IWOA-LightGBM: Hyperparameter Optimization for Sensor Data Anomaly Detection*

Rong Huang[1], Qiqiang Wu[12], Mingwei Yang[12], Yanhua Liu[12], and Baokang Zhao[3†]

[1] Fuzhou University, Fuzhou, China
{front, 241027059, lyhwa}@fzu.edu.cn, 1412617067@qq.com
[2] Engineering Research Center of Big Data Intelligence, Ministry of Education, Fuzhou, China
[3] National University of Defense Technology, Changsha, China
bkzhao@nudt.edu.cn

Abstract

Anomaly detection performance in sensor data is highly sensitive to model hyperparameters, which is central to reliable monitoring in mobile Internet security and industrial IoT (IIoT) scenarios. We propose an IWOA-LightGBM based anomaly detection method for sensor data. For machine learning-based anomaly detection methods, hyperparameter selection often determines model performance, so we propose an Improved Whale Optimization Algorithm (IWOA) and further use it to optimize the hyperparameters of the LightGBM algorithm. To avoid falling into local optima and accelerate algorithm convergence, the WOA is improved by integrating nonlinear convergence factor, adaptive inertia weight factor and stochastic differential mutation strategy. Experimental results show that during hyperparameter optimization for LightGBM model training, the IWOA achieves faster convergence and higher computational efficiency compared to the Whale Optimization Algorithm (WOA), with anomaly detection accuracy exceeding 90%.

Keywords: Anomaly Detection, Industrial Sensor Data, IWOA, LightGBM

1 Introduction

Modern industrial systems contain numerous interconnected devices and sensors that generate vast amounts of sensor data. Monitoring this data is essential to ensure facility safety and stable production operations. At the same time, as industrial systems expand in scale, industrial processes have become increasingly complex, with sensor data exhibiting characteristics such as high dimensionality, temporal continuity, and parameter correlations [1]. In mobile Internet and IIoT deployments, sensor telemetry is transported over wireless links and processed on resource-constrained edge nodes. Ensuring the integrity of these streams is a core security requirement. Consequently, how to extract meaningful information from complex high-dimensional sensor data and detect anomalies that reflect abnormalities in sensor-equipped physical entities presents a challenging problem [2]. Currently, researchers have proposed various anomaly detection methods for industrial sensor data.

Due to the continuity characteristics of sensor data, anomaly detection methods based on time series features are usually divided into reconstruction-based and prediction-based ones [3][4]. Prediction-based methods mainly discriminate anomalies based on the prediction error

*Proceedings of the 9th International Conference on Mobile Internet Security (MobiSec'25), Article No. 18, December 16-18, 2025, Sapporo, Japan.

†Corresponding Author

of the algorithm [5][6][7]. They characterize sequence data as node information and predict the value of the sequence data at the next moment through the combination of graph structure learning and graph neural network. However, the sequence association in the paper is limited by the cosine similarity of the original sequence [5]. The reconstruction-based anomaly detection algorithm learns the feature representation of the data by reconstructing the input data and discriminates the anomalies based on the error value between the reconstructed data and the original data[8][9][10]. By combining Transformer and GANs [11], where GANs are used for original data reconstruction while Transformer extracts the contextual features of time series data, experiments illustrate that the detection performance can be improved. However, the anomaly detection performance in the paper is very sensitive to the sliding window length.

In practical industrial settings, datasets often contain labeled data. Effectively incorporating label information into anomaly detection often improves the accuracy of the algorithm [12]. Moreover, in unsupervised anomaly detection algorithms, such as iForest, OCSVM and LSTM, the model is affected by noise and anomalies, and the accuracy and robustness are weaker than those of other algorithms [13]. Similarly, in the field of intrusion detection in industrial control systems, the learning efficiency of supervised learning is better than of unsupervised learning algorithms among traditional machine learning algorithms such as LightGBM (Light Gradient Boosting Machine), SVM and HMM [14].

While numerous scholars have conducted extensive research on anomaly detection for industrial data from diverse perspectives and achieved significant results, challenges persist in hyperparameter selection processes. Due to algorithm sensitivity to parameters such as learning rates and time windows, random selection fails to yield optimal analytical outcomes or support accurate decision-making. Therefore, further research on the hyperparameter optimization method of the model is needed.

To address the above problems, we propose a LightGBM-based anomaly detection method for sensor data using an Improved Whale Optimization Algorithm (IWOA). In the data processing stage, statistical features (e.g., mean, maximum, and minimum values) within time windows are calculated as temporal contextual features, while Kaiser Window smoothing is applied for data preprocessing. For hyperparameter optimization, in order to speed up the training of the model, the optimization algorithm is improved by combining the nonlinear convergence factor, adaptive inertia weight factor and stochastic differential mutation strategy to reduce the time spent on manual hyperparameter selection. This work focuses on hyperparameter optimization of LightGBM via IWOA, rather than proposing a new anomaly detector. The aim is practical: accelerate convergence and stabilize accuracy for mobile/IIoT sensor integrity checking. The main contributions of this work are as follows:

1. An anomaly detection model based on IWOA-LightGBM is proposed, which leverages the IWOA to optimize the hyperparameters of LightGBM, thereby enhancing detection accuracy and efficiency.

2. The IWOA is developed by integrating nonlinear convergence factor, adaptive inertia weight strategy, and stochastic differential mutation strategy. These improvements aim to strengthen the global search capability, accelerate convergence, and prevent premature trapping in local optima.

3. Hyperparameters of the LightGBM algorithm are optimized using the IWOA, resulting in faster convergence during model training and a balanced computational speed and detection performance. Experimental results demonstrate that IWOA outperforms the original WOA in terms of convergence speed and robustness for hyperparameter tuning.

