Journal of Network Intelligence ©2026 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **11**, Number **1**, February **2026**

Relay Protection Anomaly Diagnosis Based on Bayesian Optimization Multicore Extreme Learning Machine

Jun-Yang Tian[1], Hai-Yong Li[1], Lian-Dian Jiang[1], Zi-Zheng Shen[1], Cong-Yun Wu[1]
[1]Power Dispatching Control Center,
GUANGXI Power Grid Co., Ltd, Nanning 530023, P. R. China
sumtjy@163.com, li_hy.dd@gx.csg.cn, gxzdbhk6700@163.com,
shen_zz.dd@gx.csg.cn, wu_cy.dd@gx.csg.cn

Yue Zhou[2]
[2]Johns Hopkins University, Baltimore 21210, USA
cocokeji@163.com

Jian Wang[3,*]
[3]Shandong University, Jinan 250100, P. R. China
wjianvip@126.com

*Corresponding author: Jian Wang

Received September 05, 2024, revised December 09, 2024, accepted March 21, 2025.

Abstract. *Traditional relay protection systems are often difficult to meet the requirements for fast and accurate diagnosis of anomalies when facing the complex and changing modern power networks. With the expansion of power system scale and the development of smart grid, the demand for efficient and reliable relay protection anomaly diagnosis methods is becoming more and more urgent. To address these challenges, this study proposes a Multi-Core Extreme Learning Machine (BO-MKELM) model based on Bayesian optimisation. Firstly, through the multicore learning technique, the model is able to effectively process high-dimensional, nonlinear and time-varying operational data in power systems and extract features that can accurately reflect the system state. Secondly, by combining the fast learning characteristics of the extreme learning machine and the parameter tuning advantages of Bayesian optimisation, the BO-MKELM model is able to meet the millisecond response requirements of the relay protection system while ensuring high diagnostic accuracy. The experimental results show that on the complete dataset, the BO-MKELM model achieves a diagnostic accuracy of 97.8%, which is 1.7% higher than the best-performing comparison method, CNN. On the small-scale dataset, the BO-MKELM model still maintains an accuracy of 94.9%, which is at least 3% higher than other methods. Meanwhile, the average diagnosis time of the BO-MKELM model is only 4.5ms, which is much faster than other deep learning methods, and fully meets the real-time requirements of relay protection systems.*
Keywords: relay protection; anomaly diagnosis; multicore extreme learning machine; Bayesian optimisation; real-time monitoring.

1. **Introduction.** The continuous expansion of the scale and complexity of the power system makes the role of relay protection in ensuring the safe and stable operation of the power grid more and more critical. As the guardian of the power system, relay protection shoulders the important responsibility of quickly identifying and isolating fault areas [1],

preventing the spread of faults, and safeguarding the safety of equipment and personnel. However, the wide application of smart grids [2], renewable energy [3] and distributed generation [4] has brought unprecedented challenges to the relay protection system, which is required to maintain efficient and reliable operation in a more complex and changing environment.

Rapid advances in artificial intelligence and big data technologies provide new ideas and methods to address these challenges. Machine learning and deep learning algorithms show great potential in fault detection [5], classification [6] and localisation [7]. Nevertheless, these advanced methods still face many obstacles in practical applications. The high-dimensional, non-linear and time-varying characteristics of power system operation data make it difficult for traditional machine learning methods to handle them effectively. Although deep learning models perform well in complex data processing, their demand for large amounts of training data and computational resources is difficult to meet in certain practical scenarios. In addition, how to meet the millisecond response requirement while ensuring diagnostic accuracy is still a difficult problem to be solved.

The field of relay protection anomaly diagnosis faces multiple challenges and is in dire need of innovative solutions. The first task is to efficiently process high-dimensional, nonlinear and time-varying operational data in power systems and extract features that accurately reflect the system state [8]. Secondly, it is crucial to design a diagnostic model that can adapt to the complex and variable grid environment while maintaining a high accuracy rate in the face of data scarcity. At the same time, the stringent real-time requirements of relay protection systems mean that the diagnostic model must achieve millisecond response speed while ensuring high accuracy. In addition, the model needs to have the ability to adapt to the dynamic changes of the power system and maintain stable performance under all types of abnormal conditions [9]. Improving the interpretability of the diagnostic model is also an important issue, and a transparent decision-making process can help enhance the trust of power system operation and maintenance personnel in the model. Finally, achieving efficient training and deployment of models under limited computational resources is the key to ensure the practicality of the scheme.

1.1. **Related work.** Traditional methods for relay protection anomaly diagnosis mainly rely on techniques such as threshold judgement, signal processing and expert systems. Guo et al. [10] proposed a mathematical model-based fault detection method to avoid blind tripping by quantifying the boundaries of fault detection. It is shown that this method improves the accuracy of fault detection, but the performance in complex grid environments and measurement uncertainty conditions still needs to be further verified. Rahmati and Adhami [11] developed a fault diagnosis technique using phase angle and magnitude of positive and zero sequence voltages. Although this method shows good potential in identifying different types of faults, the lack of adaptivity in the selection of fault classification thresholds limits its application in dynamically changing grids. Dashti and Sadeh [12] proposed a node impedance matrix based fault ranging algorithm for distribution networks. This method demonstrated high accuracy in locating fault points, however, its effectiveness and applicability are still challenged when dealing with modern grids containing distributed energy sources and complex branch circuits.

With the development of artificial intelligence and big data technology, the mainstream methods of relay protection anomaly diagnosis gradually shift to machine learning and deep learning techniques. Panigrahi et al. [13] proposed a microgrid fault protection system combining wavelet energy and neural networks. It was found that this method significantly improves the accuracy and speed of fault identification, but there is still room for improving its robustness and computational efficiency when dealing with complex and rapidly changing fault scenarios. Gashteroodkhani et al. [14] developed an intelligent protection scheme based on Deep Belief Networks (DBN) and Time-Time Transform. Although this approach shows advantages in dealing with large-scale data and complex failure modes, the large amount of data and computational resources required for model training may limit its generalisation in some practical application scenarios. Kim et al. [15] proposed a Long Short-Term Memory (LSTM)-based fault diagnosis method for relay protection testing. This method performs well in processing sequential data and capturing long-term dependencies, but still has some limitations in real-time diagnosis and model interpretability. Qu et al. [16] investigated an insulated overhead conductor fault detection method based on wavelet transform and LSTM. This method achieved significant results in extracting fault features and identifying complex fault patterns, but its generalisation ability under different environmental conditions needs to be further verified. Rai et al. [17] developed a convolutional neural network-based fault classification method for distribution networks, especially for systems integrating distributed generation. Although this method performs well in handling high-dimensional data and extracting spatial features, there is still room for improvement in handling temporal information and adapting to network topology changes.

