Design of Industrial IoT Intrusion Security Detection System Based on LightGBM Feature Algorithm and Multi-layer Perception Network

Yongsheng Deng

College of Information Engineering, Chongqing Vocational and Technical University of Mechatronics, Bishan,402760, Chongqing, China
E-mail: shenying-521@sohu.com

Received 25 September 2023; Accepted 24 October 2023;
Publication 13 February 2024

Abstract

Continuous improvement of machine learning technology has provided more support for intrusion security detection in the industrial Internet of Things (IoT). The intrusion security detection system based on LightGBM feature algorithm and multi-layer perception network fusion provides more options for improving security detection, further enhancing the effectiveness of industrial IoT intrusion security detection. By comparing the applications of different models in the security detection process, the model constructed based on the LightGBM feature algorithm can achieve higher accuracy and precision in industrial IoT intrusion security detection, as well as higher F1-score and AUC values; A more reasonable detection time also lays the foundation for improving the overall efficiency of industrial IoT intrusion security detection. Therefore, in the field of industrial IoT intrusion security

Journal of Cyber Security and Mobility, Vol. 13_2, 327–348.
doi: 10.13052/jcsm2245-1439.1327

detection, the detection model constructed in this article can provide more support for further improvement and improvement of IoT intrusion security detection performance.

Keywords: LightGBM, multi-layer perception network, industrial IoT, intrusion detection, feature selection.

1 Introduction

The rapid development of machine learning has brought earth shattering changes to people's lifestyles, but it has also brought increasingly serious network security issues. The highly transparent internet has made the manifestations of network security threats more complex [1, 2]. At the same time, computer network security protection is becoming increasingly important, and the confrontation between network intrusion and defence, infiltration and reverse osmosis will continue to escalate. This has also led to the rapid development of IoT technology, which has laid an important foundation for technological progress and continuous innovation. At the same time, the IoT industry has become one of the important protection areas for data security, gradually forming a network attack and defence situation regarding IoT security. Especially in the field of industrial IoT, relevant industrial IoT data is the core link that needs to be focused on storage and protection, and achieving intrusion security detection in industrial IoT has become a content that must be faced and valued in this field [3]. Therefore, in the industrial IoT, security intrusion has become a unique manifestation in the modern network security system. It can not only disrupt and penetrate the commercial network of the industrial IoT, but also attack and penetrate enemy government, civilian, and educational IoT network systems, posing a serious threat to the industrial IoT security system.

In the field of the Industrial IoT, network intrusion behaviour is a behaviour aimed at obtaining or controlling relevant data of the Industrial IoT, thereby attempting to damage or disrupt the integrity, availability, and confidentiality of computer data in the Industrial IoT [4, 5]. The main methods used in the process of intrusion security detection in the industrial Internet of Things are to check and analyse the data collected from samples, based on relevant analysis results, and detect whether it is an intrusion behaviour through data characteristics and behaviour types, in order to discover whether the industrial Internet of Things system is being invaded and violating security policies, targeting different types of security intrusions, Targeted

implementation of subsequent protection strategies, also known as the Industrial Internet of Things intrusion security detection mechanism. Intrusion security detection technology is a commonly used technology that can greatly ensure the security and integrity of industrial IoT system data. Combined with the IoT network intrusion security detection mechanism, it generates alarm data and quickly presents data behaviour to security technicians to avoid damage and loss [6, 7].

In the field of industrial IoT intrusion security detection, Industrial Control System (ICS) is a comprehensive concept of industrial production control systems, mainly including SCADA systems, DCS systems, and other common small control systems. ICS has become the core of critical infrastructure, and its security is related to national security and social stability, making it an important component of national security strategies [8, 9]. Industrial Control Network (ICN), as the core component of ICS, usually uses controllers, sensors, and measurement and control instruments with network communication capabilities as network nodes, and uses industrial Ethernet or fieldbus as communication media to complete the measurement and control functions of industrial data through digital, open, and multi node communication methods [10].

At present, the continuous progress and updates of industrial IoT technology have greatly improved the efficiency, but the security issues faced by industrial control networks are also becoming increasingly prominent. It is even more important to conduct in-depth analysis and research on them [11]. Therefore, this article focuses on the problems faced by network security detection in the field of industrial Internet of Things. Through the application of multi-layer perception networks, a industrial Internet of Things intrusion security detection system based on LightGBM feature algorithm and multi-layer perception networks is proposed. This system combines the LightGBM algorithm and integrates multi-layer perception networks to construct an industrial IoT intrusion security detection system with higher detection efficiency. As a major security defines measure for the industrial IoT, this system can quickly collect and analyse relevant data on network security behaviour, and use known models to determine the harm of intrusion behaviour in the industrial IoT. At the same time, relevant alarm signals are fed back before the intrusion occurs, and relevant defines measures are taken to respond. Advanced industrial IoT intrusion detection systems can largely compensate for the shortcomings of security mechanisms themselves, and their impact on ensuring the safe operation of the industrial IoT is even more crucial.

