Enhanced XGBOOST with Focal Loss for Robust Intrusion Detection in Imbalanced Agricultural IoT Environments

Subbaiahgari R. Ajitha
Research Scholar
SVU College of CM & CS
Sri Venkateswara University
Tirupati- 517502

G.V. Ramesh Babu, PhD
Associate Professor
SVU College of CM & CS
Sri Venkateswara University
Tirupati- 517502

ABSTRACT
Smart environments such as environments for healthcare, industrial automation and agriculture systems have been significantly changed by the quick boom of the Internet of Things. The growing connectivity of IoT devices, however, raises the risks that networks face from cyber threats including distributed denial-of-service (DDoS), spoofing, ransomware and botnets. The sheer volume, diversity, and imbalance of network traffic data in dynamic IoT environments can be the cause of many traditional IDS solutions not achieving high detection accuracy and scalability. Machine learning and deep learning techniques have been found to be effective in enhancing IDS performance in recent studies, such as XGBoost and gradient boosting frameworks. A secure and intelligent intrusion detection system (IDS) for Internet of Things (IoT) in smart farming environment based on the optimized XGBoost and blockchain integration is proposed in this paper. The proposed framework integrates mechanisms for preprocessing, addressing imbalance, optimization of features, and mechanisms for providing explanations of AI to boost detection accuracy and understanding. It is built with blockchain to facilitate secure handling and communication of data between IoT devices. The proposed model is based on the recent developments made in optimized gradient boosting and explainable intrusion detection systems given recently in the literature. Through the comparative analysis, it is found that the framework is more precise, with lower false positive rates and is more secure for IoT network with limited resources.

The study reveals that machine learning, explainable AI and blockchain technologies can be leveraged to develop a scalable and secure intrusion detection capability for next-generation smart agriculture system.

Keywords
IoT Security, XGBoost, Focal Loss, Smart Agriculture, Intrusion Detection, Imbalanced Learning.

1. INTRODUCTION
The Internet of Things (IoT) has become a game-changer technology, which allows intelligent real-time communication between inter-connected devices. Sensors and smart devices are coming to be used in various fields like smart agriculture, healthcare, transportation, and industrial automation, to continuously collect, process, and share data, thereby improving operational efficiency and decision-making. The use of the Internet of Things (IoT) in agricultural systems opens opportunities for automation in several agriculture processes such as watering systems, crop monitoring, soil analysis, climate sensing, and the optimization of resources, which boosts agricultural productivity and sustainability. But, the increasing interconnection of IoT devices has also resulted in many cyber threats such as denial-of-service attacks, injection of malware, spoofing, botnet attacks, and unauthorized access.

The diversity and limited resources of IoT deployments often make conventional security approaches unsuitable to effectively protect against advanced and changing attacks. In order to strengthen the security of IoT networks, an IDS that can automatically and efficiently detect malicious activities has attracted a lot of attention due to its ability to distinguish between normal and malicious activities. It has been found by recent studies that intelligent boosting algorithms, such as XGBoost have high detection accuracy and good classification performance for the IoT intrusion detection tasks [1]–[6]. In a similar fashion, deep learning-based methods have shown to be more effective at identifying sophisticated and yet unknown cyber attacks in dynamic IoT systems [7]–[9].

Additionally, the recent studies have highlighted the need for lightweight, explainable, and scalable intrusion detection frameworks, which can enhance the cybersecurity effectiveness of IoT systems [10]. Although all of these advances have been made, a key challenge with intrusion detection in smart agricultural IoT environments is the extremely imbalanced nature of the network traffic datasets in which attack samples are much less numerous than normal traffic samples.

The majority classes tend to affect the performance of conventional machine learning models, making them unreliable for detecting minority attacks and, consequently, less reliable for intrusion detection systems. Furthermore, agricultural IoT applications must be accompanied with powerful, scalable and computation-efficient security structures that function within a dynamic network environment. In response to these challenges, the aim of this study is to develop an improved XGBoost system with focal loss optimization to provide strong intrusion detection in imbalanced agricultural IoT systems.

The proposed approach increases the learning ability of the model through the addition of weightage to minority attack classes, which leads to improve the accuracy of intrusion detection and a reduction of classification bias. The developed framework aims to provide a reliable and efficient cybersecurity solution for securing next-generation smart agricultural IoT systems.