1.2. **Motivation and contribution.** Existing methods for relay protection anomaly diagnosis mainly have the following limitations: firstly, traditional methods usually rely on fixed thresholds or preset rules, which are difficult to adapt to the complex and changing grid environment, while machine learning-based methods often require a large amount of labelled data for training, and may face the problem of difficult data acquisition in practical applications. Secondly, many methods are still deficient in model parameter optimisation and real-time performance, which makes it difficult to find the optimal model configuration, and at the same time may limit their application in real power systems. Finally, the lack of in-depth consideration of model robustness and interpretability affects the reliability and credibility of the diagnostic results.

To solve the above problems, this paper proposes a Multi-Kernel Extreme Learning Machine (MKELM) model based on Bayesian optimisation, which improves the effectiveness of relay protection anomaly diagnosis in intelligent substations. The main innovations and contributions of this work include:

(1) Aiming at the challenges of the complex and variable grid environment and the difficulty in acquiring training data, the proposed MKELM model improves the flexibility and generalisation ability of the model through the combination of multiple kernel functions. The use of extreme learning machine as the base model has the advantages of fast training speed and small required sample size. This approach can better adapt to different types of faults and changing network conditions, while effectively alleviating the problem of data scarcity and significantly improving the diagnostic performance of the model in complex grid environments.

(2) In order to solve the problems of insufficient optimisation of model parameters and real-time performance requirements, a Bayesian optimisation algorithm is introduced. This method automatically tunes the key parameters of MKELM through an intelligent search strategy, which not only improves the model performance, but also reduces manual intervention. Combining the fast learning characteristics of the extreme learning machine and the efficient search capability of Bayesian optimisation, the real-time response capability of the model is significantly improved while the diagnostic accuracy is guaranteed, making it more suitable for application in real power systems.

(3) In order to improve the robustness and interpretability of the model, the model's ability to handle noisy data and abnormal samples is enhanced by the multicore technique. At the same time, the credibility of the diagnostic results is enhanced by analysing the multicore weights and the Bayesian optimisation process, which provides partial interpretability of the model decisions. This approach not only improves the stability of the model in complex environments, but also provides more explanatory basis for the diagnostic results, which contributes to fault analysis and decision support in practical applications.

2. Analysis of abnormal relay protection diagnostic problems.

2.1. Introduction to intelligent substation relay protection testing. Intelligent substation relay protection devices play a crucial role in the stability and reliability of power system operation. Testing and inspection of these devices is an important part of ensuring their safe and reliable operation. In the field commissioning and factory inter-commissioning stage of intelligent substation relay protection devices, the digital relay protection test device is the most important debugging tool.

Current test devices can diagnose their own faults during use, but there is a significant limitation: it is not possible to realise automatic diagnosis of anomalies or faults occurring in the whole system, especially in the tested protection device, during the testing process. This results in technicians needing to spend a lot of time on manual troubleshooting, which significantly reduces test efficiency.

Relay protection testing in smart substations is more complex than traditional methods, mainly in the following aspects:

(1) Data interaction: The widespread use of digitised information has made it necessary for testing to take into account a wider range of data handling processes.

(2) Diversity of devices: The variety of smart devices increases the complexity of testing and requires consideration of compatibility between devices of different manufacturers and models.

(3) Flexible configuration: the characteristics of intelligent substations require test programmes with strong adaptability and scalability.

(4) Communication protocols: The implementation of communication protocols such as IEC 61850 needs to be verified.

(5) System co-ordination: It is necessary to check the co-ordination and cooperation between the protection device and other equipment of the intelligent substation.

To cope with these challenges, modern relay protection test systems usually adopt digital and intelligent test equipment that can simulate complex fault scenarios, automatically generate test reports, and support remote operation and monitoring. However, despite the continuous progress in testing technology, accurately identifying and diagnosing anomalies in relay protection systems is still an urgent problem to be solved.

Improving the efficiency of intelligent substation relay protection testing and achieving automatic diagnosis of faults occurring during the entire testing process has become one of the key directions of current research. This can not only significantly reduce the manual fault troubleshooting time, but also improve the accuracy and consistency of fault diagnosis, thus further enhancing the overall reliability and security of the power system.

2.2. Relay protection anomaly types and characteristics. Abnormalities in the relay protection system of intelligent substations can be classified into various types, each of which has its own unique characteristics. Understanding these types of anomalies and their characteristics is essential for designing effective diagnostic methods.

Firstly, hardware failures are a common type of anomaly. This includes, but is not limited to, EEPROM errors, FLASH errors, RAM self-test errors, SRAM errors, and CPU plug-in errors of the protection device. These faults are usually manifested in the form of failure of the device to start up normally, abnormal data read/write, or reduced processing capability. Let H_f denote the hardware fault characteristic vector, it can be expressed as:

$$H_f = [h_1, h_2, \ldots, h_n] \tag{1}$$

where h_i denotes the eigenvalue of the ith hardware fault and n is the total number of hardware fault types.

Second, software anomalies are another important type of anomaly. This includes problems such as self-test alarms, operational anomalies and software version mismatches of the protection device. Software anomalies may result in protection failure, trip logic errors, or communication interruptions. Define the software anomaly feature vector S_f as:

$$S_f = [s_1, s_2, \ldots, s_m] \tag{2}$$

where s_j denotes the eigenvalue of the jth software anomaly and m is the total number of software anomaly types.

Communication failures are another key type of anomaly in smart substations. These include SV (Sampled Value) bus anomalies, GOOSE (Object Oriented Substation Event) communication anomalies, and clock synchronisation failures. Communication failures may result in delayed data transmission, packet loss, or inconsistent information. Define the communication fault feature vector C_f as:

$$C_f = [c_1, c_2, \ldots, c_k] \tag{3}$$

where c_k denotes the eigenvalue of the kth communication fault and k is the total number of communication fault types.