2 LightGBM Feature Algorithm Fusion with Multi-layer Perceptual Network Theory

2.1 LightGBM Feature Algorithm Theory

In neural network algorithms, LightGBM is an ensemble learning algorithm based on GBDT, which is a variant of GBDT algorithm and follows the basic framework of GBDT. GBDT uses CART regression tree as the base learner. Therefore, before introducing the LightGBM algorithm, analyse the CART regression tree [12, 13].

For the CART regression tree, assuming that the input space has been divided into *N* units R_1, R_2, R_N, as well as the fixed output value c_n on each corresponding partition unit, can obtain the predicted output of the model [14]. The CART regression tree model can be described using formula (1):

$$f(x) = \sum_{n=1}^{N} c_n I \quad I = \begin{cases} 1, & x \in R_m \\ 0, & x \notin R_m \end{cases} \tag{1}$$

After the partition of the input space is determined, square error is used to evaluate the error of the training data. Therefore, a binary tree is constructed to minimize square error as a criterion [15, 16]. Since the output of each unit is a fixed value, the average value of that unit is used as the optimal result, as shown in formula (2):

$$c_n = avg(y_i | x_i \in R_n) \tag{2}$$

In order to find the optimal segmentation variable *j* and segmentation point *s*, traverse all variables *j* and scan all segmentation points for each segmentation variable *j* to find the optimal segmentation point s. Construct a (*j, s*) and then find the optimal (*j, s*) among the entire (*j, s*) [17, 18]. According to the principle of minimizing square error, the optimal *(j, s*) can be obtained by using the minimum loss function formula (3).

$$\min_{j,s} = \min_{c_1} \sum_{x \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x \in R_2(j,s)} (y_i - c_2)^2 \tag{3}$$

Among them, c_1 and c_2 are the average output values within the R_1 and R_2 intervals, respectively.

In this process, assuming that based on the previous iteration, $f_{t-1}(x)$ is the obtained strong learner, where $L[y, f_{t-1}(x)]$ is defined as the loss function, then finding a suitable weak learner $h_t(x)$ for the CART regression

tree model becomes an important step and goal of this iteration. Among them, the loss function obtained for this round is shown in formula (4):

$$L[y, f_t(x)] = L[y, f_{t-1}(x)] - h_t(x) \tag{4}$$

Based on the requirement of fitting a CART regression tree, the negative gradient of the loss function can be used to fit the approximate value of the current round of losses. Therefore, formula (5) provides the loss function for the *i-th* sample in the *t-th* round:

$$r_{ti} = -\frac{\partial L[y_i, f(x_i)]}{\partial f(x_i)} f(x) = f_t(x) \tag{5}$$

Based on the Taylor expansion, the loss residual of the model is approximated and a regularization term is added to the loss function. At this point, the residual of the model can be analyzed using formula (6):

$$Obt^t = \sum_{i=1}^{n} L[(y, \hat{y}^{t-1}) + f_t(x) + \Omega(f_t) + C] \tag{6}$$

Among them, L represents the loss value obtained based on the true and predicted values when establishing the *t-th* tree, and Ω represents its corresponding regularization term. Select the regularization term given in formula (7) based on LightGBM:

$$\Omega(f_t) = \gamma^T + \frac{1}{2}\lambda\|w\|^2 \tag{7}$$

2.2 Multi-layer Perception Network Model

Multi-layer perceptual networks, also known as multi-layer perceptual neural networks, are traditional supervised learning methods that simulate human neurons in the field of artificial intelligence. This method can effectively handle the problem of multi classification. During the processing, nonlinear functions are learned to infinitely approximate the actual mapping relationship between the input feature space *X* and the output label vector *Y* [19, 20]. The input layer, hidden layer, and output layer are the three important stages of partitioning on the multi-layer perception network structure. As shown in Figure 1, the basic structure of a multi-layer perception network model is presented. The application of multi-layer perception networks in the industrial

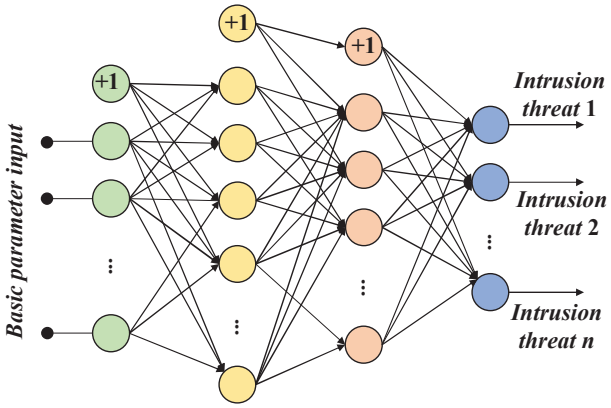


Figure 1 Schematic diagram of multi-layer perception network structure.

IoT can further promote the improvement and improvement of intrusion security detection system performance in this field, thereby providing more guarantee measures for achieving the security of industrial IoT data.

In the design process of an industrial IoT intrusion security detection system, a multi-layer perception network accumulates the effect of linear partitioning of feature attributes by each neuron, ultimately achieving non-linear abstraction of initial feature attributes, thereby enabling better linear segmentation of abstracted features [21, 22]. Considering the complexity of the model and the overfitting phenomenon caused by excessive learning, when using a multi-layer perception mechanism to create a classification model, the number of hidden layers and neurons is not necessarily better. Hiding layers and neurons too deep will exponentially increase the overall parameters of the model, greatly increasing the complexity of the model, but the classification effect of the model will not be significantly improved, and it can easily lead to excessive learning of multiple training data in the model, greatly reducing its generalization ability [23].