2 Background

2.1 Whale Optimization Algorithm

The Whale Optimization Algorithm (WOA) is an intelligent population optimization algorithm based on natural animal behavior, proposed by Professor Mirjalili in 2016 [15]. Based on the group behavior of whales during feeding and migration, the WOA searches for the optimal solution in the parameter space by simulating the searching and chasing behavior of whale groups. The specific process is shown below:

- Encircling prey: In the d-dimensional feature space, the position of the whale is a solution in the parameter space. Assume the location of individual whale is $X^i = (X_1^i, X_2^i, \ldots, X_d^i)$ and the current location of the best individual whale is $X^* = (X_1^*, X_2^*, \ldots, X_d^*)$. Individual whales move toward the optimal individual, and during this process, their position at the next moment is updated using the following equation:

$$X_k^{i+1} = X_k^* - A_1 \cdot D_k \tag{1}$$

$$D_k = |C_1 \cdot X_k^* - X_k^i| \tag{2}$$

$$C_1 = 2 \cdot r_2 \tag{3}$$

$$A_1 = 2a \cdot r_1 - a, a = 2 - \frac{2t}{T_{max}} \tag{4}$$

  a is a convergence factor that decreases linearly from 2 to 0. t denotes the current iteration number, and T_{max} denotes the maximum number of iterations.r_1, r_2 are random numbers within $[0, 1]$. A_1 and C_1 are coefficient factors that guide the movement direction and distance.

- Bubble-net attacking: Bubble-net attacking includes contraction encirclement and spiral updating, where the behavior of contraction encirclement is similar to that in encircling prey, but the range of A_1 is adjusted from $[-a, a]$ to $[-1, 1]$. Spiral positional updating mainly simulates individual whales approaching the current best whale individual in a spiral fashion. The updated formula is shown below:

$$X_k^{i+1} = X_k^* + D_k \cdot e^{bl} \cdot \cos 2\pi l \tag{5}$$

$$D_k = |X_k^* - X_k^i| \tag{6}$$

  b is the spiral shape constant and l is a random number in the range $[-1, 1]$. At the same time, whales need to swim toward their prey in a spiral form while shrinking their encirclement; thus, each behavior has a 50% probability, and the individual position update formula is as follows:

$$X_k^{i+1} = \begin{cases} X_k^* - A_1 \cdot D_k & p < 0.5 \\ X_k^* + D_k \cdot e^{bl} \cdot \cos 2\pi l & p \geq 0.5 \end{cases} \tag{7}$$

  p is a random number uniformly sampled in $[0, 1]$.

- Search for prey: In the process of searching for prey, the whale individual is not in the contraction envelope when the coefficient $A \geq 1$. At this point, the whale seeks other prey, and thus does not move toward the current optimal individual, it moves randomly in the direction of one of the whale partners in the group, and with the potential to find a better solution than the current individual, thereby endowing the group with a certain global search capability. The search for predation process is similar to Eq.1, but the update target is a random individual whale $X^T = (X_1^T, X_2^T, \ldots, X_d^T)$.

$$X_k^{i+1} = X_k^T - A_1 \cdot D_k \tag{8}$$

2.2 Light Gradient Boosting Machine

LightGBM is a framework proposed to implement the Gradient Boosting Decision Tree (GBDT) algorithm, designed to avoid memory constraints when training data [16]. When applied, LightGBM handles large-scale and high-dimensional data better than other machine learning methods through Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB).

GOSS is a down-sampling technique, and its core idea is to prioritize training samples with large gradients, which typically contribute more to information gain, and to down-sample those with small gradients in a controlled manner.

Specifically, given a dataset of size N, the absolute gradient values of all training instances are computed and sorted in descending order. A proportion a of the instances with the largest gradients is retained entirely. From the remaining instances, a random subset is selected with a sampling ratio b. To maintain the overall data distribution and unbiased estimation of information gain, a compensation weight is applied to each selected instance from the small-gradient subset. The weight is computed as:

$$w = \frac{1-a}{b} \tag{9}$$

$a \in (0, 1)$ denotes the retention ratio of large-gradient samples,$b \in (0, 1)$ denotes the sampling ratio for small-gradient samples, and w is the amplification coefficient assigned to each sampled small-gradient instance. This approach ensures that important data points are preserved for accurate split gain calculation, while redundant information from low-gradient instances is efficiently reduced, achieving both training speed and generalization performance.

The core idea of EFB is to reduce the total number of features by bundling mutually exclusive features together. In high-dimensional data, features are often mutually exclusive, i.e., it is rare for certain features to appear at the same time. Taking advantage of this property, EFB bundles these mutually exclusive features together to form a new feature, and uses the bundled new feature as an input during model training, thus reducing the dimensionality of the features. In addition, LightGBM allows multi-threaded parallel execution, which greatly improves model efficiency. Due to its high performance and efficiency, LightGBM has been applied to many IoT data analytics tasks, such as cyber-attack and malware detection in IoT systems [17][18]. The hyperparameters of machine learning methods directly affect the training speed and prediction accuracy of the model[19]. In the LightGBM algorithm, there are a wide variety of hyperparameters, constituting a large parameter space. This often face the dilemma of excessive time consumption and inefficiency in practical applications when relying only on manual selection or random picking alone. Therefore, intelligent optimization algorithms are needed to solve the optimal hyperparameters of the model.

3 Our Method

We propose an anomaly detection model based on IWOA-LightGBM, which uses IWOA to optimize the hyperparameter selection problem of LightGBM. The structure of our anomaly detection framework is shown in Fig.1.