In addition, protection action abnormality is also a type of abnormality that needs to be focused on. This includes protection malfunction, refusal to act and abnormal action time. These types of anomalies directly affect the safe and stable operation of the power system. Define the protection action anomaly feature vector P_f as:

$$P_f = [p_1, p_2, \ldots, p_l] \tag{4}$$

where p_l denotes the eigenvalue of the lth protective action exception and l is the total number of protective action exception types.

Considering the complexity and variability of the grid environment, these anomaly types may occur simultaneously or interact with each other. Therefore, we propose a composite anomaly feature vector F, which combines the features of all the above anomaly types:

$$F = [H_f, S_f, C_f, P_f] \tag{5}$$

This integrated feature vector will be used as input to subsequent multicore extreme learning machine models for accurate diagnosis of complex anomalies.

By introducing this comprehensive feature vector, our model is able to consider multiple anomaly types at the same time, which improves the flexibility and generalisation ability of the model. This approach can not only adapt to different types of faults and changing network conditions, but also effectively deal with data scarcity problems that may occur in practical applications. By analysing the weight distribution of different anomaly types in the comprehensive feature vector, we can provide the interpretability of some model decisions, which is important for improving the credibility of diagnostic results and supporting practical fault analysis.

3. Theory of Multicore Extreme Learning Machines.

3.1. **ELM.** As an efficient single hidden layer feedforward neural network [18, 19], ELM has attracted much attention for its fast-learning speed and excellent generalisation performance. In this study, we choose ELM as the base model to cope with the complex and variable environment and data scarcity problem in relay protection anomaly diagnosis.

The core idea of ELM is to randomly generate input weights and biases [20], and then directly compute the output weights by parsing method, avoiding the time-consuming iterative training over in traditional neural networks, as shown in Figure 1. For a given N samples $(\mathbf{x}_i, \mathbf{y}_i)|\mathbf{x}_i \in \mathbb{R}^d, \mathbf{y}_i \in \mathbb{R}^m, i = 1, \ldots, N$, where $\mathbf{x}_i$ is the input feature vector and $\mathbf{y}_i$ is the corresponding output label, the mathematical model of ELM can be expressed as [21]:

$$\sum_{j=1}^{L} \beta_j g(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = \mathbf{y}_i, \quad i = 1, \ldots, N \tag{6}$$

where L is the number of hidden layer neurons; $g(\cdot)$ is the activation function; $\mathbf{w}_j$ and b_j are the input weight and bias of the jth hidden layer neuron, respectively; β_j is the output weight.

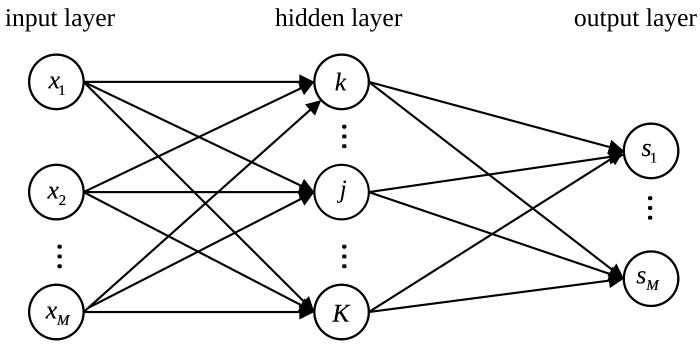

Figure 1. Structure of the ELM

Reduce the above equation to matrix form:

$$\mathbf{H}\beta = \mathbf{Y} \tag{7}$$

where β is the output weight matrix, $\mathbf{Y}$ is the target output matrix, and $\mathbf{H}$ is the hidden layer output matrix.

$$\mathbf{H} = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \ldots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \ldots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \tag{8}$$

The training process of ELM can be reduced to solving the following least squares problem:

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{Y}\|^2 \tag{9}$$

The analytical solution to the problem is:

$$\beta = \mathbf{H}^{\dagger}\mathbf{Y} \tag{10}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalised inverse of $\mathbf{H}$.

In the context of relay protection anomaly diagnosis, the input feature vector $\mathbf{x}_i$ corresponds to the comprehensive anomaly feature vector F defined in the previous section, and the output $\mathbf{y}_i$ represents the anomaly diagnosis result. In this way, ELM is able to quickly learn the mapping relationship between anomaly features and diagnosis results.

This property of ELM makes it particularly suitable for dealing with the challenges we face. Firstly, its fast learning capability helps to improve the real-time performance of the model, which meets the requirement of real-time diagnosis in power systems. Second, ELM also has good learning ability for small sample datasets, which helps to alleviate the data scarcity problem that may be faced in practical applications.

However, despite the above advantages of ELM, a single ELM model may still have limitations when dealing with highly nonlinear and complex relay protection anomaly diagnosis problems. Therefore, we will introduce the multi-core technique in the next section to construct a multi-core Extreme Learning Machine model to better adapt to the complex and changing grid environment.

3.2. **MKELM model.** In order to further enhance the flexibility and generalisation of the model, we introduce the multi-core technique to construct the MKELM model. MKELM is able to better capture the complex non-linear relationships of relay protection anomaly features by combining multiple core functions [22].

In MKELM, the hidden layer output matrix $\mathbf{H}$ is replaced by the kernel matrix $\boldsymbol{\Omega}$ whose elements are defined by the kernel function $K(\mathbf{x}_i, \mathbf{x}_j)$. The multikernel function $K(\mathbf{x}_i, \mathbf{x}_j)$ is a linear combination of several elementary kernel functions:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \sum_{p=1}^{P} \mu_p K_p(\mathbf{x}_i, \mathbf{x}_j) \tag{11}$$

where P is the number of elementary kernel functions and $\mu_p \geq 0$ is the weight of the pth kernel function and satisfies $\sum_{p=1}^{P} \mu_p = 1$.

The output of MKELM can be expressed as:

$$\mathbf{y}(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i K(\mathbf{x}, \mathbf{x}_i) \tag{12}$$

where α_i is the output weight, which can be obtained by solving the following optimisation problem:

$$\min_{\alpha} \|\boldsymbol{\Omega}\alpha - \mathbf{Y}\|^2 + \lambda\alpha^T\boldsymbol{\Omega}\alpha \tag{13}$$

Here, $\lambda > 0$ is the regularisation parameter, which is used to control the model complexity and generalisation ability. The solution of this optimisation problem is:

$$\alpha = (\boldsymbol{\Omega} + \lambda\mathbf{I})^{-1}\mathbf{Y} \tag{14}$$

where $\mathbf{I}$ is the unit matrix.