3 Construction of Intrusion Security Detection System Model for Industrial IoT

3.1 Application of LightGBM Feature Algorithm Fusion with Multilayer Perception Networks in Industrial IoT

In the industrial IoT, industrial control networks mostly have obvious structural systems, from bottom to top are the on-site control layer network,

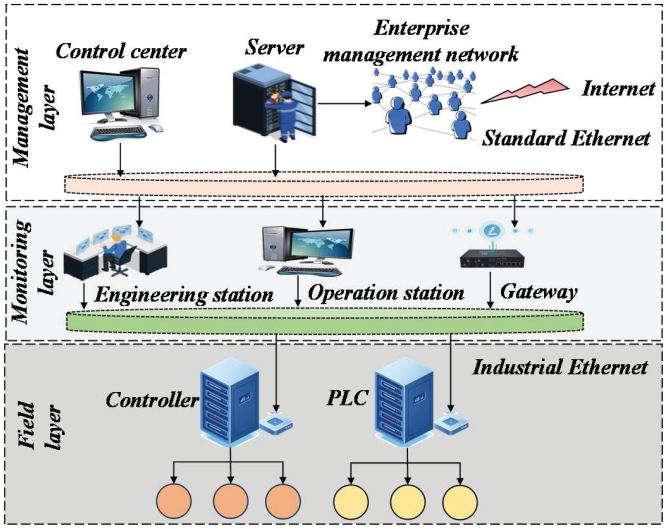


Figure 2 Typical industrial IoT security control network structure.

process control layer network/data acquisition and monitoring layer, and enterprise management layer. As shown in Figure 2, a typical industrial IoT security control network structure is presented [24].

In the on-site control layer network, distributed control systems (DCS) or programmable logic controllers (PLC) are used to achieve data sampling, instruction execution, and logic control operations for on-site devices. As the medium for the lowest level data interaction, the on-site control layer network not only needs to complete data interaction between on-site devices and achieve various collection and control tasks, but also completes data communication with the upper layer network [25].

The process control layer network is the core of ICS, and users can monitor and control the entire production process of ICS in real-time through the human-computer interaction interface. Operators collect operational data of on-site equipment through the data collection and monitoring network and report it, with high reliability. The enterprise management network is at the top level of the entire structural system, most of which are directly connected to the external internet, using TCP/IP technology and Ethernet for communication. Enterprise managers use monitoring network data to provide information services, decision support, scheduling execution, and production management for administrators, thereby completing some business-related activities.

In the field of industrial Internet of Things, the functionality of industrial control networks is divided into layers, and data transmission and exchange between layers are carried out according to agreed communication protocols, thereby achieving hierarchical management of ICS. Although the hierarchical network structure design makes the management of ICS more convenient and efficient, the overall security of the network has been threatened to a certain extent. Moreover, due to the different control tasks completed by each network layer, the requirements for security protection intensity vary greatly, so this hierarchical network structure even increases the cost of overall network security protection [26, 27].

The system characteristics of the Industrial IoT determine that intrusion detection requires high accuracy and real-time performance. Existing intrusion detection methods cannot find a good balance between the two, resulting in either poor real-time performance or low accuracy [28]. Therefore, combining the multi-layer perception network algorithm in machine learning, an industrial IoT intrusion security detection system based on LightGBM feature algorithm and multi-layer perception network was studied. LightGBM is a new integrated learning model with fast speed, high accuracy, and good performance in processing high-dimensional and large sample data, making it very suitable for industrial environment application scenarios. Building an industrial IoT intrusion detection model based on LightGBM can further improve and enhance its detection performance on traditional industrial IoT intrusion security detection systems, achieve early judgment and prediction of intrusion threats, and provide more guarantees for data security of the industrial IoT.

3.2 Construction of an Industrial IoT Intrusion Security Detection Model Based on LightGBM Feature Algorithm and Fusion of Multilayer Perception Networks

3.2.1 Intrusion security detection model framework for industrial IoT

In the design of industrial IoT intrusion security detection systems, the LightGBM algorithm abandons the decision tree growth strategy of Level wise used by most GBDT algorithms, and uses the Leaf wise algorithm with depth constraints. Compared with Level wise, the Leaf wise algorithm has better accuracy under the same number of splits, while ensuring efficiency and anti-fitting under the maximum depth limit. For this algorithm, assuming a given training dataset $D = \{(x_1, y_1), \ldots, (x_N, y_N)\}$, the lifting tree model

can represent an additive model with a decision tree as the basic learner, as shown in formula (8):

$$F(x) = \sum_{t=1}^{T} \alpha_t f_t(x) \tag{8}$$

During the model learning process, the following regularization strategy can be used to alleviate overfitting, as shown in formula (9):

$$F_t(x) = F_{t-1}(x) - v f_t(x) \tag{9}$$

Among them, $0 < v \leq 1$ is the learning rate used to update the model. The smaller the learning rate, the more regression trees need to be generated, resulting in higher accuracy. However, it also increases the training time, so it can be controlled γ, λ, ν parameters are used to construct a fast and high-precision LightGBM model.