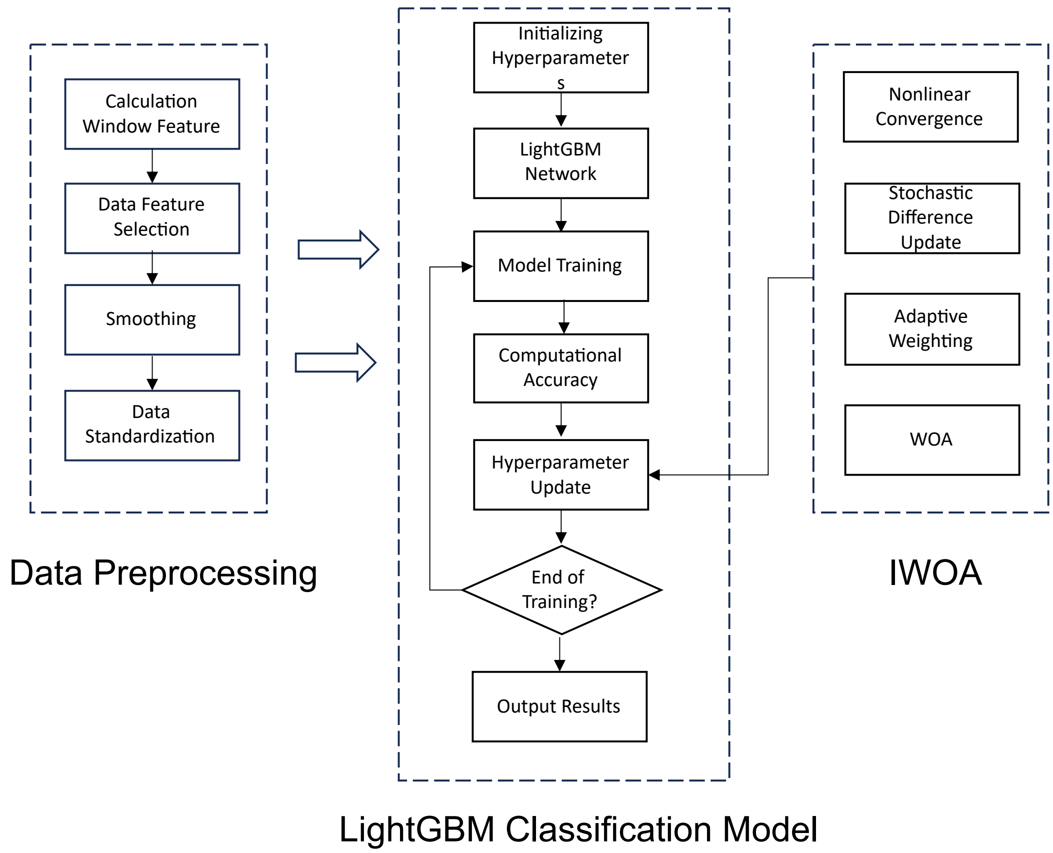

Figure 1: Anomaly Detection Framework

The proposed anomaly framework contains three parts: data preprocessing, the LightGBM classification model and the IWOA. In data preprocessing, this paper mainly focuses on the processing of residuals and temporal context information in the sensing data. After data processing, the LightGBM model is trained to distinguish abnormal samples from normal samples. The IWOA is used to optimize the hyperparameters of the LightGBM model. In this paper, the WOA is improved by the nonlinear convergence factor, the adaptive inertia weight factor and the stochastic differential mutation strategy, and the corresponding fitness function is employed to continuously optimize the model's hyperparameters until the termination conditions are satisfied.

3.1 Data Preprocessing

The data preprocessing stage mainly includes feature selection, smoothing and data normalization. [20] In the feature processing stage, given the temporal correlation in sensor data samples, a time window parameter is set. For each data sample, a window is formed with the current sample and a specified number of previous samples, and statistics such as the window mean and extreme values are calculated as contextually relevant features. To avoid random selection from destroying the data's contextual information, feature extraction and fusion are performed on data with known anomaly categories, which are then divided into the training set, validation set, and test set in a 7:2:1 ratio.

Residual error exists in industrial sensing data, and if this error is large, it is difficult to capture meaningful patterns from raw time series data. Therefore, the data is smoothed using the Kaiser window function with a window length of 11 and a β parameter of 2 to eliminate residual errors. [21]

Finally, to ensure the processed data conforms to a standard normal distribution, the data is normalized, and the normalization process is defined as:

$$X^* = \frac{X - \mu}{\sigma} \tag{10}$$

μ is the mean of all samples and σ is the standard deviation.

3.2 Improved Whale Optimization Algorithm

3.2.1 Nonlinear Convergence Factor

In the iterative process of the original WOA algorithm, the update speed of the individual whale position is often affected by the convergence factor a, and the unreasonable convergence factor may cause the algorithm to prematurely fall into a local optimal solution, leading to slow convergence in the late iteration stage[22]. In the original whale optimization algorithm, the linear decrease of the convergence factor fails to ensure effective global search in the early stage and accurate local optimization in the late stage, which may result in slow algorithm convergence. Therefore, we adopt a nonlinear convergence strategy, as shown in Eq.11:

$$a = a_0 \cdot e^{-(\tan(1.2 \times \frac{t}{T_{max}})^2)} \tag{11}$$

a_0 denotes the initial convergence factor, t denotes the current iteration number, and T_{max} is the maximum number of iterations. This nonlinear function allows the convergence factor a to decrease slowly in the early stage of the search process, thereby enhancing the global exploration capability. As the number of iterations increases, the factor decreases rapidly and approaches zero, which promotes local exploitation in the later stage. This ensures algorithm's convergence speed while enhancing its search accuracy and global search capability.

3.2.2 Adaptive Inertia Weight Factor

In the optimization process of intelligent algorithms, it is necessary to balance the algorithm's global search ability and local search ability to avoid falling into local optimal solutions while improving the algorithm's efficiency [23]. To enhance the algorithm's global search capability and improve its convergence performance, we propose an adaptive inertia weight factor. When updating its position, an individual calculates the corresponding inertia factor ω , and the calculation formulas are shown in Eq.12:

$$\omega = \begin{cases} rand(0.8, 1.2) & f(i) < f_{avg1} \\ 1 & f_{avg1} \leq f(i) < f_{avg2} \\ rand(0.6, 0.8) \ or \ rand(1.2, 1.4) & f(i) \geq f_{avg2} \end{cases} \tag{12}$$

Specifically, in the iterative process, individual whales are first sorted in ascending order by their fitness, then divided into two parts. The average fitness of these two parts is calculated as f_{avg1} and f_{avg2} respectively, where $f_{avg1} < f_{avg2}$. By comparing the fitness of the current whale individuals $f(i)$ with f_{avg1} and f_{avg2}, the corresponding inertia factor classification and value for the individual can be determined.

After adopting the whale optimization algorithm with an adaptive inertia weight factor, the process of individual roundup prey and spiral updating is shown in Eq.13:

$$X_k^{i+1} = \begin{cases} \omega \cdot X_k^* - A_1 \cdot D_k & p < 0.5 \\ \omega \cdot X_k^* + D_k \cdot e^{bl} \cdot \cos 2\pi l & p \geq 0.5 \end{cases} \tag{13}$$

3.2.3 Stochastic Differential Mutation Strategy

Since the traditional WOA can only rely on random individuals in the population to guide updates during global search, its global search ability is limited [24]. Further, to enhance the algorithm's global search capability in finding the optimal solution, we employ the stochastic differential mutation to optimize the position update strategy for whale individuals. Specifically, this is shown in Eq.14:

$$X^{t+1} = r_1 \times (X^* - X^t) + r_2 \times (X'^t - X^t) \tag{14}$$

r_1 and r_2 are random numbers in $[0, 1]$, and X' is a random individual.