3.3. **Selection and combination of multiple kernel functions.** In MKELM, the selection and combination of kernel functions are crucial to the model performance. Considering the complexity of relay protection anomaly features, we choose to combine Gaussian Radial Basis Function (RBF) kernel and polynomial kernel to take into account both local and global features.

The RBF kernel function is defined as:

$$K_{\mathrm{RBF}}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\|\mathbf{x}_i - \mathbf{x}_j\|^2) \tag{15}$$

where $\gamma > 0$ is the kernel width parameter.

The polynomial kernel function is defined as:

$$K_{\mathrm{Poly}}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T\mathbf{x}_j + c)^d \tag{16}$$

where d is the order of the polynomial and $c \geq 0$ is a constant term.

Our multicore function is denoted as:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mu_1 K_{\mathrm{RBF}}(\mathbf{x}_i, \mathbf{x}_j) + \mu_2 K_{\mathrm{Poly}}(\mathbf{x}_i, \mathbf{x}_j) \tag{17}$$

where μ_1 and μ_2 are the weights of the kernel function satisfying $\mu_1 + \mu_2 = 1$.

Through the combination of MKELM and multicore functions, we construct a flexible and efficient relay protection anomaly diagnosis model with good generalisation capability. This approach not only adapts to the complex and changing grid environment, but also effectively handles the data scarcity problem in practical applications, while providing a certain degree of model interpretability.

This multi-core combination strategy has the following advantages:

(1) Improve model flexibility: the RBF kernel can effectively capture local features, while the polynomial kernel is good at extracting global features. By combining these two kernel functions, the model can better adapt to different types of relay protection anomalies.

(2) Enhanced generalisation: the combination of multiple cores can extract features at different scales, which helps to improve the generalisation ability of the model when dealing with unseen anomaly types.

(3) Provide interpretability: by analysing the weights of different kernel functions, we can understand the relative importance of local and global features in anomaly diagnosis, and enhance the interpretability of the model.

(4) Coping with data scarcity: multi-core combination can extract richer feature information on limited training data, which helps to alleviate the data scarcity problem.

However, the introduction of multiple kernel functions also brings new challenges in determining the optimal kernel function parameters (e.g. γ, d, c) and kernel function weights (μ_1, μ_2). To address this problem, we will introduce a Bayesian optimisation algorithm in the next chapter to automatically tune the key parameters of MKELM to further improve the model performance and reduce manual intervention.

4. **Bayesian optimisation based model for multi-core extreme learning machines.**

4.1. **Model structure design.**

Based on the MKELM theory introduced in the previous section, we designed a Bayesian optimisation-based multicore extreme learning machine (BO-MKELM) model for relay protection anomaly diagnosis. The model structure includes three main parts: input layer, multicore mapping layer and output layer.

The input layer receives the comprehensive anomaly feature vector F defined in the previous section. The multicore mapping layer maps the input features to a high-dimensional feature space using a multicore function. The output layer then gives the final anomaly diagnosis results. To improve the adaptability and generalisation of the model, we introduce a dynamic weight adjustment mechanism. Instead of fixed values, the kernel function weights μ_1 and μ_2 are dynamically adjusted according to the input features:

$$\mu_1(\mathbf{x}) = \sigma(\mathbf{w}_1^T \mathbf{x} + b_1) \tag{18}$$

$$\mu_2(\mathbf{x}) = 1 - \mu_1(\mathbf{x}) \tag{19}$$

where $\sigma(\cdot)$ is the sigmoid function, $\mathbf{w}_1$ and b_1 are the parameters to be optimised. This dynamic weight adjustment mechanism enables the model to adaptively adjust the importance of the kernel function according to different types of anomalous features, further improving the flexibility of the model.

The modified model output expression is:

$$y(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i(\mu_1 K_{\text{RBF}}(\mathbf{x}, \mathbf{x}_i) + \mu_2 K_{\text{Poly}}(\mathbf{x}, \mathbf{x}_i)) \tag{20}$$

where $\mathbf{x}$ is the input feature vector, $y(\mathbf{x})$ is the model output, and N is the number of training samples.

4.2. **Bayesian optimisation of multi-kernel parameters.** To automatically tune the key parameters of the BO-MKELM model, we introduce the Bayesian optimisation algorithm [23, 24]. In our model, the parameters to be optimised include the width parameter γ of the RBF kernel, the order d and constant term c of the polynomial kernel, the regularisation parameter λ, and the parameters $\mathbf{w}_1$ and b_1 in the dynamic weight adjustment mechanism.

Define the parameter vector $\theta = [\gamma, d, c, \lambda, \mathbf{w}_1, b_1]$ and the goal of Bayesian optimisation is to find the optimal θ^* that makes the model perform best on the validation set:

$$\theta^* = \arg\max_{\theta} f(\theta) \tag{21}$$

where $f(\theta)$ is the model's performance metric function (e.g., accuracy or F1 score) on the validation set.

The core idea of Bayesian optimisation is to construct a probabilistic model of the objective function $f(\theta)$, usually using a Gaussian Process (GP). In each iteration, the algorithm performs the following steps:

Step 1: Update the GP model based on the observed data points $\{(\theta_i, f(\theta_i))\}_{i=1}^{t}$.

Step 2: Select the next evaluation point θ_{t+1} using the capture function (Expected Improvement):

$$\theta_{t+1} = \arg\max_{\theta} EI(\theta) \tag{22}$$

where $EI(\theta)$ is the expected improvement function.

Step 3: Evaluate the new point $f(\theta_{t+1})$ and add it to the observation dataset.

Step 4: Repeat steps 1-3 until the predefined number of iterations is reached or the stop condition is met.

The introduction of Bayesian optimisation algorithms brings the advantages of automated parameter tuning, global optimisation, high sample efficiency and adaptability. This approach not only improves the adaptability and generalisation ability of the model, but also reduces manual intervention, making the model easier to apply in real engineering.