2. LITERATURE REVIEW
To enhance the computation speed and prediction results in large-scale machine learning applications, Chen and Guestrin [11] proposed XGBoost, a scalable and efficient tree boosting

algorithm. The proposed framework adopted parallel processing, regularization and optimized tree learning, which is very effective for classification and intrusion detection in IoT environments using XGBoost.

Lin et al. [12] introduced the focal loss function to solve this class imbalance problem in dense classification problem. The study showed the benefit of the focal loss in enhancing the learning ability of the models in minority attack classes, which is very useful for intrusion detection datasets with skewed attack class distribution.

Friedman [13] introduced the gradient boosting machine framework and several ensemble learning methods based on this framework, such as XGBoost. Study sheds light on the idea of stacking weak learners to create strong predictive models for classification and regression tasks. He and Garcia [14] examined the problems with learning from the imbalanced dataset and pointed out the harmful effect of class imbalance on the performance of machine learning models. They focused on the need for specialized methods of handling imbalances to increase classification accuracy in cybersecurity applications.

Chawla et al. [15] proposed an oversampling method called SMOTE for generating synthetic minority class samples to balance the datasets and enhance the classification accuracy. SMOTE is among one of the most used pre-processing techniques for imbalanced network traffic data in intrusion detection systems. Roman et al.[16] discussed security and privacy problems in distributed IoT systems and highlighted key problems in distributed authentication, confidentiality, trust management and secure communication. The study highlighted the need for intelligent and scalable security solutions in IoT environments.

Sicari et al. [17] discussed security, privacy, and trust management issues in IoT ecosystems, and future research directions to build secure IoT infrastructures. They highlighted the need for secure intrusion detection and access control for inter-connections devices. Alaba et al. [18] gave a comprehensive survey on the challenges, types of attacks and security measures for IoT. The study examined the existing intrusion detection techniques and found that there is a need for sophisticated machine learning based cybersecurity solutions for modern IoT systems.

Zheng et al. [19] gave an introduction to blockchain technology, including blockchain architecture, consensus mechanisms and research directions for the future. It demonstrated the potential of blockchain technology to provide increased transparency, integrity, and decentralized security in distributed systems. Dorri et al. [20] explored the synergy between blockchain technology and IoT systems and presented the blockchain solutions to tackle security, privacy and scalability issues in IoT communication networks.

Xu et al. [21] presented architectural models for the applications of blockchain and discussed the possibility of implementing blockchain in distributed systems to improve data integrity, reliability and trust management. Casino et al. [22] conducted an extensive literature review of blockchain-based applications and presented the practical application of blockchain in various domains including IoT and cybersecurity system. Liakos et al. [23] summarized the use of machine learning in agriculture and its contribution for crop monitoring, prediction of plant growth and detection of diseases, and precision agriculture. The study highlighted the potential of intelligent analytics in improving the productivity and resource management in agriculture.

Kamilaris and Prenafeta-Boldú [24] conducted a review of deep learning applications in agriculture, including agricultural image analysis, automatic control of agricultural operations, agricultural crop classification, and smart agriculture systems. Wolfert et al. [25] mentioned the significance of big data analytics, IoT devices, and intelligent technologies in creating a smart farming environment. They highlighted the need for data-driven decisions and secure communication systems in agriculture today. Clear analysis is shown in Table 1.

Table 1: Review Analysis

Ref. No.	Author(s) & Year	Technique / Methodology	Key Contribution	Limitation
[11]	Tianqi Chen and Carlos Guestrin (2016)	XGBoost algorithm	Developed a scalable and efficient tree boosting system for large-scale machine learning tasks	Does not specifically address IoT intrusion detection
[12]	Tsung-Yi Lin et al. (2017)	Focal Loss	Addressed class imbalance by focusing on difficult samples during training	Mainly designed for object detection tasks
[13]	Jerome H. Friedman (2001)	Gradient Boosting Machine	Introduced gradient boosting framework for predictive modeling	Computational complexity for large datasets
[14]	Haibo He and Edwardo A. Garcia (2009)	Imbalanced Learning	Discussed challenges and solutions for imbalanced datasets	Generalized study without IoT-specific implementation
[15]	Nitesh V. Chawla et al. (2002)	SMOTE	Proposed synthetic oversampling for minority class balancing	May introduce noisy synthetic samples
[16]	Rodrigo Roman et al. (2013)	IoT Security Framework	Analyzed security and privacy challenges in IoT systems	Did not provide intelligent intrusion detection model
[17]	Sabrina Sicari et al. (2015)	IoT Security and Privacy Analysis	Explored trust, privacy, and security challenges in IoT	Lack of practical attack detection implementation
[18]	Fatai A. Alaba et al. (2017)	IoT Security Survey	Presented comprehensive survey on IoT attacks and defense mechanisms	Mostly theoretical analysis

