

# CatBoost-based Intrusion Detection Method for the Physical Layer of Smart Agriculture

Zizhong Wei<sup>1,2,†</sup>, Fanggang Ning<sup>3,†</sup>, Kai Jiang<sup>2</sup>, Yang Wang<sup>2</sup>, Zixiang Bi<sup>2</sup>, Qiang Duan<sup>2</sup>, Jichen Zhang<sup>2</sup>, and Rui Li<sup>2,\*</sup>

<sup>1</sup>Key Laboratory of Agricultural Blockchain Application, Ministry of Agriculture and Rural Affairs, Beijing, China

<sup>2</sup>Inspur Academy of Science and Technology, Jinan, Shandong, China

<sup>3</sup>Inspur Software Co. Ltd, Jinan, Shandong, China

**Abstract.** Agriculture holds a pivotal role in the progress of human society. The challenges stemming from a burgeoning population, land degradation, water scarcity, and urbanization have intensified the need for more efficient agricultural production. While smart farming brings significant benefits to farmers and agricultural output, it also introduces complex cybersecurity risks to agricultural production. The security of the physical layer in smart agriculture is intricately tied to crop growth and yield, with indirect implications for the security of the network and application layers. This paper introduces a novel intrusion detection scheme based on CatBoost for the physical layer and evaluates its effectiveness using the publicly available ToN\_IOT dataset. In binary classification results, the scheme achieves a remarkable recognition accuracy of 99.94%, along with a precision and recall of 99.88%. In multi-classification results, the scheme outperforms other existing solutions across all metrics. The experimental findings clearly illustrate the exceptional recognition accuracy of this implemented method against physical layer attacks within the domain of smart agriculture. Furthermore, the system's implementation ensures the security of input data for the smart agriculture network layer, cloud, and blockchain applications.

## 1 Introduction

Agriculture has been integral to human development. Firstly, agriculture's emergence ensured sufficient food supplies to meet basic subsistence needs. The enhancement of agricultural yield and quality played critical roles in preventing famines and guaranteeing societal stability. Secondly, agriculture has served as the economic foundation for several nations. Agriculture supports significant employment opportunities and contributes to regional economic growth through the production and trade of agricultural goods. Moreover, agricultural development is inextricably linked to land, water resources, and ecosystems. Implementing scientific methods in agriculture can prevent soil erosion, lessen water pollution and mitigate ecosystem damage, promoting sustainable development over time.

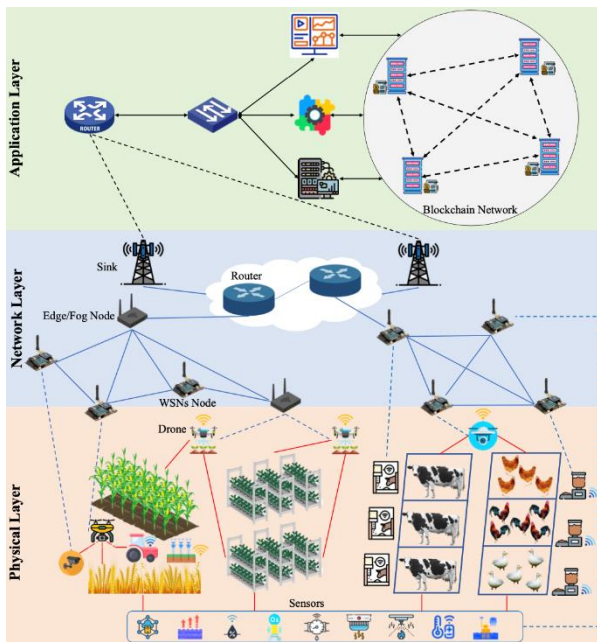
Nevertheless, numerous real-world challenges demand increased efficiency in agricultural production. Worldwide, 500 million people remain undernourished, and 821 million experience famine – a situation that will exacerbate with population growth [1]. Furthermore, overexploitation of land and pollution of water have hindered crop yields. Urbanisation has resulted in a decrease in arable land and agricultural labour, leading to a significant impact on the effectiveness and scope of agricultural production [2]. Consequently, finding a solution to the food supply problem brought about by population growth in the context of a diminishing

number of farmers and limited agricultural production resources has become a primary concern in agricultural development today.

Increasing agricultural productivity is crucial for solving current industry challenges. To enhance efficiency and reduce operational expenses, agriculture has seen an increased adoption of cutting-edge and innovative technologies, including artificial intelligence, Internet of Things (IoT), blockchain, and cloud computing [3]. These implementations have formed a sophisticated, smart agriculture system, as shown in Fig. 1. The smart agriculture system's physical layer comprises drones and IoT devices, which gather and monitor data related to crops, soil, weather, and livestock. When certain conditions are met, the relevant controllers are prompted to carry out automated management actions, such as irrigation or automatic feeding. In addition, IoT devices merge collected data into edge or fog nodes to form various network types at the network layer. Cloud computing and blockchain are utilised at the application layer for modelling and trusted storage of agricultural data. Predictive analytics based on agricultural data and management decision-making with regards to agricultural production are enabled via cloud computing. Transparency and traceability of the agricultural supply chain is provided by blockchain.

