28] N. Shahbazi *et al.*, "Identification and Classification of Cross-country Faults in Transformers using K-NN and Tree-based Classifiers", *Electric Power Systems Research*, vol. 204, p. 107690, 2022. http://dx.doi.org/10.1016/j.epsr.2021.107690

[29] H. B. Divyashrec *et al.*, "Multi Objective Energy based Hybrid Optimization Algorithm for Clustering and Routing in WSN", *Journal of System and Management Sciences*, vol. 12, no. 1, pp. 480–497, 2022. http://dx.doi.org/10.33168/JSMS.2020.0131

Received: August 2023
Revised: September 2023
Accepted: September 2023

Contact addresses:
Deqian Kou
Inner Mongolia Ehv Power Supply Company
Hohhot
Inner Mongolia
China
e-mail: 18702732489@163.com
*Corresponding author

Yan Su
Inner Mongolia Ehv Power Supply Company
Hohhot
Inner Mongolia
China
e-mail: suy235@126.com

Deqian Kou received his MSc degree in industrial engineering. He is a senior engineer and a senior political engineer. He is also the director of the enterprise management department of the Inner Mongolia Ehv Power Supply Company. His activities mainly include business management, marketing, and energy metering technology.

Yan Su received his MSc degree in electrical engineering. He is s senior engineer and the director of the technical office of the energy measurement center of the Inner Mongolia Ehv Power Supply Company. His activities mainly include marketing and energy metering technology.