Through Bayesian optimisation, our BO-MKELM model is able to find the optimal parameter configurations automatically, thus maintaining high performance in different types of relay protection anomaly diagnosis tasks.

4.3. **Model training and optimisation process.** The pseudo-code for the training and optimisation process of the BO-MKELM model is shown in Algorithm 1.

This training and optimisation process fully reflects the innovative points of this paper. Firstly, through the introduction of Bayesian optimisation, we achieve automatic tuning of the key parameters of the model, which reduces manual intervention and improves the adaptability of the model in the complex and changing grid environment. Secondly, the introduction of the dynamic weight adjustment mechanism enables the model to adaptively adjust the importance of the kernel function according to different types of anomaly features, which further improves the flexibility and generalisation ability of the model.

Algorithm 1 Pseudo-code for BO-MKELM training and optimisation procedure

Input: training set $\mathcal{D}_{\text{train}}$, validation set $\mathcal{D}_{\text{val}}$, maximum number of iterations $T_{\max}$, performance thre

Output: optimised BO-MKELM model

begin
Initialise parameter search range and Gaussian process model GP
$\mathcal{D}_0 = \emptyset, t = 0, f_{\text{best}} = -\infty$
while $t < T_{\max}$ and $f_{\text{best}} < \eta$ **do**
  if $t = 0$ **then**
    $\theta_t =$ random sampling from search range
  else
    $\theta_t = \arg\max_\theta EI(\theta)$ //Use Desired Improvement to select the next evaluation point
  end if
  Calculate the kernel matrix $\mathbf{\Omega}$ using θ_t.
  Solving $\alpha = (\mathbf{\Omega} + \lambda\mathbf{I})^{-1}\mathbf{Y}$
  $f(\theta_t) = \text{F1}(\mathcal{D}_{\text{val}}, \theta_t)$ // Evaluate performance on validation set
  $\mathcal{D}_t = \mathcal{D}_{t-1} \cup \{(\theta_t, f(\theta_t))\}$
  Updating the gaussian process model GP using $\mathcal{D}_t$
  if $f(\theta_t) > f_{\text{best}}$ **then**
    $f_{\text{best}} = f(\theta_t)$
    $\theta_{\text{best}} = \theta_t$
  end if
  $t = t + 1$
end while
Retraining the MKELM model on $\mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{val}}$ using θ_{best}
return Optimised BO-MKELM model
end

In addition, this method can effectively deal with the problem of data scarcity. Bayesian optimisation is sample efficient and can find better parameter configurations on limited training data. The use of multicore techniques then allows the model to extract features at different scales and get more information from limited data.

Finally, by analysing the importance of different parameters in the optimisation process and the final kernel function weights, we can obtain a certain degree of model interpretability. This not only helps to understand the decision-making process of the model, but also provides insights for further improvement of the anomaly diagnosis method.

5. Implementation of relay protection anomaly diagnosis methods.

5.1. **Anomaly diagnosis index selection.** In order to comprehensively assess the operation status of the relay protection system, we selected a series of key indicators as the basis of abnormality diagnosis. These indicators cover hardware status, software operation, communication quality and protection action. Specifically, we define the anomaly diagnosis indicator vector I as follows:

$$I = [I_H, I_S, I_C, I_P] \tag{23}$$

The subvectors I_H, I_S, I_C, and I_P represent hardware status indicators, software operation indicators, communication quality indicators, and protection action indicators, respectively. Each sub-vector contains several specific indicators, e.g., I_H: CPU utilisation, memory usage, storage space usage, etc.; I_S: software version number, running time, self-test result, etc.; I_C: Communication delay, packet loss, clock synchronisation error, etc.; I_P: Action time, return value, number of actions, etc.

This multi-dimensional index selection method can comprehensively reflect the operational status of the relay protection system and help improve the accuracy and reliability of abnormality diagnosis.

5.2. **Data preprocessing.** Raw data often have problems such as noise, missing values and different magnitudes, in order to improve the performance of the BO-MKELM model, we use the following data preprocessing steps:

(1) Missing value treatment: For a small amount of missing data, we use k-nearest neighbour interpolation to fill in. Define the filled indicator vector as I'.

(2) Outlier detection and processing: the quartile-based method is used to detect outliers and replace them with boundary values in the upper and lower quartiles.

(3) Normalisation: In order to eliminate the differences in the scales of the different indicators, we use the Min-Max normalisation method to map all indicators to the interval [0, 1]:

$$I''_j = \frac{I'_j - \min(I'_j)}{\max(I'_j) - \min(I'_j)} \tag{24}$$

where I''_j is the jth indicator value after normalisation.

(4) Feature engineering: based on the domain knowledge, we constructed some composite features, such as the deviation rate of the protection action time from the set value, to enhance the expressive ability of the model.

The preprocessed feature vector is denoted as X, which will be used as input to the BO-MKELM model.

5.3. **Abnormality diagnosis algorithm flow.** Based on the BO-MKELM model proposed in the previous section, we design a complete relay protection anomaly diagnosis algorithm flow. The flow includes two phases: offline training and online diagnosis:

Offline training phase:

Step 1: Collect historical data and preprocess it to get the training dataset $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^N$, where X_i is the preprocessed feature vector, and y_i is the corresponding anomaly type label.

Step 2: Optimise the parameters of the MKELM model using the Bayesian optimisation algorithm to obtain the optimal parameter configuration θ^*.

Step 3: Train the MKELM model on the whole training set using the optimal parameter θ^* to get the final anomaly diagnosis model.

Online diagnostic phase:

Step 1: Collect the operation data of relay protection system in real time, and preprocess it to get the feature vector X_{new}.

Step 2: Input X_{new} into the trained BO-MKELM model to get the anomaly diagnosis result y_{pred}.

Step 3: Based on the diagnostic result y_{pred}, the system gives the corresponding warning or processing suggestion.

Step 4: Regularly update the model with newly collected data to accommodate dynamic changes in the system.

The algorithmic process fully embodies the innovative points of this paper. Firstly, the model is provided with high-quality input data through multi-dimensional indicator selection and fine data pre-processing. Secondly, the introduction of Bayesian optimisation achieves the automatic tuning of model parameters and improves the adaptability of the model in the complex grid environment. Further, the use of MKELM model not only improves the accuracy of anomaly diagnosis, but also effectively handles the problem of data scarcity. Finally, the regular update mechanism ensures that the model can continuously adapt to the dynamic changes of the system.