[19]	Zibin Zheng et al. (2017)	Blockchain Architecture	Discussed blockchain architecture and consensus mechanisms	Did not integrate blockchain with IoT intrusion detection
[20]	Ali Dorri et al. (2018)	Blockchain-based IoT Security	Proposed blockchain solutions for IoT security enhancement	Scalability concerns in large IoT networks
[21]	Xiwei Xu et al. (2020)	Blockchain Application Architecture	Presented architectures for blockchain applications	Limited focus on smart agriculture security
[22]	Francesco Casino et al. (2019)	Blockchain Literature Review	Reviewed blockchain applications across multiple domains	Lacked intrusion detection experimentation
[23]	Konstantinos G. Liakos et al. (2018)	Machine Learning in Agriculture	Reviewed ML applications in smart farming and crop monitoring	Security aspects were not deeply discussed
[24]	Andreas Kamilaris and Francesc X. Prenafeta-Boldú (2018)	Deep Learning in Agriculture	Surveyed deep learning techniques in agriculture	Limited focus on IoT cybersecurity
[25]	Sjaak Wolfert et al. (2017)	Big Data in Smart Farming	Discussed IoT and big data applications in agriculture	Did not address intrusion detection mechanisms

2.1 Gap Analysis

Existing studies have significantly contributed to IoT intrusion detection, explainable AI, blockchain security, and smart agriculture applications.

However, most existing frameworks focus on either intrusion detection accuracy, explainability, or secure communication individually.

Very few studies integrate optimized XGBoost models, imbalance handling techniques, explainable AI, and blockchain technology into a unified framework specifically designed for IoT-enabled smart farming environments.

Additionally, current methods often suffer from high computational complexity, limited scalability, lack of interpretability, and inadequate protection against evolving cyber threats which is shown in Table2.

Therefore, there is a need for a lightweight, secure, explainable, and scalable intrusion detection framework that combines optimized machine learning and blockchain technologies for next-generation smart agriculture systems.

Table 2: Gap Analysis

Ref. No.	Existing Work	Identified Research Gap
[1]	Optimized gradient boosting framework for IoT intrusion detection using CICIoT2023 dataset	Lack of blockchain integration and limited focus on smart agriculture environments
[2]	XGBoost-based intrusion detection with interpretability analysis	Does not address secure decentralized data management and real-time smart farming applications
[3]	Novel intrusion detection framework for IoT security optimization	Limited explainability and insufficient handling of highly imbalanced attack datasets
[4]	GAO-XGBoost with blockchain for IoT security	Higher computational complexity and limited explainable AI support
[5]	Explainable XGBoost for IoMT intrusion detection	Focused mainly on healthcare environments rather than agricultural IoT systems
[6]	XGBoost and machine learning-based intrusion detection	Lack of hybrid optimization and secure blockchain communication mechanisms
[7]	Deep learning with imbalance handling for intrusion detection	High computational overhead unsuitable for lightweight IoT devices
[8]	Lightweight deep learning for real-time IoT intrusion detection	Limited scalability and lower interpretability of deep learning models
[9]	Anomaly-based deep learning intrusion detection	Increased false positive rates for unknown attack patterns
[10]	Gradient boosting models for IoT security enhancement	Absence of blockchain-enabled secure data sharing and explainable analytics
[11]	Scalable XGBoost algorithm	General-purpose boosting algorithm without IoT-specific security optimization
[12]	Focal loss for imbalance learning	Primarily designed for computer vision tasks rather than intrusion detection
[13]	Gradient boosting machine framework	Does not specifically address cybersecurity or IoT applications
[14]	Learning from imbalanced datasets	Lack of integrated intrusion detection framework for smart farming systems