3.2.4 Algorithmic Process

During the encircling predation or spiral updating process of the IWOA, the algorithm calculates the inertia factor using the adaptive factor strategy, and then updates the individual position. Further, during search predation, individuals combine random and optimal individuals in the population to select a new moving target via the stochastic differential mutation strategy. This strategy not only accelerates population convergence, enabling the algorithm to find the optimal solution faster, but also effectively prevents the population from falling into local optimization and enhances the algorithm's global search capability. The specific workflow of our IWOA is shown in Fig.2, which can be divided into the following main steps:

Initialization: Set the parameters of the algorithm, such as the initial population, number of iterations and parameter bounds.

Update individual position: Update the position of the selected whale individual. The update rule involves two behaviors: encircling prey and bubble-net attacking. As shown in Fig.2, when $p \geq 0.5$, spiral updating is performed as in Eq.13; when $p < 0.5$ further selection is made by the value of A. When $|A| < 1$, prey rounding-up is performed, and the individual's position is updated according to Eq.13 when $p < 0.5$. When $|A| \geq 1$, a random search is performed, where individuals select a random individual X^T and update their position. This is shown in Eq.15:

$$X_k^{i+1} = \omega \cdot X_k^T - A_1 \cdot D_k \tag{15}$$

Finally, the individual position is updated according to the stochastic differential mutation strategy, following Eq.14.

Calculate the degree of adaptation: To distinguish the goodness of individual whales, their fitness needs to be defined. To optimize the classification prediction network, classification accuracy is used as the basis for the fitness function; specifically, we select the average of the training set accuracy and validation set accuracy as the fitness function. A larger fitness function

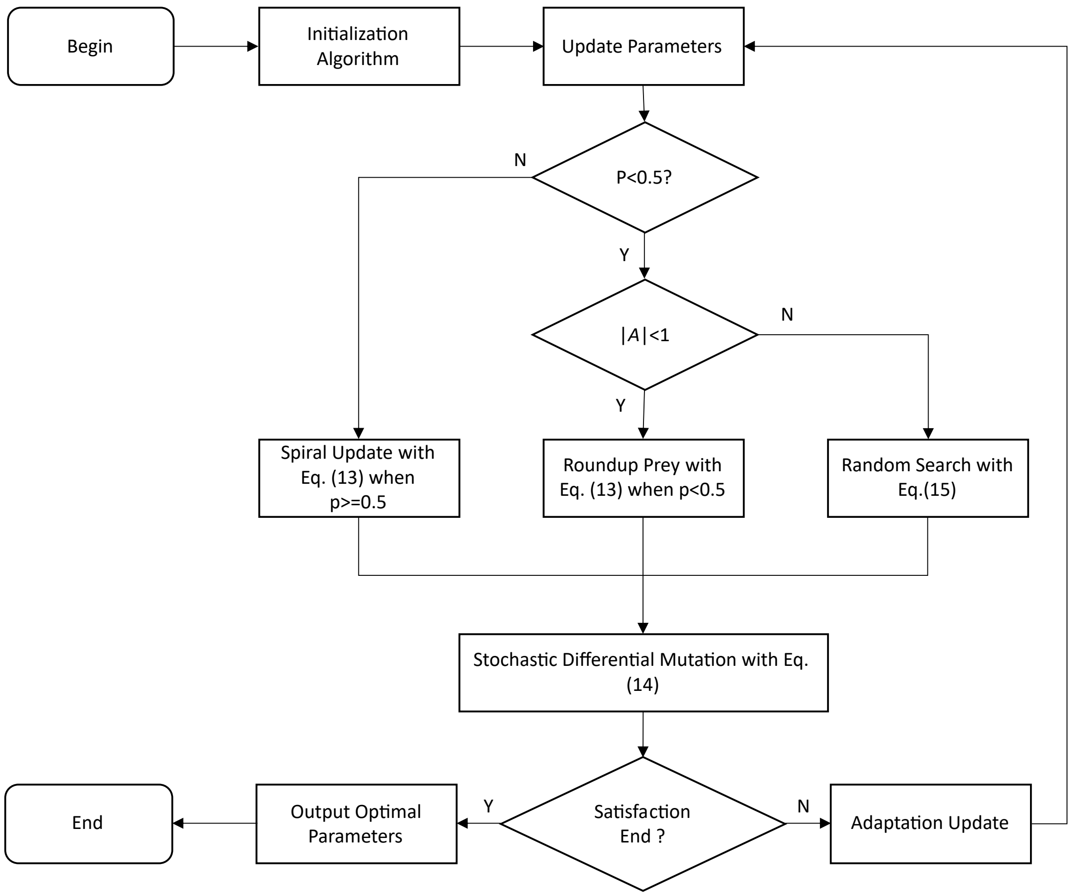

Figure 2: Flowchart of IWOA

value indicates higher model classification accuracy and better anomaly detection performance, as shown in Eq.16:

$$Fitness = \frac{Acc_{train} + Acc_{valid}}{2} \tag{16}$$

Individuals are sorted by their calculated fitness values in ascending order. Adaptive weights of individuals are updated according to Eq.12, and the individual with the largest fitness value is considered the optimal individual.

Finally, determine whether the termination conditions are reached, such as the number of iterations reaches the preset upper limit, or find the optimal solution that meets the preset accuracy requirements. If the termination condition is reached, output the optimal solution; otherwise, return to continue iteration.

3.2.5 Distinction from Other WOA Improvements

IWOA exhibits distinct characteristics compared to other recent WOA improvements, primarily in the design of key optimization mechanisms:

1. The IWOA incorporates a nonlinear convergence factor (Eq.11) that adjusts dynamically based on iteration progress. This factor decreases slowly in early iterations to enhance global exploration and rapidly in later stages to focus on local exploitation. This differs from the Lévy flight mutation-based improvement in [25], which emphasizes random jumps for diversity but does not adjust convergence dynamics in a nonlinear manner.

2. Our algorithm introduces an adaptive inertia weight factor (Eq.12) that varies with individual fitness, enabling differentiated search behavior for suboptimal and optimal solutions. This adaptive strategy is distinct from the hybrid genetic mechanism in [26], which integrates genetic algorithm operators (crossover, mutation) for optimization.