6. Simulation experiments and analysis of results.

6.1. Experimental platform and data set. The experimental platform of this study is constructed based on the RTDS (Real Time Digital Simulator) [25] and the IEC 61850 standard intelligent substation relay protection test system [26]. The RTDS system is used to simulate the operation status of the power system, including normal operation and various fault conditions. The intelligent substation relay protection test system is used to collect and process the operation data of relay protection devices. The hardware configuration of the experimental platform is as follows.

Processor: Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz

Memory: 128GB DDR4

Graphics: NVIDIA Tesla V100 16GB

Operating system: CentOS 7.6

The software environment mainly includes Python 3.8, TensorFlow 2.4, Scikit-learn 0.24 and GPy 1.9.9. The experimental dataset is derived from the actual operation data of a 500kV intelligent substation and the data generated by RTDS simulation system. The dataset consists of normal operation data and 14 typical protection abnormalities, such as hardware faults, software abnormalities, communication faults and protection action abnormalities. The detailed composition of the dataset is shown in Table 1.

In order to simulate the data scarcity problem that may be encountered in real engineering, we also constructed a small-scale dataset containing only 2,000 samples, with the proportions of each type of data remaining consistent with the full dataset.

This dataset design takes into account the various anomalies that may be encountered in the relay protection system, which helps to comprehensively evaluate the performance of our proposed BO-MKELM model in dealing with the complex and changing grid environment. The setting of the anomaly types allows us to evaluate the model's ability to identify different types of anomalies in a more detailed manner, which is in line with the innovations of the model's flexibility and generalisation capability emphasised in this paper.

Meanwhile, the introduction of a small-scale dataset allows us to validate the model's performance under data scarcity, which further demonstrates the advantages of the BO-MKELM model in dealing with the problem of insufficient data in real engineering.

6.2. Model performance assessment metrics. In order to comprehensively assess the performance of the BO-MKELM model in relay protection anomaly diagnosis, we choose four common assessment indicators, including Accuracy, Precision, Recall and F1 score. In addition to the above common indicators, considering the specificity of relay protection anomaly diagnosis, we also introduced two specialised assessment indicators:

Table 1. Composition of the experimental data set

No.	Data type	Subtype	Sample size	Percentage
1	Normal operation	-	5000	50%
2	Hardware failure	CPU failure	250	2.5%
3		Memory failure	250	2.5%
4		Storage device failure	250	2.5%
5		Power module failure	250	2.5%
6	Software anomaly	Programme crash	250	2.5%
7		Version mismatch	250	2.5%
8		Misconfiguration	250	2.5%
9		Algorithmic exception	250	2.5%
10	Communications failure	Network outage	500	5%
11		Protocol error	500	5%
12		Clock synchronisation failure	500	5%
13	Abnormal protective action	Malfunction	500	5%
14		Prevent from moving	500	5%
15		Motion delay	500	5%
Total	-	-	10000	100%

Average Diagnosis Time (ADT):

$$ADT = \frac{1}{N}\sum_{i=1}^{N} t_i \tag{25}$$

where N is the number of test samples and t_i is the diagnosis time of the model for the ith sample.

Reliability Index (RI):

$$RI = \frac{1}{N}\sum_{i=1}^{N} w_i \times c_i \tag{26}$$

where w_i is the weight of the ith category of anomalies (based on its severity) and c_i is the correct diagnosis rate for that category.

These evaluation indexes not only reflect the classification performance of the model comprehensively, but also take into account the special requirements of relay protection system on real-time and reliability. By evaluating the performance from multiple perspectives, we can more objectively analyse the performance of the BO-MKELM model in relay protection anomaly diagnosis.

In the following experiments, based on these evaluation metrics, we will compare and analyse the performance of the BO-MKELM model with that of other typical methods on both complete and small-scale datasets to validate the effectiveness and superiority of this method proposed in this paper. In particular, we will focus on the performance of the model in dealing with complex and changing grid environments, coping with data scarcity problems, and maintaining real-time diagnosis to fully reflect the innovative points of this paper.

6.3. **Experimental results and analysis.** In order to comprehensively evaluate the performance of the Bayesian optimisation-based multi-core extreme learning machine (BO-MKELM) model in relay protection anomaly diagnosis, a series of experiments have been conducted and compared with a number of comparative methods including the most recent ones.

First, we trained and tested the BO-MKELM model and other methods on the full dataset, using a 5-fold cross-validation method to ensure the reliability of the results. Table 2 shows the average performance of the BO-MKELM model and other methods on each assessment metric.

Table 2. Performance Comparison of Different Methods on the Complete Dataset

Method	Accuracy	Precision	Recall	F1	ADT(ms)	RI
SVM	0.923	0.918	0.923	0.92	8.5	0.915
RF	0.941	0.939	0.941	0.94	6.2	0.933
FNN	0.945	0.943	0.945	0.944	10.7	0.938
DBN	0.958	0.956	0.958	0.957	13.5	0.951
LSTM	0.956	0.954	0.956	0.955	12.3	0.948
CNN	0.961	0.959	0.961	0.96	11.8	0.955
ELM	0.932	0.929	0.932	0.93	3.8	0.925
MKELM	0.962	0.960	0.962	0.961	4.2	0.957
BO-MKELM	0.978	0.977	0.978	0.977	4.5	0.973

As can be seen from Table 2, the BO-MKELM model outperforms other methods in all evaluation metrics, including the deep learning methods such as DBN, LSTM and CNN analysed in Section 2.1. In particular, BO-MKELM achieves 0.978 and 0.977 on Accuracy and F1, respectively, which is about 1.7 % higher than the best-performing comparison method, CNN. This indicates that the BO-MKELM model can identify various relay protection anomalies more accurately.

In terms of ADT, the BO-MKELM model is slightly slower than ELM, but still remains within 5ms, much faster than deep learning methods such as DBN, LSTM and CNN, which fully meets the real-time requirements of the relay protection system and ensures high diagnostic accuracy. In terms of RI, the BO-MKELM model reaches 0.973, which is higher than all the compared methods, indicating that the model performs particularly well in dealing with important and critical anomalies.