† These authors contributed equally to this work

\* Corresponding author: [lirui01@inspur.com](mailto:lirui01@inspur.com)



**Fig. 1.** Smart agriculture system schematic.

The incorporation of technology across all aspects of smart agriculture has augmented the effectiveness of agricultural production but created various cybersecurity obstacles. Cyber threats targeting agriculture, exemplified by the hacking of the Australian wool trading software and remote access to the Florida water supply system, attest to the potential dangers of hacking smart agriculture systems. Physical layer sensors and control devices are highly susceptible to attacks. These devices are deployed directly in the agricultural production environment, making them vulnerable to hijacking by attackers through physical contact or wireless networks. Furthermore, sensing and control devices lack strong security protections against robust cyberattacks. Once an attacker gains control of a physical layer device, they could launch a range of attacks. These include distributed denial-of-service attacks, which prevent access to the network service where the device is located, and false data injections. Attacks on the physical layer could result in the complete destruction of agricultural production. In addition, the spread of malicious data through the wireless sensor network to the network layer and application layer can have adverse effects on the entire system's normal operation.

The cybersecurity sector has been working to identify rapid and efficient resolutions to the security obstacles confronted in agricultural production. Intrusion detection systems, a type of malicious behaviour detection system, is well-suited to the intricate network environment and unpredictable network attacks of agricultural IoT devices. An intrusion detection system is capable of monitoring and evaluating the inputs to both the network and devices. It swiftly detects the presence of intrusion behaviours and subsequently initiates alarms or activates dependable security protection measures to lessen the consequences provoked by the attack behaviours.

Signature-based schemes are the traditional means of detecting intrusions. They are limited to identifying

attacks with predetermined attack patterns or specified signatures, and their effectiveness is dependent on the volume and variety of signatures. Unknown intrusions, however, cannot be identified by such schemes. To address unknown and fluctuating network attacks, academia has also proposed anomaly-based intrusion detection methods. Such schemes offer greater flexibility, by constructing models based on intrusion behaviour features to detect and identify anomalies. To enhance the security protection of the smart agriculture system, this article suggests a CatBoost-based intrusion detection scheme for the physical layer of smart agriculture, considering that the physical layer has complex attack surfaces, diverse network structures and large amounts of data. This approach follows the concept of anomaly-based intrusion detection and utilizes the precision and generalizability of the CatBoost algorithm for the prompt and efficient recognition of intrusions in agricultural IoT systems. To validate the efficacy of the proposed scheme in detecting physical layer intrusions amidst diverse and complex security threats, the ToN\_IoT dataset serves as the training model and validation source, comprising 9 categories of typical IoT attacks including backdoor, man-in-the-middle, denial-of-service, scanning, and ransomware, among others. This paper's principal contributions are:

1. The physical layer intrusion detection system for smart agriculture presented in this research paper, which utilises CatBoost technology, attains a 99.95% accuracy in identifying anomalies and 99.38% accuracy in identifying attack types in the ToN\_IoT dataset.
2. In this study, the CatBoost model's categorical variable processing method was employed to address the issue of over-categorisation in agricultural IoT data. Additionally, the weighted cross-entropy loss function was utilised to tackle the problem of extremely unbalanced attack category distribution.
3. In this study, two intrusion detection models were developed using the CatBoost model. To detect intrusions in physical layer devices that have limited computational resources, a binary classification model has been created. Additionally, a multiclassification model has been designed to detect attack types in data forwarding nodes, thus enabling the implementation of protective measures.

## 2 Related Work

This section evaluates intrusion detection methods proposed for smart agriculture, as well as those concerning the physical layer. First, the different types of solutions for smart agriculture intrusion detection are analysed chronologically. Then, as the section categorises smart agricultural physical layer devices into three types: IoT devices, drones, and unmanned tractors, it also summarises the corresponding intrusion detection solutions.

## 2.1 Smart Agriculture Intrusion Detection Solutions

The initial intrusion detection method suggested for smart agricultural systems is Thakur et al.'s intrusion detection scheme that utilises sensor data to detect intrusions before forwarding the data to the cloud. The scheme's effectiveness is demonstrated by the experimental findings. In 2021, Ferrag et al. introduced a deep learning-based intrusion detection system to identify distributed denial of service attacks for fog computing [4]. The study implemented three deep learning frameworks to train and test the system, demonstrating its remarkable ability to detect DDoS attacks. In 2022, Raghuvanshi et al. devised a machine learning intrusion detection scheme for smart agriculture irrigation systems [5]. The scheme transforms symbolic features into numerical ones via principal component analysis, and assesses the intrusions' accuracy using various machine learning algorithms. In 2023, Kethineni and colleagues introduced an advanced deep learning framework to identify intrusions into the fog layer in smart farming systems [6]. The model combines a merged CNN with a bidirectional gated recurrent unit (Bi-GRU) and achieves high accuracy in intrusion detection across a publicly available dataset.