The current typical industrial IoT network security architecture is becoming mature and widely used, and its stability and reliability have been verified. This architecture with typical characteristics such as openness and layering make system vulnerabilities vulnerable to exploitation by criminals, viruses, etc., posing a great threat to network security [29]. Moreover, this layered structure design ignores the protection of data transmission and information flow between layers, allowing attackers to attack different layers from different channels, thereby increasing the possibility of security threats to related links in the industrial Internet of Things field.

The industrial IoT intrusion security detection system based on LightGBM feature algorithm and multi-layer perception network fusion adopts a machine learning intrusion detection model that relies on existing datasets to train the model. Then, the trained model can be used to analyse network traffic and provide early warning for network attack behaviour. In the detection stage, network data packets are captured using network packet capturing tools, and relevant data features are extracted based on the input requirements of the intrusion detection model. They are input into the trained intrusion detection model, which returns the judgment result of the data packet and provides an alarm signal, so that operators can take timely measures to reduce losses. Optimizing the training set can retrain the model by independently collecting new data; During the model design process, a timed automatic retraining function can also be added to continuously adapt to new scenarios, and the traffic results detected by the host can be written into the training set for model updates [30]. As shown in Figure 3, the basic structure of the industrial IoT security control network is presented.

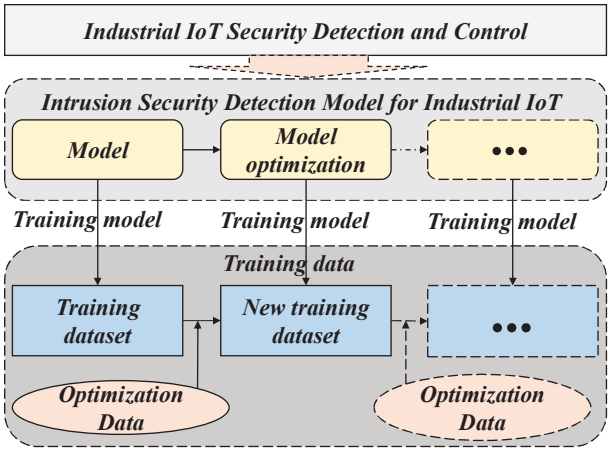


Figure 3 Machine learning based intrusion security detection for industrial IoT.

3.2.2 Construction and training of intrusion security detection model for industrial IoT

In the LightGBM algorithm, the lifting tree model uses a decision tree as the base learner and an additive model, which is a forward distribution algorithm. Initialize the first lifting tree as $f_0(x) = 0$ and gradually accumulate it through the addition model. The formula for step *n* is shown in (10):

$$f_n(x) = f_{n-1}(x) + T(x', \theta_n) \tag{10}$$

To determine the parameters of the next tree θ_n. Minimize the empirical risk value using formula (11):

$$\theta_n = arg \min_{\theta_m} \sum_{i=1}^{n} L[y_i, f_{m-1}(x_i)] + T(x_i', \theta_m) \tag{11}$$

Based on the above analysis, it can be seen that the LightGBM algorithm is an improved form of the GBDT algorithm, which uses discrete numerical outputs to solve classification problems. When the results are output as sample categories, there are certain shortcomings in the process of fitting errors. There are two methods to solve this problem: firstly, using the difference between predicted and true probability values to fit the loss; The second is to use an exponential loss function to fit the error. The following is an introduction to multi classification GBDT using logarithmic likelihood loss function.

Assuming the classification category $K(K > 2)$, the logarithmic likelihood loss function is shown in formula (12):

$$L[y, f(x)] = -\sum_{k=1}^{K} logp_k(x) \tag{12}$$

The probability $p_k(x)$ expression for the *k-th* category of sample output is shown in formula (13):

$$p_k(x) = \frac{e^{f(x)}}{\sum_{i=1}^{K} e^{f(x)}} \tag{13}$$

According to Equations (12) and (13), the negative gradient error corresponding to the *i-th* sample in category l after the *t-th* iteration can be calculated as shown in formula (14). Among them, the error is the difference between the true probability of sample *i* in the corresponding category *l* and the predicted probability of $t-1$ round.

$$r = -\left[\frac{\partial L[y_i, f(x_i)]}{\partial f(x_i)}\right]_{f_k(x)=f_{l,t-1}(x)} = y_{i,l} - p_{l,t-1}(x_i) \tag{14}$$

The LightGBM algorithm belongs to the boosting framework, so the overall idea of the LightGBM algorithm meets the overall idea of the boosting framework. Based on the core improvement steps of LightGBM algorithm for GBDT, the overall idea of LightGBM algorithm is summarized.

3.2.3 Basic process of industrial IoT intrusion security detection model

In the industrial IoT intrusion security detection system based on LightGBM feature algorithm and multi-layer perception network, the main components of the intrusion security detection process for the IoT include input dataset, data preprocessing, feature engineering, and model training and accuracy comparison [31]. As shown in Figure 4, the basic process of an industrial IoT intrusion security detection system based on the LightGBM model is presented.