3. The IWOA employs a stochastic differential mutation strategy (Eq.14) that combines random individuals and the current optimal solution to introduce controlled perturbations. This differs from the bat algorithm-local search integration in [27], which uses random interference on local optima, and the chaos theory-based mutation in [28], which relies on chaotic maps for diversity. Our mutation strategy directly links perturbations to population dynamics, enhancing relevance to the current search state.

These mechanisms collectively define the unique approach of IWOA, focusing on dynamic balance between exploration and targeted perturbation, distinct from the improvement paths of other works.

3.3 LightGBM Hyperparameter Optimization

In this section, the hyperparameters of the LightGBM model are optimized using the IWOA to solve for the optimal parameters. In the LightGBM model, there are many hyperparameters that affect the training of the model, and the hyperparameters that need to be adjusted include Learning Rate, Max Depth of the tree, and Min Data in Leaf. The selection of these hyperparameters is crucial to model performance. For example, the Learning Rate controls the step size during model training, affecting the convergence speed of the model and the stability of the algorithm; the Max Depth of the decision tree, where a tree with too large a depth may lead to overfitting, while a tree with too small a depth may not be able to learn the data adequately; and Min Data in Leaf, which is the minimum number of samples required in the leaf nodes. Smaller values may lead to a more complex model, but may also increase the risk of overfitting the model; Lambda (L1 regularization) is a parameter that controls model complexity and avoids overfitting by penalizing such complexity. The purpose of the IWOA is to optimize these hyper-parameters to help the model fit the data better and to improve the performance of the model.

The overall anomaly detection process is shown in Algorithm 1, where the IWOA is used to select hyperparameters to optimize the training of the LightGBM network during the LightGBM training process. The hyperparameters optimized by the algorithm improve the model's predictive accuracy and shorten its training time. Specifically, data preprocessing is needed first, including calculation of window features, smoothing processing and data normalization. Then, the IWOA and LightGBM method need to be initialized and enter the iterative process. In the iterative process, the corresponding population solutions and datasets are fed into the

LightGBM classification framework. Classification results of the samples are obtained to compute fitness and select the optimal individuals, and individual positions are updated based on adaptive weights and differential variances. Finally, the LightGBM classification model with optimal hyperparameters is output.

Algorithm 1 Anomaly detection algorithm based on IWOA-LightGBM

Input: Dataset D.
Output: The trained Model

Processing D
Initialize the population, Number of Iterations max_iter, Boundaries of the solution space
for i in range(max_iter) **do**
    if the individual is out of bounds **then** change the out of bounds value to upper and lower bounds
    end if
    Train the LightGBM model with the corresponding parameters
    use Eq.(11) to update the convergence factor
    use Eq.(14) to update individual position
    use Eq.(12) to update adaptive weights for individuals:
    use Eq.(13) to choose encircling prey or spiraling update
end for
Output optimal model

4 Experiments and Results

4.1 Dataset

The Skoltech Anomaly Benchmark (SKAB) dataset is a benchmark test set designed to evaluate the performance of anomaly detection algorithms [29]. The benchmark currently contains over 30 different datasets and associated Python modules for algorithm evaluation. Each dataset contains multivariate time-series data collected by sensors installed on the testbed, validated for algorithm performance across various scenarios, and labeled for evaluating outlier and change-point detection problems. The valve1_data is used as the experimental data, specifically a dataset of abnormal waveforms generated when the valve is closed. The valve1_data set consists of 18,162 data points, including 11,853 normal data points and 6,309 abnormal data points. All instances in the dataset are labeled with supervisory data for each time point (normal: 0, abnormal: 1), which are used to evaluate and verify the accuracy of anomaly detection.

The data are 8-dimensional time series data obtained from sensors characterized by parameters: acceleration_1RMS, acceleration_2RMS, current, pressure, temperature, thermocouple, voltage and volume flowrate. Fig.3 [29] illustrates the structural layout of the water circulation system used in the SKAB dataset. The system simulates an industrial environment consisting of multiple interconnected components: Water is drawn from a storage tank (3) via a pump (4) and circulated through pipes controlled by solenoid valves (1,2). A motor (6) controlled by a frequency converter (7) drives the circulation process. Various sensors are installed throughout the system to measure vibration, electrical parameters, fluid dynamics, and temperature. Control and data collection are managed via a Compact RIO module (8), while safety mechanisms–such as the emergency stop (5) and the shaft deviation level (9)–ensure operational reliability.