## 2.2 Smart Agriculture Physical Layer Related Intrusion Detection Solutions

This section examines intrusion detection methods concerning agricultural drones, autonomous tractors, and agricultural IoT devices. These aforementioned devices constitute the primary device types in the physical layer of smart agriculture.

### 2.2.1 Intrusion Detection System for Drones

As unmanned aerial vehicle (UAV) technology continues to develop, certain studies have suggested methods for utilising UAV swarms to work together towards defined agricultural objectives, ultimately increasing productivity and reducing manual labour. However, these systems are vulnerable to cybersecurity breaches, potentially resulting in malicious takeover of UAVs and disruption of operations. Current drone intrusion detection schemes can be divided into two categories: those using machine learning and those based on blockchain.

In 2021, Ramadan et al. implemented a real-time data analysis framework using recurrent neural networks to identify intrusion detection schemes in self-organising networks for UAVs [7], demonstrating the superiority of their approach based on simulation data. In 2019, Arthur proposed a self-learning multi-class support vector machine (SVM)-based intrusion detection scheme for UAV networks [8], which achieved high accuracy, sensitivity, and specificity against security attacks. In 2022, Ihekoronye and colleagues put forward an anomaly-based hierarchical optimised random forest intrusion detection system (IDS) [9]. Simulation results indicate that the model

obtains the highest F1 scores and the smallest mean squared error in predicting various lethal attacks, when compared to other systems. In 2023, Subbarayalu and coauthors proposed a timed probabilistic-based automata IDS [10], aiming to emulate normal UAV swarm behaviour and recognise intrusion instances. The results from the experiment demonstrate the effectiveness of this approach for identifying various types of attacks with high adaptability.

With regards to the blockchain methodology, Ferrag et al. introduced an intrusion detection system named DeliveryCoin [11], which employs hash functions and short signatures. The system identifies autopilot network attacks and counterfeit transactions between autopilot nodes at each node. The simulation results validate the framework's accuracy and latency. In 2023, Heidari put forward a radial basis function neural network model which employs blockchain technology to enhance data integrity and storage [12]. The model is geared towards facilitating refined intelligent decision-making across distinct UAV networks.

### 2.2.2 Intrusion Detection System for Autonomous Tractors

The use of automated tractors in smart agriculture has rapidly increased. Similar to self-driving cars, these tractors operate within large networked environments, which present potential cybersecurity risks. In order to mitigate these risks, IDS proves to be an effective class of solutions.

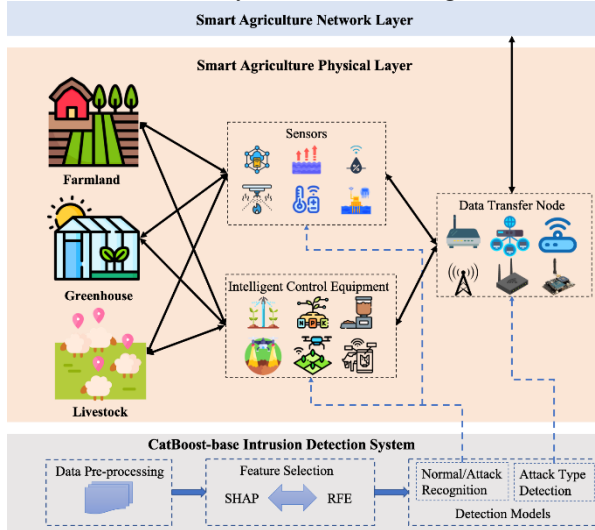
Song and colleagues propose a deep convolutional neural network-based intrusion detection system aimed at safeguarding the control network of vehicles [13]. The architecture reduces superfluous complexity in the Inception-ResNet model and specifically optimises data traffic. The scheme exhibits exceptionally low miss and false alarm rates, as evidenced by actual vehicle data. Wyk and colleagues utilised a convolutional neural network alongside an anomaly detection scheme and Kalman filtering [14], employing an  $X^2$  detector to recognise and classify unusual behaviour in autonomous vehicles. Experimental findings demonstrate the efficacy of this method, exhibiting superior precision and sensitivity.

### 2.2.3 Intrusion Detection System for IoT Devices

Current intrusion detection techniques for Internet of Things (IoT) devices can be categorized into three types: deep learning-based, hybrid machine learning-based, and artificial bee colony-based.