(1) Input dataset: Construct a network intrusion detection system framework based on the LightGBM model according to the intrusion detection system framework. The input dataset includes KDD CUP 99, NSL-KDD, and UNSW-NB15 network intrusion detection datasets, which are divided into training and testing sets for subsequent experiments.

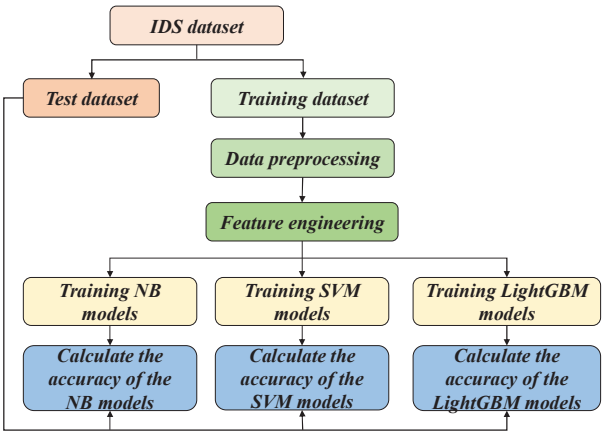


Figure 4 Network intrusion detection system based on LightGBM model.

(2) Data preprocessing: For the input training set data, it is first necessary to supplement relevant missing values for the records in the network connection records, clarify duplicate data, and standardize processing. At the same time, the main task in the detection process is to perform dimensionless data processing on the detected connection records. Due to the different measurement methods and units used for continuous attributes, there are significant differences in the values of different continuous attributes. In order to avoid deviations when measuring anisotropy, this type of attribute has been standardized, which means that the attribute values can be mapped to the standard value space using the Z-score method. The transformation process is shown in formula (15):

$$y_j = \frac{(x_j - w_j)}{s_{ij}} \tag{15}$$

Where, x_j is the nth attribute of the sample data, w_j is the average value of the *n-th* attribute of the sample data, and s_{ij} is the standard deviation of the *n-th* attribute of the sample data.

(3) Feature engineering: Machine learning models can only handle numbers, and numerical variables are variables that can take any value within a finite or infinite interval. They can be naturally represented by numbers, so they can be directly used in the model. However, the original category variable usually exists in the form of a string, so it needs to be transformed before being passed into the model, that is, feature encoding. Common feature encoding methods

include label encoder label encoding, onhotencode unique hot encoding, label binarize binary encoding, histogram encoding, and count encoding.

(4) Comparison of training and accuracy of each model: After data preprocessing and feature encoding, the system uses the naive Bayesian NB model, support vector machine SVM model, and ensemble learning LightGBM model to learn and train the training data, and compares the accuracy of each model through experiments to determine the advantages and disadvantages of different machine learning models.

4 Model Experiment and Result Analysis

4.1 Model Evaluation Indicators

The security elements in network security situational awareness have the problem of uneven data distribution. Therefore, in the industrial IoT intrusion security detection system based on LightGBM feature algorithm and multilayer perception network fusion, in order to accurately reflect the true effect of the constructed model, appropriate evaluation indicators should be selected. The confusion matrix is often used to evaluate the effective indicators of network security models, as shown in Table 1, providing commonly used evaluation effective indicators.

In the industrial IoT intrusion security detection system constructed in this article, in order to evaluate the evaluation effect of the model accurately and effectively in real environments, the following indicators are introduced to evaluate the effectiveness of the model. Precision represents the ratio of correctly detected negative samples to the true negative samples, and the calculation process is shown in formula (16):

$$P = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N} (TP_i + FP)} \tag{16}$$

Recall Rate represents the proportion of correctly detected negative samples to all negative samples, and the calculation process is shown in

Table 1 Confusion matrix

	Forecast Category	
True Category	Positive	Negative
Positive	True Positive	False Positive
Negative	False Negative	True Negative

formula (17):

$$R = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FN_i)} \tag{17}$$

Accuracy represents the proportion of correctly detected negative and positive samples to the total number, as shown in formula (18):

$$A = \frac{\sum_{i=1}^{N}(TP_i + TN_i)}{\sum_{i=1}^{N}(TP_i + FN_i + TN_i + FP_i)} \tag{18}$$

In addition, F1-score is the harmonic mean of accuracy and recall, commonly used to comprehensively evaluate the accuracy of classification models. The specific calculation process is given by formula (19):

$$F1\text{-}score = \frac{2P \times R}{P + R} \tag{19}$$

4.2 Feature Extraction

Considering the differences in the features required to identify different categories, the LightGBM algorithm has the function of balancing data distribution and extracting features. This article uses the LightGBM algorithm to learn and analyze different categories, and combines the importance of the obtained features in different categories to extract more important features in different categories based on this. In this process, first set the label of a certain attack category in the dataset to 0, and set the remaining labels to 1; Next, input the data into LightGBM for training, and the GOSS algorithm enables the sample data of each category to be learned more reasonably and fairly; Then, the EFB algorithm calculates the importance of each feature to generate a decision tree, obtaining a feature subset Hi. Finally, all feature subsets Hi are merged into a training set H that includes important feature training for all categories, as shown in formula (20):

$$H = \sum_{i=1}^{n} H_i \tag{20}$$

Apply the LightGBM algorithm to different datasets to obtain the importance of different categories of features. Filter feature importance for each category in three datasets, setting an importance threshold of 0.1. If the threshold is greater than the threshold, it is considered that the feature contains valid information, and if it is less than the threshold, it is considered invalid.