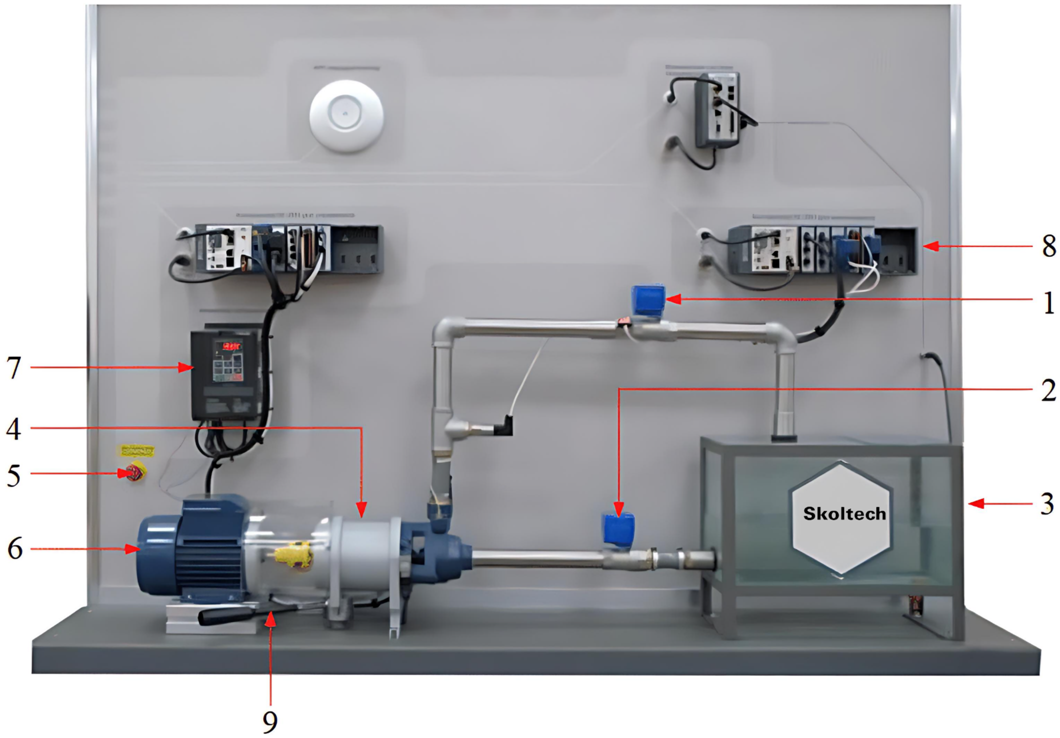

Figure 3: The Water Circulation, Control and Monitoring System

4.2 Evolution Curve

In this section, the superiority of IWOA is verified by comparing the evolutionary curves of IWOA and WOA. In the anomaly detection model, the initial experimental settings are a population size of 30 and a maximum iteration count of 50. The search range for LightGBM hyperparameters [16] is set as follows: learning rate [0.001, 0.2], maximum tree depth [2, 30], minimum data in leaf [3, 2000], and number of leaves [1, 400].

The evolution curves of the experimental results are shown in Fig.4. During the solving process of IWOA and WOA, the fitness value remains unchanged on multiple occasions, leading to local optimal solutions. However, IWOA can escape local optima more quickly, enabling faster algorithm convergence. In Fig.4, the WOA falls into a local optimum around the 20th round iteration, and struggles to find other valid solutions in the subsequent 20 iterations, indicating a need for further optimization. The IWOA algorithm also finds this local optimum at the 15th round, and jumps out of the optimum at around the 25th round. Meanwhile, it can be observed that in the early stage of the optimization process, the fitness function value of IWOA improves more significantly. This indicates that, with the same population size and number of iterations, the algorithm achieves efficient searching and accelerates the optimal solution finding process due to the stochastic differential mutation strategy.

To avoid the influence of randomness on the experimental results. We conducted 15 independent repeated experiments for both IWOA and WOA. The algorithm parameters were

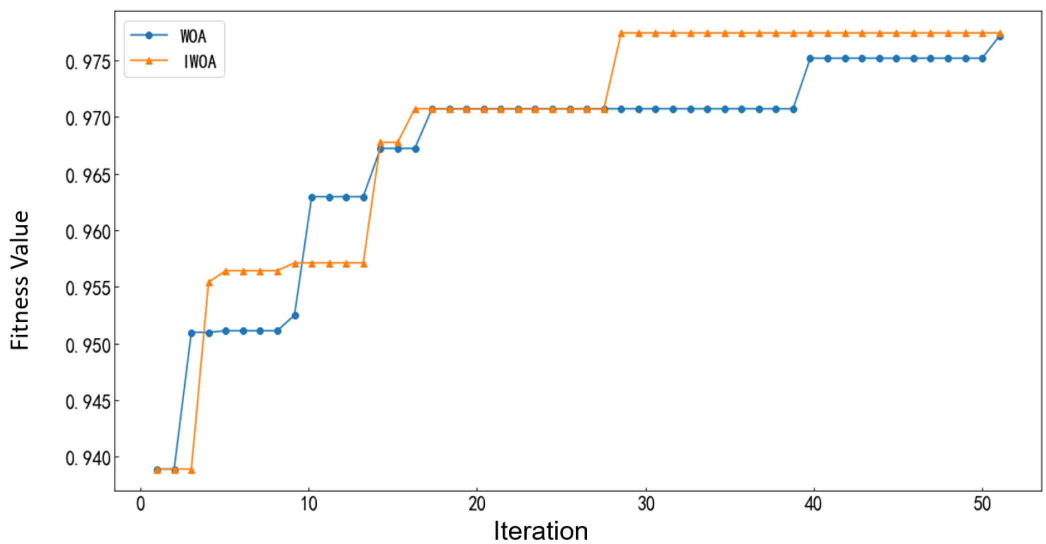

Figure 4: Comparison of Evolutionary Curves for IWOA and WOA

set the same as in the single experiment, with the maximum number of iterations was set to 100, and the mean value of the 15 results was calculated. As shown in Fig.5, IWOA exhibits a more significant improvement in the fitness function value in the early stage, with a faster convergence speed, and also converges rapidly in the later stage of the algorithm.

4.3 Anomaly Detection

In the above experiment, the optimal hyperparameters optimized by the IWOA algorithm are as follows: learning rate = 0.1, minimum data in leaf = 20, maximum depth = 25, and number of leaves = 30. In order to validate the detection performance of the proposed model, we validate it on a test set, with the following results: the test set contains 1,750 samples, and the algorithm achieves a accuracy = 92.17%, precision = 98.67%, F1-score = 90.20% and recall = 83.06%. The sample distribution and algorithm labeling results are shown in Fig.6. Samples in the intervals 250 and 1500 are abnormal, while the rest are normal. The detection algorithm in this paper can cover most actual abnormal intervals, verifying the validity of the anomaly detection model.

To further validate the superiority of the proposed IWOA-LightGBM model, we compared its performance with several baseline models, including traditional machine learning algorithms and deep learning-based methods. All baselines in Table 1 are evaluated under the same dataset split and preprocessing pipeline described in Section 3.1. For XGBoost and CatBoost,