To further analyse the performance of the BO-MKELM model, we plotted the confusion matrix as shown in Figure 2.

From the confusion matrix, it can be seen that the BO-MKELM model has a high recognition accuracy for most types of anomalies. Especially for some serious anomalies, such as CPU faults, program crashes and protection malfunctions, the model's recognition accuracy is close to 100%. This proves that the BO-MKELM model has excellent performance in dealing with the complex and changing grid environment.

To verify the performance of the BO-MKELM model in the presence of data scarcity, we also conducted experiments on small-scale datasets. Table 3 demonstrates the performance comparison of different methods on small-scale datasets.

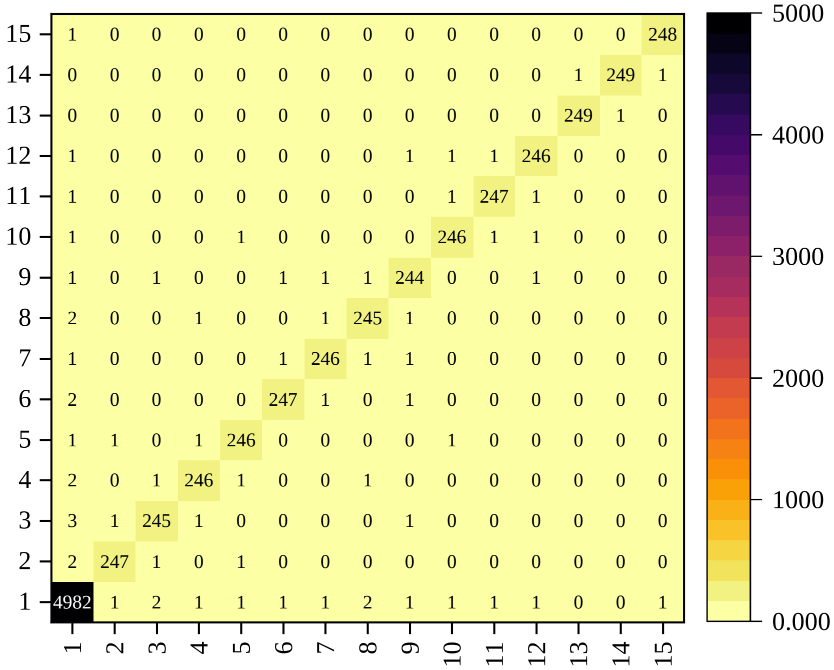

Figure 2. Confusion matrix for the BO-MKELM model on the full dataset

Table 3. Performance Comparison of Different Methods on Small Datasets

Method	Accuracy	Precision	Recall	F1	ADT(ms)	RI
SVM	0.862	0.858	0.862	0.860	7.8	0.855
RF	0.885	0.882	0.885	0.883	5.9	0.878
FNN	0.893	0.890	0.893	0.891	10.1	0.886
DBN	0.907	0.904	0.907	0.905	12.8	0.900
LSTM	0.903	0.900	0.903	0.901	11.5	0.896
CNN	0.912	0.909	0.912	0.910	11.2	0.906
ELM	0.891	0.888	0.891	0.889	3.6	0.884
MKELM	0.928	0.925	0.928	0.926	4.0	0.922
BO-MKELM	0.949	0.947	0.949	0.948	4.3	0.944

On the small dataset, the performance of all methods decreased, but the BO-MKELM model showed the least decrease. Especially on Accuracy and F1, BO-MKELM still maintains a high level of 0.949 and 0.948, which is at least 3% higher than the other methods. This proves the superiority of the BO-MKELM model in dealing with the data scarcity problem.

It is worth noting that deep learning methods such as DBN, LSTM and CNN show a more pronounced performance degradation on small-scale datasets, while the BO-MKELM model maintains a better stability. This highlights the advantage of the BO-MKELM model in coping with the data scarcity problem, which echoes the innovation proposed in this paper.

In summary, the experimental results fully validate the superiority of the BO-MKELM model in the relay protection anomaly diagnosis task. The model not only performs well on the complete dataset, outperforming all methods including those analysed in

Section 1.1, but also maintains a high diagnostic accuracy despite data scarcity. These results demonstrate the advantages of the BO-MKELM model in dealing with complex and changing grid environments, coping with the problem of data scarcity, and maintaining real-time diagnosis, which fully reflects the innovative points of this paper.

In particular, the BO-MKELM model significantly reduces the diagnosis time while maintaining high accuracy compared to deep learning methods such as DBN, LSTM, and CNN, which is crucial for the real-time response of relay protection systems. Meanwhile, the excellent performance of the BO-MKELM model on small-scale datasets solves the problem of high data demand in existing research, and provides an effective solution to the situation of data scarcity in practical applications.

7. **Conclusions.** In this study, a multicore limit learning machine (BO-MKELM) model based on Bayesian optimisation is proposed for solving the key problems in anomaly diagnosis of relay protection systems. By integrating multicore learning, limit learning machine and Bayesian optimisation techniques, the model effectively overcomes the limitations of traditional methods in processing complex power system data and meeting real-time diagnosis requirements. The experiments based on the actual operation data of a 500kV intelligent substation and the data generated by the RTDS simulation system lead to the following conclusions:

(1) The BO-MKELM model excels in handling high-dimensional, nonlinear and time-varying power system operational data, and is able to effectively extract key features reflecting the system state.

(2) Compared with existing deep learning methods, the BO-MKELM model significantly reduces the computation time while maintaining a high diagnostic accuracy, which better meets the real-time requirements of relay protection systems.

(3) In the case of data scarcity, the BO-MKELM model shows excellent generalisation ability and stability, providing an effective solution to the problem of insufficient data in practical engineering.

(4) The application of Bayesian optimisation in parameter tuning of multi-core extreme learning machines effectively improves the overall performance of the model and achieves a good balance between accuracy and computational efficiency.

(5) Compared with the single kernel function approach, the multiple kernel learning strategy significantly improves the model's ability to recognise different types of anomalies and enhances the model's adaptability and robustness.

The experimental data in this study are mainly from a single substation and simulation system, which may limit the generalisation capability of the model. Future work should consider introducing more power system data from different voltage levels, geographic locations and operating environments to verify the effectiveness of the BO-MKELM model in a wider range of application scenarios. In addition, further improving the interpretability of the model as well as exploring the extension of this method to other power system intelligence application areas are also interesting research directions.