4.3 Model Test Results

In the industrial IoT intrusion security detection system based on LightGBM feature algorithm and multi-layer perception network fusion, in order to analyse the differences in the relevant evaluation parameters of the model, the model is trained and tested. The dataset is divided into a stratified sampling ratio of 8:2, with 80% of the data as the training set and 20% as the testing set.

Considering that different features in the industrial control network traffic dataset used in this article may use different measurement units, in order to avoid the problem of features with smaller value ranges being drowned by features with larger values during the training process and becoming invalid features. In order to explore the potential information of the data as much as possible, it is necessary to normalize the samples during the preprocessing stage. This article uses Z-score to normalize the data.

Accuracy and precision can reflect the effectiveness of different models. This article analyses the changes in accuracy and precision of different models in industrial IoT intrusion security detection based on Decision Tree, Random Forest, XBoot, kNN, SVM, and the LightGBM algorithm model constructed in this article. As shown in Figure 5, the differences between different models are presented. As shown in the figure, the model constructed in this article based on the LightGBM algorithm has the highest accuracy and 96.2% and 97.4%, respectively. Its accuracy is about 5.6% higher than the Decision Tree algorithm (90.6%) with the lowest value; The accuracy is about 4.2% higher than the lowest value of the kNN algorithm (93.2%).

For the Recall Rate and F1-score of the model, it can also reflect the performance differences between different models, and there are also certain

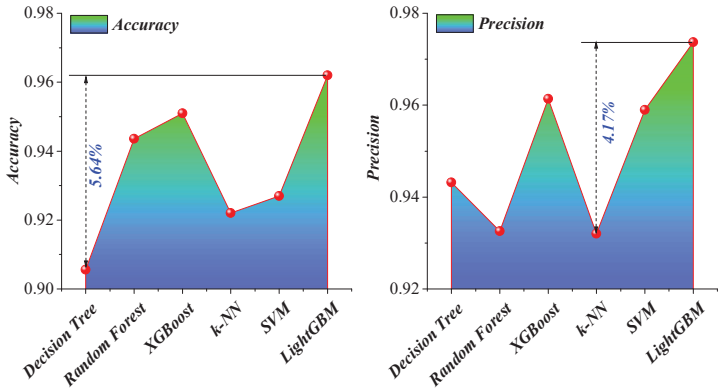


Figure 5 Comparative analysis of accuracy and precision of different models.

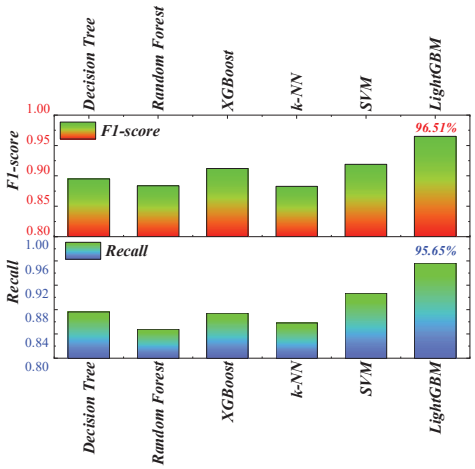


Figure 6 Comparative analysis of recall rates and F1-score of different models.

differences between models. As shown in Figure 6, the Recall Rate and F1-score between the six different models mentioned above are presented. As shown in the figure, the LightGBM algorithm-based model constructed in this article can achieve greater advantages in the intrusion security detection process of the industrial IoT. Among them, the Recall Rate of this model is about 95.6%, and the F1-score is about 96.5%, which has a higher Recall Rate and F1-score compared to other models. Therefore, from the above two evaluation indicators, the model constructed in this article has more obvious security detection advantages in industrial IoT intrusion security.

In addition, the detection time of different models was selected for comparative analysis with the model constructed in this article. As shown in Figure 7, a comparison of intrusion security detection time for industrial IoT is presented. Among them, the intrusion security detection time based on the LightGBM algorithm model constructed in this article is 1.01 s, which is about 0.78 s longer than the detection time of Decision Tree (which is 0.23 s). Although the detection time of this model is relatively lagging, its overall detection efficiency can meet the real-time requirements of security detection and achieve more efficient detection performance. Therefore, considering the real-time and accuracy indicators of the model, the method proposed in this article can be effectively applied to the detection and analysis of industrial IoT intrusion security.