Almiani et al. proposed a completely automated intrusion detection system utilizing multi-layered recurrent neural networks for identifying attacks in IoT environments within the deep learning scheme [15]. The model's stability and robustness with respect to various performance metrics are demonstrated through experimental results and simulations. Li et al. utilise convolutional neural networks to overcome image security detection problems and adapt to diverse media

types [16]. This method has exhibited exceptional classification accuracy on a standard image database.



**Fig. 2.** Smart Agriculture Physical Layer Intrusion Detection Overview

For the hybrid learning scheme, GARUDA, an IoT anomaly detection scheme based on incremental clustering of feature patterns, was proposed by Aljawarneh et al [17]. Jiang et al. presented a new multi-channel intelligent attack detection method that relies on long and short-term memory recurrent neural networks [18]. The results of the experiments confirmed the superiority of this attack detection method over several others that use feature detection as well as Bayesian or support vector machine classifiers for attack detection.

For the artificial swarm scheme, Murali and colleagues have devised a lightweight intrusion detection approach using the artificial swarm model. This method is capable [19], with flexibility and precision, to detect Sybil attacks.

### 3 Methodology

The physical layer of smart agriculture comprises a range of devices, which include controllers, nodes, and data collectors. These devices are susceptible to attacks by adversaries due to weak security measures and easy access. Incidents of cyber-attacks on agricultural devices may lead to severe crop damage through the execution of malicious agricultural practices, such as inappropriate irrigation or fertiliser application. Hijacked physical layer devices may enable attackers to carry out internal attacks on network and application layers, resulting in network layer paralysis, mispredictions at the application layer, and attacks on the agricultural blockchain. To safeguard smart agricultural systems, it is vital to secure the physical layer. Hence, an intrusion detection system should be put in place at this layer.

This chapter provides an overview of a CatBoost-based intrusion detection scheme for the smart agriculture application layer. The figure depicted in Fig. 2 displays the overall scheme, whereby the intrusion detection system is implemented at the physical layer data aggregation nodes, command issuing, and message

forwarding nodes, to identify any physical layer intrusions. The construction process of the intrusion detection system involves the following steps: 1. Process the IoT network traffic data; 2. Use SHAP value and recursive feature elimination methods for feature selection; 3. Train two Catboost models, one specifically for binary classification of normal records and attack records, and one for multiple classification to identify the specific attack type to which each record belongs; 4. Based on demand and computing power, deploy the binary CatBoost model on terminal nodes with low computing power, and deploy the multivariate classification model on core nodes. The above process is described in detail below.

### 3.1 Data Processing

Agricultural IoT devices produce substantial volumes of network traffic data during intercommunication, and hidden intrusions can be found within this data. The current investigation employs an enhanced CatBoost algorithm to develop an intrusion detection system relying on network traffic data occurring within the physical layer of agricultural IoT. To guarantee the system's efficacy, categorical features and class imbalance must undergo processing. The forthcoming segment presents the data processing approach of this intrusion detection system.

#### 3.1.1 Categorical Features Processing

An important feature of network traffic data is that it contains many non-numeric characteristics. For example, network protocol type, domain name, etc. There are also features that appear to be numerical values but are actually category numbers, such as network ports. These features are all categorical features. If they are directly converted into numerical values, it is likely to reduce the detection accuracy of the system.

Two main methods are used to address categorical data, with the classical approach being one-hot coding of categorical features. Although, if a feature has an excessive number of categories, one-hot coding creates a high quantity of dummy features, hampering training and inference. Alternative methods are employed by various algorithms to handle categorical features such as LightGBM [20], which takes advantage of optimal segmentation techniques to digitise categorical features. The CatBoost [21] model utilised in this study applies a combined method, incorporating one-hot coding for variables with fewer categories and numerical coding for variables with more categories.

#### 3.1.2 Class Imbalance

Class imbalance is another problem faced when building our intrusion detection system. The amount of data generated by different intrusion methods is very different. For example, DDOS attacks will generate a large amount of network data, while other attack methods may only leave a small amount of records. This leads to an extreme imbalance in the amount of collected

data on different attack types, resulting in poor identification of attack types with fewer records.

There are two main solutions to this problem. One is to process the data, undersample categories with a large number of records, and oversample categories with a small number of records, for example the SMOTE [22] method; one is to process the loss function, increasing the weights of categories with a small number of records. Both undersampling and oversampling will affect the distribution of data, and when there are many categorical features, oversampling may not be effective. For example, in the study of Gad et al. [23], the F1-score decreased slightly after using the SMOTE method. Therefore, in this study, the weight of the loss function is adjusted to deal with the class imbalance problem.

### 3.2 Feature Engineering

In this study, we drop features such as timestamps, IP addresses, etc. that may lead to overfitting. We use SHAP values (SHapley Additive exPlanations) [24] and recursive feature elimination [25] methods to perform feature selection on the remaining features, retaining only a specified number of features to improve the inference speed while ensuring the accuracy of the results. SHAP shows the contribution or the importance of each feature on the prediction of the model. Recursive feature elimination is a method of continuously dropping features with poor contribution in iterations.