Acknowledgment. This work is supported by the Science and Technology Project of Guangxi Power Grid Corporation (No. GXKJXM20230110): Research and engineering application of panoramic display and decision analysis technology for power grid fault relay protection based on artificial intelligence.

REFERENCES

[1] Z. Bo, X. Lin, Q. Wang, Y. Yi, and F. Zhou, "Developments of power system protection and control," *Protection and Control of Modern Power Systems*, vol. 1, no. 1, pp. 1-8, 2016.

[2] H. C. Kiliçkiran, İ. Şengör, H. Akdemir, B. Kekezoğlu, O. Erdinç, and N. G. Paterakis, "Power system protection with digital overcurrent relays: A review of non-standard characteristics," *Electric Power Systems Research*, vol. 164, pp. 89-102, 2018.

[3] B. Zhang, Z. Hao, and Z. Bo, "New development in relay protection for smart grid," *Protection and Control of Modern Power Systems*, vol. 1, pp. 1-7, 2016.

[4] C. So, and K. Li, "Time coordination method for power system protection by evolutionary algorithm," *IEEE Transactions on Industry Applications*, vol. 36, no. 5, pp. 1235-1240, 2000.

[5] K. Bae, and J. S. Thorp, "A stochastic study of hidden failures in power system protection," *Decision Support Systems*, vol. 24, no. 3-4, pp. 259-268, 1999.

[6] Z. Jiao, H. Gong, and Y. Wang, "A DS evidence theory-based relay protection system hidden failures detection method in smart grid," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 2118-2126, 2016.

[7] V. F. Pires, L. S. Martins, T. G. Amaral, R. Marçal, R. Rodrigues, and M. M. Crisóstomo, "Distance-learning power-system protection based on testing protective relays," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 6, pp. 2433-2438, 2008.

[8] M. Andreev, Y. Bay, and B. Malyuta, "Novel method for setting up the relay protection of power systems containing renewable energy sources and hydrogen energy storage systems," *International Journal of Hydrogen Energy*, vol. 48, no. 76, pp. 29425-29445, 2023.

[9] S. Tamronglak, S. Horowitz, A. Phadke, and J. Thorp, "Anatomy of power system blackouts: preventive relaying strategies," *IEEE Transactions on Power Delivery*, vol. 11, no. 2, pp. 708-715, 1996.

[10] W. Guo, f. Wen, G. Ledwich, Z. Liao, X. He, and J. Liang, "An analytic model for fault diagnosis in power systems considering malfunctions of protective relays and circuit breakers," *IEEE Transactions on Power Delivery*, vol. 25, no. 3, pp. 1393-1401, 2010.

[11] A. Rahmati, and R. Adhami, "A fault detection and classification technique based on sequential components," *IEEE Transactions on Industry Applications*, vol. 50, no. 6, pp. 4202-4209, 2014.

[12] R. Dashti, and J. Sadeh, "Accuracy improvement of impedance-based fault location method for power distribution network using distributed-parameter line model," *International Transactions on Electrical Energy Systems*, vol. 24, no. 3, pp. 318-334, 2014.

[13] B. K. Panigrahi, P. K. Ray, P. K. Rout, A. Mohanty, and K. Pal, "Detection and classification of faults in a microgrid using wavelet neural network," *Journal of Information and Optimization Sciences*, vol. 39, no. 1, pp. 327-335, 2018.

[14] O. Gashteroodkhani, M. Majidi, and M. Etezadi-Amoli, "A combined deep belief network and time-time transform based intelligent protection Scheme for microgrids," *Electric Power Systems Research*, vol. 182, pp. 106239, 2020.

[15] W.-H. Kim, J.-Y. Kim, W.-K. Chae, G. Kim, and C.-K. Lee, "LSTM-based fault direction estimation and protection coordination for networked distribution system," *IEEE Access*, vol. 10, pp. 40348-40357, 2022.

[16] N. Qu, Z. Li, J. Zuo, and J. Chen, "Fault detection on insulated overhead conductors based on DWT-LSTM and partial discharge," *IEEE Access*, vol. 8, pp. 87060-87070, 2020.

[17] P. Rai, N. D. Londhe, and R. Raj, "Fault classification in power system distribution network integrated with distributed generators using CNN," *Electric Power Systems Research*, vol. 192, pp. 106914, 2021.

[18] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19-46, 2024.

[19] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10. 2339, 2023.

[20] T.-Y. Wu, H. Li, and S.-C. Chu, "CPPE: An Improved Phasmatodea Population Evolution Algorithm with Chaotic Maps," *Mathematics*, vol. 11, no. 9. 1977, 2023.

[21] Y. Huérfano-Maldonado, M. Mora, K. Vilches, R. Hernández-García, R. Gutiérrez, and M. Vera, "A comprehensive review of extreme learning machine on medical imaging," *Neurocomputing*, pp. 126618, 2023.

[22] J. Zhang, Y. Li, W. Xiao, and Z. Zhang, "Non-iterative and fast deep learning: Multilayer extreme learning machines," *Journal of the Franklin Institute*, vol. 357, no. 13, pp. 8925-8955, 2020.

[23] X. Wang, Y. Jin, S. Schmitt, and M. Olhofer, "Recent advances in Bayesian optimization," *ACM Computing Surveys*, vol. 55, no. 13s, pp. 1-36, 2023.

[24] R. Roussel, A. L. Edelen, T. Boltz, D. Kennedy, Z. Zhang, F. Ji, X. Huang, D. Ratner, A. S. Garcia, and C. Xu, "Bayesian optimization algorithms for accelerator physics," *Physical Review Accelerators and Beams*, vol. 27, no. 8, pp. 084801, 2024.

[25] M. Milton, C. De La O, H. L. Ginn, and A. Benigni, "Controller-embeddable probabilistic real-time digital twins for power electronic converter diagnostics," *IEEE Transactions on Power Electronics*, vol. 35, no. 9, pp. 9850-9864, 2020.

[26] M. Glatt, C. Sinnwell, L. Yi, S. Donohoe, B. Ravani, and J. C. Aurich, "Modeling and implementation of a digital twin of material flows based on physics simulation," *Journal of Manufacturing Systems*, vol. 58, pp. 231-245, 2021.