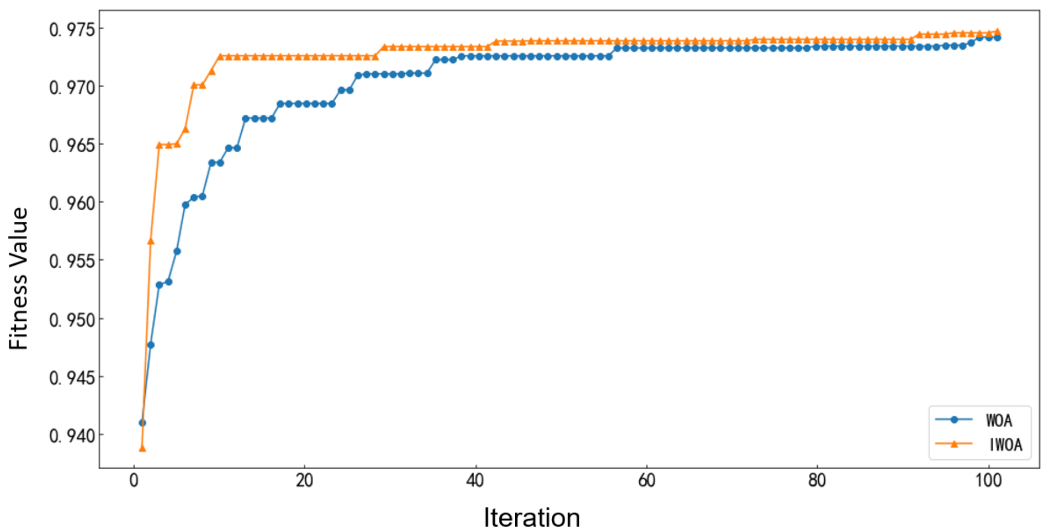

Figure 5: Comparison of Evolutionary Curves after averaging 15 independent experiments

Table 1: Performance Comparison of Anomaly Detection Models

Model	Accuracy	F1-score
IWOA-LightGBM(ours)	0.92	0.90
MEST	0.58	0.78
T-squared	0.79	0.66
T-squared+Q	0.71	0.76
Conv_AE	0.83	0.78
LSTM_AE	0.70	0.74
Vanilla_LSTM	0.85	0.53
XGBoost	0.92	0.90
CatBoost	0.88	0.83

hyperparameters are selected by random search using the identical validation protocol; the best validation configuration is then applied to the test set. The evaluation metrics include accuracy and the F1-score, both of which are widely adopted in anomaly detection tasks[30]. The results are summarized in Table 1.

As shown in Table 1, the proposed IWOA-LightGBM model achieves the highest accuracy (0.92) and a competitive F1-score (0.90) among all the methods compared, matching the scores of the XGBoost baseline. However, in mobile/IIoT settings with constrained compute,

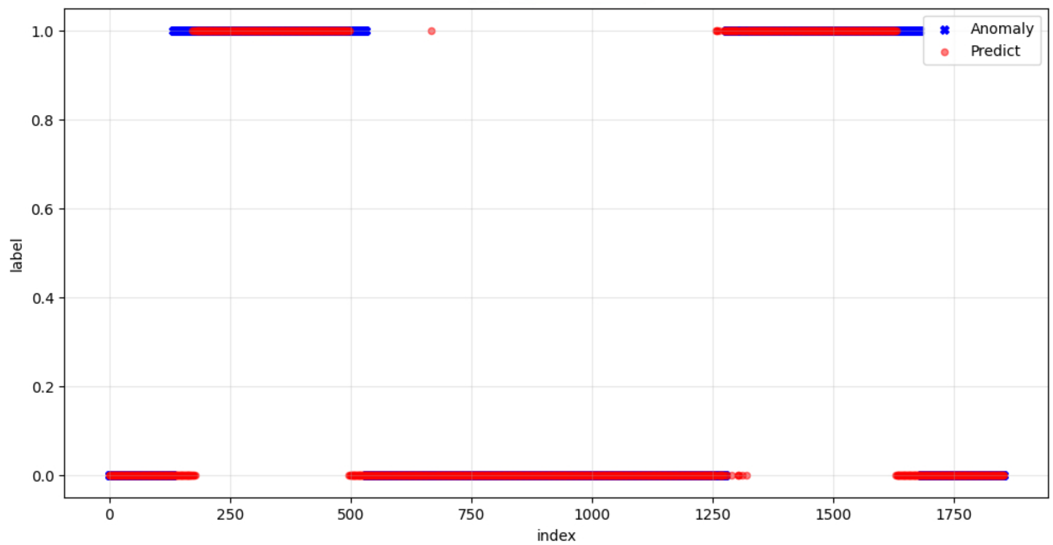

Figure 6: Label Comparison of Test Set

anomaly detectors must be updated and executed promptly to ensure timely detection. The histogram-based, leaf-wise growth used by LightGBM is typically advantageous for shortening wall-clock training time while maintaining comparable detection quality[16]. These results confirm that the IWOA-LightGBM model effectively uses the improved optimization strategy to adjust LightGBM hyperparameters. This enables the model to better adapt to the temporal continuity and high dimensionality of sensor data, thereby achieving more reliable anomaly detection.

5 Conclusion

In this paper, we construct an IWOA-LightGBM anomaly detection model with the IWOA reducing hyperparameter optimization time. IWOA enhances convergence speed via nonlinear convergence factor, adaptive inertia weight factor and stochastic differential mutation, while time-window-related features added in preprocessing improve data quality. Verified on public sensor data, the model quickly finds optimal hyperparameters and effectively detects anomalies.

In the data collected by the industrial system, more rich and relevant information can be further extracted. Different from the traditional sensor process measurements, such as flow, temperature, pressure, etc., there are also some different data types, such as images, vibration signals, and acoustic data, in real industrial scenarios, which make the industrial datasets present multimodal and heterogeneous characteristics. In mobile Internet security and industrial IoT (IIoT) deployments, where resources are constrained and timely detection is required, these multimodal characteristics are particularly salient for on-device/edge monitoring. Therefore, there is a need to study the integrated multi-modal anomaly detection framework to make full use of various data, further information extraction and feature fusion of different types of data, and to develop detection strategies based on these features. Future work will explore integrated multi-modal anomaly detection frameworks to utilize diverse data types, conduct further information extraction and feature fusion, and develop corresponding detection strategies.

Acknowledgments

The authors would like to express appreciation to the National Natural Science Foundation of China(Grant No.U21A20472 and No.U22B2005), the Natural Science Foundation of Fujian Province(Grant No.2021J01625), the Major Science and Technology Project of Fuzhou(Grant No.2023-ZD-003), and the Industrial Technology Development and Application Program of Fujian Province(Grant No.2025H0043). The authors would like to thank the editor-in-chief, editor, and reviewers for their valuable comments and suggestions.

References

[1] A. Nechibvute and H. D. Mafukidze. Integration of scada and industrial iot: Opportunities and challenges. *IETE Technical Review*, 41(3):312–325, 2024.

[2] Gen Li and Jason J. Jung. Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges. *Information Fusion*, 91:93–102, 2023.