### 3.3 CatBoost

The intrusion detection system constructed in this study uses the CatBoost algorithm. CatBoost is a machine learning framework based on the Gradient Boosting Decision Tree (GBDT). The GBDT algorithm uses decision trees as weak learners. As shown in Fig. 3, when a weak learner completes learning, the current gradient of the loss function is calculated, and the next weak learner is used to fit the gradient. Eventually these weak learners are added together to form a strong learner. CatBoost uses an improved GBDT algorithm, its main process is shown in Algorithm 1.

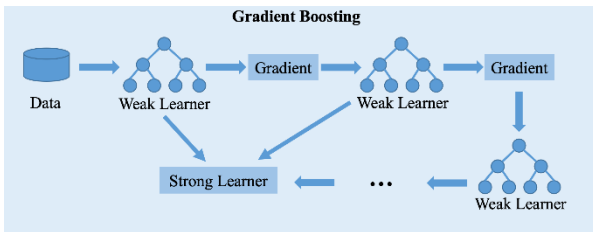


Fig. 3. Gradient Boosting Schematic.

CatBoost has the following main features:

(1) Catboost introduces a categorical variable processing method that combines one-hot encoding and numerical encoding, which can efficiently process categorical features and solve the problem of too much categorical data in our intrusion detection system. For category features whose number of unique values exceeds the threshold, CatBoost uses the following method to encode them. Namely, assume that we are

given a dataset of observations  $D = (\mathbf{X}_i, Y_i), i = 1 \dots n$ , where  $\mathbf{X}_i = (x_{i,1}, \dots, x_{i,m})$  is a vector of  $m$  features, some numerical, some categorical, and  $Y_i \in \mathbb{R}$  is a label value. Let  $\sigma = (\sigma_1, \dots, \sigma_n)$  be the permutation, then  $x_{\sigma_p,k}$  is encoded by

#### Algorithm 1 CatBoost Updating Algorithm [21]

**Input :**  $\{(\mathbf{X}_k, Y_k)\}_{k=1}^n$  ordered according to  $\sigma$ , the number of trees  $L$ ;  
**Output:**  $M_1, \dots, M_n; M_1(\mathbf{X}_1), \dots, M_n(\mathbf{X}_n)$   
 1:  $M_i \leftarrow 0$  for  $i = 1, \dots, n$ ;  
 2: **for**  $iter \leftarrow 1$  to  $L$  **do**  
 3:     **for**  $i \leftarrow 1$  to  $n$  **do**  
 4:         **for**  $j \leftarrow 1$  to  $i - 1$  **do**  
 5:              $g_j \leftarrow \frac{d}{da} Loss(y_j, a)|_{a=M_i(X_j)}$  ;  
 6:         **end for**  
 7:          $M \leftarrow LearnOneTree((\mathbf{X}_j, g_j) \text{ for } j=1, \dots, i-1)$  ;  
 8:          $M_i \leftarrow M_i + M$ ;  
 9:     **end for**  
 10: **end for**

$$\frac{\sum_{j=1}^{p-1} [x_{\sigma_j,k} = x_{\sigma_p,k}] Y_{\sigma_j} + a \cdot P}{\sum_{j=1}^{p-1} [x_{\sigma_j,k} = x_{\sigma_p,k}] + a}, \quad (1)$$

where  $[\cdot]$  denotes Iverson brackets, i.e.,  $[x_{j,k} = x_{i,k}]$  equals 1 if  $x_{j,k} = x_{i,k}$  and 0 otherwise.  $P$  and  $a > 0$  are prior values. From the formula, we know that records with the same feature value will receive different encodings due to different positions in the dataset.

(2) Catboost introduces a weighted cross-entropy lossfunction, which makes it easy to adjust the weight of different features in the loss function, and has great advantages when dealing with extremely imbalanced data. It can solve the problem of category imbalance in this system. The formula of the loss function is as follows:

$$\frac{\sum_{i=1}^N \sum_{j=0}^{M-1} \omega_j ([j = Y_i] \log p_{ij} + [j \neq Y_i] \log(1 - p_{ij}))}{N \sum_{j=0}^{M-1} \omega_j}, \quad (2)$$

where  $N$  is the length of the dataset,  $M$  is the number of classes,  $\omega_j$  is the weight of class  $j$ ,  $Y_i$  is the real class of the  $i$ -th record,  $p_{ij}$  is the predicted probability of record  $i$  belonging to class  $j$ , and  $[j = Y_i] = 1$  if  $j = Y_i$ ,  $[j = Y_i] = 0$  if  $j \neq Y_i$ .

(3) The reasoning speed is fast and can respond quickly to attacks. Using the Catboost algorithm, our intrusion detection system effectively solves the two main difficulties raised in Section 3.1 and achieves highly accurate detection of physical layer intrusions in smart agriculture.