In order to further compare the performance changes of different models in the industrial IoT intrusion security detection process, as shown in

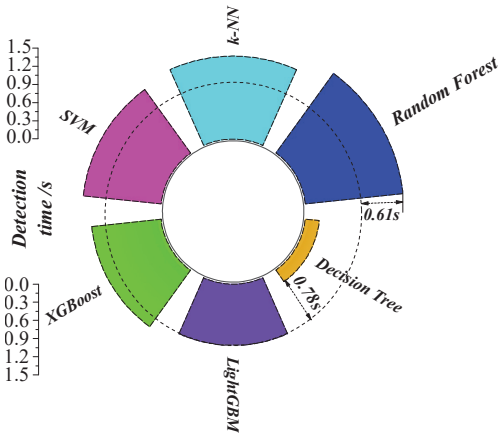


Figure 7 Comparison of detection times between different models.

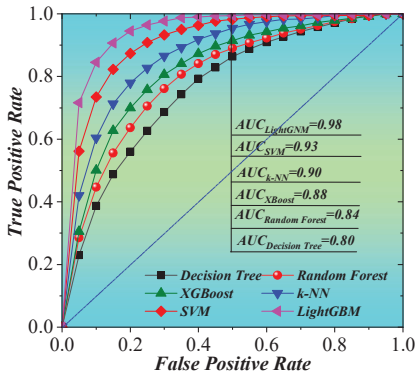


Figure 8 Comparison of ROC curves and AUC among different models.

Figure 8, the ROC curves and AUC values of each model were compared. From the figure, the ROC curve of the industrial IoT intrusion security detection model based on the LightGBM algorithm shows that it has relatively better predictive performance; At the same time, the highest AUC value of the model is 0.98, followed by SVM with an AUC value of approximately 0.93, and the smallest is Decision Tree with an AUC value of approximately 0.80.

5 Conclusions

The progress of machine learning technology has promoted the continuous improvement and progress of industrial IoT technology, laying a technical

foundation for the analysis of intrusion security detection in industrial IoT. The construction of an industrial IoT intrusion security detection system based on LightGBM feature fusion multi-layer perception network provides more effective detection measures for security detection in this field. The application of this model can further improve the accuracy, F1-score value, and other basic performance parameters of industrial IoT intrusion security detection, while also obtaining more reasonable detection time and AUC performance. Based on this, the main conclusions obtained are as follows:

(1) Combining the LightGBM feature algorithm, the constructed industrial IoT intrusion security detection system can achieve better overall performance compared to other models, achieving further improvement in predictive performance. Based on the application of this model in industrial IoT security detection, its accuracy is 96.2%, which is 5.7% higher than the decision tree model with the lowest value of 90.5%; At the same time, the accuracy of the model in intrusion security prediction reached 97.3%.
(2) In terms of Recall Rate, F1-score value, and AUC performance indicators, this model has higher comprehensive detection efficiency in the industrial IoT intrusion security detection process. Among them, the Recall Rate of this model is 95.6%, and the F1-score is about 96.5%, which has significant advantages compared to other models. At the same time, its AUC value is about 0.98, which shows a significant improvement compared to other models and demonstrates a more advantageous comprehensive detection performance, thereby providing support for the comprehensive analysis of industrial IoT intrusion security detection.

References

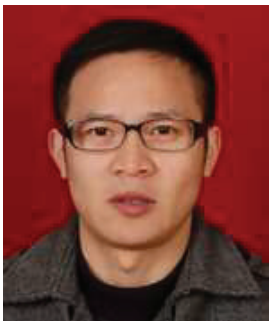
[1] Zhang, Z. Analysis of Network Security Countermeasures From the Perspective of Improved FS Algorithm and ICT Convergence. Journal of Cyber Security and Mobility, 2023, 12(01), 1–24.
[2] Yingle Y. Big data network security defense mode of deep learning algorithm[J]. Open Computer Science, 2022, 12(1):345–356.
[3] Zhao, Yan, Hu, et al. A secure and flexible edge computing scheme for AI-driven industrial IoT[J]. Cluster Computing, 2021, 26(1):1–19.
[4] Shantanu P, Zahra J. Analysis of Security Issues and Countermeasures for the Industrial Internet of Things[J]. Applied Sciences, 2021, 11(20):93–102.