[3] Hang Zhao, Yujing Wang, Juanyong Duan, Congrui Huang, Defu Cao, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, and Qi Zhang. Multivariate time-series anomaly detection via graph attention network. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 841–850, 2020.

[4] Ming Jin, Huan Yee Koh, Qingsong Wen, Daniele Zambon, Cesare Alippi, Geoffrey I. Webb, Irwin King, and Shirui Pan. A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10466–10485, 2024.

[5] Ailin Deng and Bryan Hooi. Graph neural network-based anomaly detection in multivariate time series. In *AAAI Conference on Artificial Intelligence*, pages 4027–4035, 2021.

[6] Zhao Sun, Qinke Peng, Xu Mou, and Muhammad Fiaz Bashir. Generic and scalable periodicity adaptation framework for time-series anomaly detection. *Multimedia Tools Applications*, 82(2):2731–2748, 2023.

[7] Daniel Fährmann, Naser Damer, Florian Kirchbuchner, and Arjan Kuijper. Lightweight long short-term memory variational auto-encoder for multivariate time series anomaly detection in industrial control systems. *Sensors*, 22(8):2886, 2002.

[8] Sepehr Maleki, Sasan Maleki, and Nicholas R. Jennings. Unsupervised anomaly detection with lstm autoencoders using statistical data-filtering. *Applied Soft Computing*, 108:107443, 2021.

[9] Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In *Artificial Neural Networks and Machine Learning – ICANN 2019: Text and Time Series*, pages 703–716, 2019.

[10] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2828–2837, 2019.

[11] Liyan Xu, Kang Xu, Yinchuan Qin, Yixuan Li, Xingting Huang, Zhicheng Lin, Ning Ye, and Xuechun Ji. Tgan-ad: Transformer-based gan for anomaly detection of time series data. *Applied Sciences*, 12(16):8085, 2022.

[12] Shikuan Shao, Hongjun Zhang, Qinfeng Xiao, Jing Wang, Xiaohui Liu, and Youfang Lin. Time series anomaly detection based on unsupervised adversarial learning. *Journal of nanjing University*, 57(6):1042–1052, 2021.

[13] Yongqian Sun, Ruru Zhang, Zihan Lin, Shenglin Zhang, Zhiyuan Tan, and Yuzhi Zhang. Evaluation of kpi anomaly detection methods. *Frontiers of Data and Computing*, 4(3):46–65, 2022.

[14] Qixu Liu, Yanhui Chen, Jieshuo Ni, Cheng Luo, Caiyun Liu, Yaqin Cao, Ru Tan, Yun Feng, and Yue Zhang. Survey on machine learning-based anomaly detection for industrial internet. *Journal of Computer Research and Development*, 59(5):994–1014, 2022.

[15] Seyedali Mirjalili and Andrew Lewis. The whale optimization algorithm. *Advances in Engineering Software*, 95, 2016.

[16] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[17] Jingmei Liu, Yuanbo Gao, and Fengjie Hu. A fast network intrusion detection system using adaptive synthetic oversampling and lightgbm. *Computers & Security*, 106:102289, 2021.

[18] Nissar Nabil, Naja Najib, and Jamali Abdellah. Leveraging artificial neural networks and lightgbm for enhanced intrusion detection in automotive systems. *Arabian Journal for Science and Engineering*, 49:12579–12587, 2024.

[19] Tao Lin, Han Yan, and Danyang Zhao. Research on fault early warning of spindle bearing based on iwoa-lightgbm. *China Measurement & Test*, 49(5):82–88, 2023.

[20] Adriano Liso, Angelo Cardellicchio, Cosimo Patruno, Massimiliano Nitti, Pierfrancesco Ardino, Ettore Stella, and Vito Renò. A review of deep learning-based anomaly detection strategies in industry 4.0 focused on application fields, sensing equipment, and algorithms. *IEEE Access*, 12:93911–93923, 2024.

[21] Bengie L Ortiz, Vibhuti Gupta, Rajnish Kumar, Aditya Jalin, Xiao Cao, Charles Ziegenbein, Ashutosh Singhal, Muneesh Tewari, and Sung Won Choi. Data preprocessing techniques for ai and machine learning readiness: Scoping review of wearable sensor data in cancer care. *JMIR mHealth uHealth*, 12(1):e59587, 2024.

[22] Zhi Wang, Yayun Li, Lei Wu, and Qiang Guo. A nonlinear adaptive weight-based mutated whale optimization algorithm and its application for solving engineering problems. *IEEE Access*, 12:40225–40254, 2024.

[23] Y. O. M. Sekyere, F. B. Effah, and P. Y Okyere. An enhanced particle swarm optimization algorithm via adaptive dynamic inertia weight and acceleration coefficients. *Journal of Electronics and Electrical Engineering*, pages 53–67, 2024.

[24] Qing He, Kangyuan Wei, and Qinshuai Xu. Mixed strategy based improved whale optimization algorithm. *Application Research of Computers*, 36(12):3647–3651,3665, 2019.

[25] H. Lotfi, M.H. Nikkhah, and M.E. Hajiabadi. Dynamic reconfiguration for energy management in ev and res-based grids using iwoa. *World Electric Vehicle Journal*, 16(8):412, 2025.

[26] S. Saleem, F. Hussain, and N.K. Baloch. Iwo-iga—a hybrid whale optimization algorithm featuring improved genetic characteristics for mapping real-time applications onto 2d network on chip. *Algorithms*, 17(3):115, 2024.

[27] Zhao Yong, Wu Junyue, Li Haobin, Liu Hua, Chen Zhiwer, and Zhang Juncheng. Fault section localization in distribution system based on ima-iwoa algorithm. In *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, pages 1614–1619, 2024.

[28] M.A.A. Masaud, S.A. Avcı, and J. Rahebi. Detecting cyberattacks in smart grids using vgg-16 and whale-fisher mantis optimization algorithm (woa-fmo). *The Journal of Supercomputing*, 81(7):857, 2025.

[29] I. D. Katser and V. O. Kozitsin. Skab - skoltech anomaly benchmark, 2020.

[30] A. Elhalwagy and T. Kalganova. Hybridization of capsule and lstm networks for unsupervised anomaly detection on multivariate data. *Applied Science*, 12(22):11393, 2022.