## 4 Evaluation and Discussion

This section introduces the results of evaluating the intrusion detection system built in Section 3. First, we select the ToN\_IoT dataset that is similar to the actual data of agricultural IoT, and then we process the dataset according to the method described in Section 3. And we use this dataset to train two models for intrusion detection and attack type detection respectively. Finally, we show the performance of the two models on common

evaluation metrics and compare them with other popular machine learning algorithms.

#### 4.1 Dataset and Preprocessing

To evaluate our intrusion detection system, we need a suitable dataset. In our smart agriculture architecture, the physical layer is the Internet of Things composed of many sensors. In order to reflect this feature, we chose the ToN\_IoT dataset as the evaluation dataset. The ToN\_IoT dataset [26] is a new generation of Internet of Things (IoT) and Industrial IoT (IIoT) dataset for evaluating the fidelity and efficiency of different cybersecurity applications based on Artificial Intelligence (AI). It is released by the Cyber Range Lab of the Australian Centre for Cyber Security (ACCS) in 2019. And it is a comprehensive dataset that includes the telemetry data of an IoT network. Several portions of the dataset contain different traces of IoT services, network traffic, and Operating System (OS) logs. The data was generated using a realistic network testbed. The dataset contains several attack scenarios such as backdoor, DoS, Distributed DoS (DDoS), injection, Man in the Middle (MITM), password, ransomware, scanning, and Cross-Site Scripting (XSS). All data points in the network ToN\_IoT dataset are made up of 44 attributes and an attack-type labeled as normal or attack. As presented in Table 1, we show the normal and attack statistics for network data records in the train-test ToN\_IoT dataset.

**Table 1.** The types and numbers of records in the entire ToN\_IoT dataset and its testing and training sets

Type	Dataset	
	ToN_IoT	Train Test datasets
backdoor	508116	20000
ddos	6165008	20000
dos	3375328	20000
injection	452659	20000
mitm	1052	1043
password	1718568	20000
ransomware	72805	20000
scanning	7140161	20000
XSS	2108944	20000
normal	796380	300000

During pre-processing, we dropped the time and ip features (called 'ts', 'src\_ip' and 'dst\_ip' in the dataset) that may cause overfitting. Missing values are filled using the highest frequency value of the corresponding feature.

We split the train-test ToN\_IoT dataset into two parts, with 70% of the records as the training set and 30% of the records as the test set. On the training set, we used 5-fold cross-validation to ensure the stability of the training results. On the test set, we calculated a variety of metrics based on the prediction results to evaluate the effectiveness of our intrusion detection model

#### 4.2 Metrics for Evaluation

In order to evaluate our intrusion detection system, we used the trained model to calculate a variety of metrics

on the test set, including Accuracy, Precision, Recall and F1-score, as well as the False Positive Rate. These measurements were constructed using true negative (TN), true positive (TP), false negative (FN), and false positive (FP) data. Taking binary classification as an example. True Positive (TP) is the total number of actual attack records that are correctly identified as attacks. True Negative (TN) refers to the total number of real records that are correctly classified as normal records. False Negative (FN) refers to the total number of real attack samples that are incorrectly detected as normal. False Positive (FP) refers to the total number of normal samples that are incorrectly identified as attacks.

Accuracy (Acc) refers to the proportion of correctly classified attack and normal records to all records, indicating the overall efficiency of the system. Accuracy is mathematically stated as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \quad (3)$$

Precision (Pre): Defines the percentage of genuinely detected attacks versus all records designated as attacks; arithmetically expressed as follows:

$$Pre = \frac{TP}{TP + FP}. \quad (4)$$

Recall (Rec): The system's ability to correctly detect attacks when a security breach occurs; often known as the true positive rate, and expressed mathematically as follows:

$$Rec = \frac{TP}{TP + FN}. \quad (5)$$

F1-score (F1) is theoretically defined as the harmonic average of recall and precision.

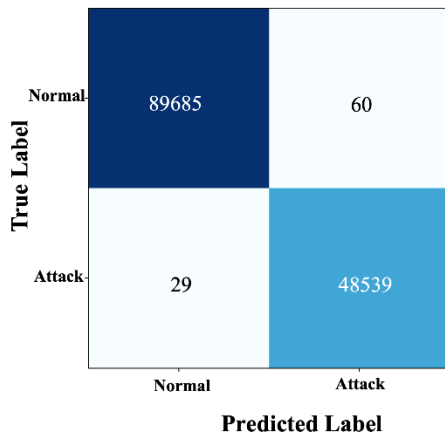
$$F1 = 2 * \frac{Rec \cdot Pre}{Rec + pre}. \quad (6)$$

The False Positive Rate (FPR) is the ratio between the number of normal records, which is predictable as an attack record and the total number of normal records.

$$FPR = \frac{FP}{FP + TN}. \quad (7)$$

#### 4.3 Intrusion Detection Binary Classification Result

Sensors and control equipment in smart agriculture frequently possess limited computing capacity, necessitating the accurate differentiation between attack behaviour and normal records. A binary classification model for normal attacks must therefore be developed. This investigation proceeds by following the outlined steps in Section 3 to train and test the normal-attack binary classification model.