[5] Thilagavathi, B., and Suthendran, K. Boosting Based Implementation of Biometric Authentication in IoT. Journal of Cyber Security and Mobility, 2018, 7(1–2), 131–144.
[6] Aron L, Waseem A, Yevgeniy V, et al. Integrating redundancy, diversity, and hardening to improve security of industrial internet of things[J]. Cyber-Physical Systems, 2020, 6(1):1–32.
[7] Alshathri S, Sayed E A, Shafai E W, et al. An Efficient Intrusion Detection Framework for Industrial Internet of Things Security[J]. Computer Systems Science and Engineering, 2023, 46(1):819–834.
[8] Kollipara, V. N. H., Kalakota, S. K., Chamarthi, S., Ramani, S., Malik, P., and Karuppiah, M. Timestamp Based OTP and Enhanced RSA Key Exchange Scheme with SIT Encryption to Secure IoT Devices. Journal of Cyber Security and Mobility, 2023, 12(01), 77–102.
[9] Zayed S, Gamal A, Ayman S E, et al. An Efficient Fault Diagnosis Framework for Digital Twins Using Optimized Machine Learning Models in Smart Industrial Control Systems[J]. International Journal of Computational Intelligence Systems, 2023, 16(1):63–84.
[10] Alzahrani A, Aldhyani H H T. Design of Efficient Based Artificial Intelligence Approaches for Sustainable of Cyber Security in Smart Industrial Control System[J]. Sustainability, 2023, 15(10):123–135.
[11] Xinyi Q, Min L, Lu Z, et al. Structural protein fold recognition based on secondary structure and evolutionary information using machine learning algorithms[J]. Computational Biology and Chemistry, 2021, 91.
[12] Meng W, Kejun S, Caiwang T, et al. Research on fault diagnosis system for belt conveyor based on internet of things and the LightGBM model.[J]. PloS one, 2023, 18(3):23–39.
[13] Xie B, Li F, Li H, et al. Enhanced Internet of Things Security Situation Assessment Model with Feature Optimization and Improved SSA-LightGBM[J]. Mathematics, 2023, 11(16):212–234.
[14] Lv Z. A novel LightGBM-based industrial internet intrusion detection method[J]. International Journal of Computer Applications in Technology, 2023, 71(3):208–216.
[15] Qingmin Y, Xin G, Yong Z, et al. The missing data filling method of the industrial internet platform based on rules and LightGBM[J]. IFAC PapersOnLine, 2020, 53(5):152–157.
[16] Leevy L J, Khoshgoftaar M T, Hancock J. Feature evaluation for IoT botnet traffic classification[J]. International Journal of Internet of Things and Cyber-Assurance, 2022, 2(1):87–102.

[17] Prajisha C, Vasudevan A R. An efficient intrusion detection system for MQTT-IoT using enhanced chaotic salp swarm algorithm and LightGBM[J]. International Journal of Information Security, 2022, 21(6):1263–1282.
[18] Meng D, Dai H, Sun Q, et al. Novel Wireless Sensor Network Intrusion Detection Method Based on LightGBM Model[J]. IAENG International Journal of Applied Mathematics, 2022, 52(4):654–668.
[19] Diana Lopez-Soto, Francisco Angel-Bello Soumaya Yacout. A multi-start algorithm to design a multi-class classifier for a multi-criteria ABC inventory classification problem[J]. Expert System with Applications, 2017, 81(C):12–21.
[20] Zhao K, Wang D. Research on Speech Recognition Method in Multi-Layer Perceptual Network Environment[J]. International Journal of Circuits, Systems and Signal Processing, 2021, 15(8):996–1004.
[21] Huang X, Gao L, Crosbie S R, et al. Groundwater Recharge Prediction Using Linear Regression, Multi-Layer Perception Network, and Deep Learning[J]. Water, 2019, 11(9):457–469.
[22] Zamil A H G M, Samarah S, Rawashdeh M, et al. Multimedia-oriented action recognition in Smart City-based IoT using multilayer perceptron[J]. Multimedia Tools and Applications, 2019, 78(21):315–329.
[23] Mahmoud H, Sayed Y M, Galal A I, et al. Smart Cognitive IoT Devices Using Multi-Layer Perception Neural Network on Limited Microcontroller[J]. Sensors, 2022, 22(14):142–163.
[24] Sankaran K S, Kim B-H. Deep learning based energy efficient optimal RMC-CNN model for secured data transmission and anomaly detection in industrial IOT[J]. Sustainable Energy Technologies and Assessments, 2023, 56(5):89–103.
[25] Kalinin M O, Krundyshev V M, Sinyapkin B G. Development of the Intrusion Detection System for the Internet of Things Based on a Sequence Alignment Algorithm[J]. Automatic Control and Computer Sciences, 2021, 54(8):993–1000.
[26] Huang Zhaojun, Zeng Mingru. Hierarchical ICS Intrusion Detection Algorithm Based on RVM Combined with GSA-SVM[J]. Control Engineering of China, 2022, 29(7):1323–1329.
[27] Doynikova E., Fedorchenko, A., and Kotenko, I. A Semantic Model for Security Evaluation of Information Systems. Journal of Cyber Security and Mobility, 202,9(2), 301–330.

[28] Knowles W, Prince D, Hutchison D, et al. A Survey of Cyber Security Management in Industrial Control Systems[J]. International Journal of Critical Infrastructure Protection, 2015, 9(3):52–80.
[29] Liu Jin, Zhao Jing, Feng Yingmin. Power Load Forecasting in Power Internet of Things Based on Gradient Boosting Decision Tree[J]. Power Planning, 2022, 50(8):46–53.
[30] Shen Yeming, Li Beibei, Liu Xiaojie, Ou Yangyuankai. Research on Active Learning-based Intrusion Detection Approach for Industrial Internet[J]. 2023, 21(1):80–87.
[31] Bagui S, Wang X, Bagui S. Machine Learning Based Intrusion Detection for IoT Botnet[J]. International Journal of Machine Learning and Computing, 2021, 11(6):122–136.

Biography



Yongsheng Deng was born in Bazhong, Sichuan, P.R. China, in 1978, he received the Master degree from Chongqing University, P.R. China. Now, he works in College of Information Engineering, Chongqing Vocational and Technical University of Mechatronics. His research interests include Internet of Things technology application, network security detection, big data application development, etc.