**Fig. 4.** Confusion matrix of CatBoost Intrusion Detection Result

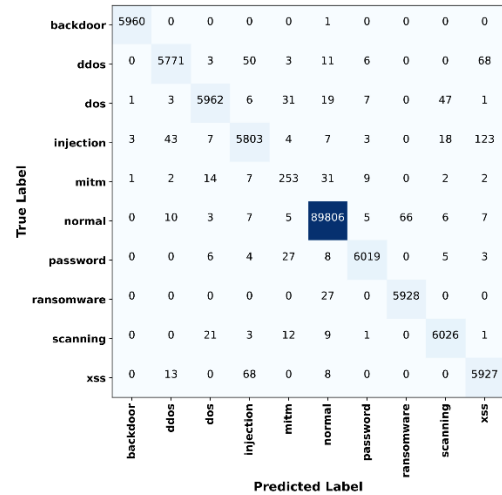
Fig. 4 shows the classification results of our intrusion detection system on the test set. Table 2 shows the comparison of the test results of the CatBoost model and the results of various other models on the ToN\_IoT network dataset. The results show that the CatBoost model achieved an accuracy of 0.9994, a Precision of 0.9988, a recall rate of 0.9994, an F1 of 0.9991, and a FPR of 0.0007 on the normal-attack binary classification task. CatBoost significantly surpasses other models. It can not only accurately identify attack behaviors, but also effectively avoid misidentification.

**Table 2.** Comparison of binary classification results between CatBoost and other methods.

Method	Accuracy	Precision	Recall	F1	FPR
Logistic Regression[23]	0.867	0.781	0.860	0.818	0.130
Naive Bayes[23]	0.464	0.395	0.998	0.566	0.882
Decision Tree[23]	0.980	0.960	0.982	0.971	0.022
Random Forest[27]	0.9749	-	-	0.99	0.029
kNN[23]	0.988	0.986	0.979	0.983	0.007
SVM[23]	0.868	0.784	0.859	0.820	0.127
AdaBoost[23]	0.909	0.827	0.935	0.878	0.105
XGBoost[23]	0.991	0.984	0.991	0.987	0.009
E-GraphSAGE[28]	0.9787	1	0.9786	0.9892	0.0192
<b>CatBoost</b>	<b>0.9994</b>	<b>0.9988</b>	<b>0.9994</b>	<b>0.9991</b>	<b>0.0007</b>

#### 4.4 Attack Multiclassification Result

As introduced in Section 4.1, the ToN\_IoT dataset contains nine different attack types, covering almost all common attack methods. As can be seen from Table 1, the ToN\_IoT dataset has a serious class imbalance problem, because the records in the dataset is collected using real equipment, which is more consistent with the reality. This study followed the steps described in Section 3 to train and test the attack type classification task.



**Fig. 5.** Confusion matrix of CatBoost Attack Type Identification Result

Fig. 5 shows the multiclass classification results of our intrusion detection system on the test set. Table 3 shows the comparison of the test results of the CatBoost model and the multiclass classification results of various other models on the ToN\_IoT network dataset. The results show that the CatBoost model achieved an accuracy of 0.9938, a Precision of 0.9938, a recall rate of 0.9938, an F1 of 0.9938, and a FPR of 0.0008 on the attack type classification task. From the above results, we noticed that the class imbalance problem did not have a great impact on the overall effectiveness of the model. CatBoost handles the class imbalance problem extremely well and achieves high recognition accuracy and low misrecognition rate.

**Table 3.** Comparison of multiclass classification results between CatBoost and other methods

Method	Accuracy	Precision	Recall	F1	FPR
Logistic Regression[23]	0.777	0.777	0.777	0.777	0.046
Naive Bayes[23]	0.712	0.712	0.712	0.712	0.136
Decision Tree[23]	0.934	0.934	0.934	0.934	0.022
Random Forest[27]	0.937	0.937	0.937	0.937	0.021
kNN[23]	0.979	0.979	0.979	0.979	0.009
SVM[23]	0.780	0.780	0.780	0.780	0.046
AdaBoost[23]	0.399	0.399	0.399	0.399	0.505
XGBoost[23]	0.983	0.983	0.983	0.983	0.008
E-GraphSAGE[28]	-	-	-	-	0.87
<b>CatBoost</b>	<b>0.9938</b>	<b>0.9938</b>	<b>0.9938</b>	<b>0.9938</b>	<b>0.0002</b>

## 5 Conclusion

Agriculture plays a crucial role in the advancement of human society. The growing population, land degradation, water scarcity, and urbanisation challenges have amplified the demand for efficient agricultural production. While smart farming delivers tremendous benefits to farmers and agricultural output, it also poses intricate cyber security risks to agricultural production. The security of the physical layer in smart agriculture is closely connected to the growth and yield of crops. Additionally, it has an indirect impact on the security of the network and application layers. This paper presents a CatBoost-based intrusion detection scheme for the physical layer, followed by the evaluation of the scheme's effectiveness using ToN\_IOT, which is a publicly available dataset. In binary classification results, this paper's scheme achieves a recognition accuracy of 99.94%, as well as a precision and recall of

99.88%. In the multi-classification results, the scheme surpasses other existing schemes in all metrics. The experimental findings indicate that the implemented method demonstrates outstanding recognition accuracy against physical layer attacks in the domain of smart agriculture. In addition, the implementation of this system ensures the security of the input data of the smart agriculture network layer, the cloud and the blockchain application.

## Acknowledgment

This study was supported by the Research and Application of Key Technologies for Blockchain-based Privacy Computing and Trusted Smart Computing project (2023KLABA03).

## References

1. A.R. De Zanella, E. Da Silva, L.C.P. Albin, *Array* **8**, 100048 (2020)
2. A. Assefa, T. Kassa, *Environmental & Socio-economic Studies* **8**, 73 (2020)
3. N. Gondchawar, R.S. Kawitkar, *IJARCCCE* **5**, 838 (2016)
4. M.A. Ferrag, L. Shu, H. Djallel, K.R. Choo, *Electronics* **10**, 1257 (2021)
5. A. Raghuvanshi, U.K. Singh, G.S. Sajja, H. Pallathadka, E. Asenso, M. Kamal, A. Singh, K. Phasinam, *Journal of Food Quality* **2022**, 1 (2022)
6. K. Kethineni, G. Pradeepini, *Cluster Computing*, 1 (2023)
7. R.A. Ramadan, A.H. Emar, M. Al-Sarem, M. Elhamahmy, *Electronics* **10**, 2633 (2021)
8. M.P. Arthur, Detecting signal spoofing and jamming attacks in UAV networks using a lightweight IDS, in 2019 International Conference on Computer, Information and Telecommunication Systems, CITS, 1 (2019)
9. V.U. Ihekoronye, S.O. Ajakwe, D.S. Kim, J.M. Lee, Hierarchical intrusion detection system for secured military drone network: A perspicacious approach, in MILCOM 2022-2022 IEEE Military Communications Conference, MILCOM, 336 (2022)
10. V. Subbarayalu, M.A. Vensuslaus, *Drones* **7**, 248 (2023)
11. M.A. Ferrag, L. Maglaras, *Computers* **8**, 58 (2019)
12. A. Heidari, N.J. Navimipour, M. Unal, *IEEE Internet Things J.* (2023)
13. H.M. Song, J. Woo, H.K. Kim, *Veh. Commun.* **21**, 100198 (2020)
14. F.V. Wyk, Y. Wang, A. Khojandi, N. Masoud, *IEEE Trans. Intell. Transp. Syst.* **21**, 1264 (2019)
15. M. Almiani, A. AbuGhazleh, A. Al-Rahayfeh, S. Atiewi, A. Razaque, *Simul. Model. Pract. Th.* **101**, 102031 (2020)
16. D. Li, L. Deng, B.B. Gupta, H. Wang, C. Choi, *Inform. Sciences.* **479**, 432 (2019)
17. S.A. Aljawarneh, R. Vangipuram, J. Supercomputing. **76**, 4376 (2020)
18. F. Jiang, Y. Fu, B.B. Gupta, Y. Liang, S. Rho, F. Lou, F. Meng, Z. Tian, *IEEE Trans. Serv. Comput.* **5**, 204 (2018)
19. S. Murali, A. Jamalipour, *IEEE Internet Things J.* **7**, 379 (2019)
20. G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.Y. Liu, *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*, *Advances in neural information processing systems* **30** (2017)
21. A.V. Dorogush, V. Ershov, A. Gulin, *arXiv preprint* (2018)
22. N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, *J. Artif. Int. Res.* **16**, 321 (2002)
23. A.R. Gad, A.A. Nashat, T.M. Barkat, *IEEE Access* **9**, 142206 (2021)
24. S.M. Lundberg, S.I. Lee, *Advances in Neural Information Processing Systems* **30**, 4765(2017)
25. I. Guyon, J. Weston, S. Barnhill, V. Vapnik, *Mach. Learn.* **46**, 38 (2002)
26. N. Moustafa, *Sustain. Cities. Soc.* **72**, 102994 (2021)
27. M. Sarhan, S. Layeghy, M. Portmann, *arXiv preprint* (2022)
28. W.W. Lo, S. Layeghy, M. Sarhan, M. Gallagher, M. Portmann, E-graphsage: A graph neural network based intrusion detection system for iot, in NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium, 1 (2022)