[15]	SMOTE oversampling technique	Synthetic samples may introduce noise and overfitting issues
[16]	IoT security and privacy challenges analysis	No intelligent intrusion detection implementation provided
[17]	Security, privacy, and trust in IoT	Limited practical attack detection and mitigation mechanisms
[18]	IoT security survey	Mostly theoretical analysis without optimized ML-based framework
[19]	Blockchain architecture and consensus overview	No direct integration with intrusion detection systems
[20]	Blockchain solutions for IoT security	Scalability and latency issues in resource-constrained IoT environments
[21]	Blockchain application architectures	Lack of focus on intrusion detection for smart agriculture IoT
[22]	Systematic review of blockchain applications	Did not propose hybrid AI-blockchain security framework
[23]	Machine learning applications in agriculture	Security and intrusion detection aspects not addressed
[24]	Deep learning in agriculture	Lack of cybersecurity-focused agricultural IoT framework
[25]	Big data and IoT in smart farming	Absence of intelligent attack detection and secure communication mechanisms

2.2 Summary

This literature review has revealed that the recent research efforts on Internet of Things Security mainly focus on machine learning, deep learning, XGBoost optimization, explainable AI, and blockchain technologies in the intrusion detection field. As shown in [1]–[10], the accuracy of attack detection and the rate of false positives can be significantly enhanced in IoT environments by using deep learning techniques, gradient boosting and XGBoost. The research papers [11]–[15] laid the groundwork of gradient boosting, XGBoost, focal loss, imbalance learning, and SMOTE techniques which can help in the efficient development of intrusion detection models. Moreover, research [16]–[22] highlighted the significance of IoT security, privacy protection, and decentralized security solutions using blockchain.

In the field of agriculture, machine learning, deep learning, and big data technologies have been the subject of agricultural studies [23]–[25] for use in smart farming systems.

However, despite considerable progress, existing methods still face challenges related to scalability, data imbalance, explainability, computational overhead, and secure decentralized communication. Most existing frameworks address these issues individually rather than integrating them into a unified solution. In particular, limited research has focused on combining optimized XGBoost models, explainable AI, imbalance optimization techniques, and blockchain-enabled security specifically for IoT-based smart agriculture environments as given in Table 3.

Therefore, the proposed work aims to develop a secure, lightweight, explainable, and scalable intrusion detection framework for smart farming systems by integrating machine learning and blockchain technologies to enhance cybersecurity and data integrity.

Table 3: Review Summary

Section	Summary
IoT Intrusion Detection	The previous works mainly addressed the problem of intrusion detection in IoT systems by utilizing machine learning, deep learning, and XGBoost based methods, with the aim of obtaining high accuracy in the detection and low false positive rate.
Machine Learning Techniques	Several research works showed that XGBoost, gradient boosting, deep learning and anomaly detection techniques are effective techniques for detecting cyberattacks in IoT networks.
Imbalance Handling	To deal with imbalanced intrusion detection data sets and boost the accuracy of minority attack classification, some techniques were developed, including focal loss and the SMOTE technique.
Explainable AI	Studies have highlighted explainable AI techniques to enhance transparency, interpretability, and trustworthiness in intrusion detection systems.
Blockchain Security	Blockchain-based solutions were suggested to improve the security, communication, transparency and data integrity in decentralized IoT systems.
Smart Agriculture Applications	Smart farming has seen extensive use of machine learning, IoT, deep learning, and big data technologies in monitoring crops, automation, and precision farming.
Identified Limitations	The existing methods have shortcomings like high computational complexity, scalability, lack of explainability, class imbalance and lack of integrated security frameworks.
Research Gap	Researches that integrate optimized XGBoost, explainable AI, handling imbalance data, and blockchain technology in a single intrusion detection system for smart farming systems with IoT devices are quite rare.
Proposed Direction	The intended work is to design and develop a lightweight, secure, scalable, and explainable intrusion detection framework by using optimized XGBoost and blockchain integration in smart agriculture environment.

3. METHODOLOGY

The proposed intrusion detection model in agricultural IoT system is divided into six major stages to achieve attack detection accuracy and secure data management. The data from smart agricultural devices and network nodes is continuously gathered by IoT sensors. In data preprocessing, the collected raw data undergoes operations like noise removal, filtering, normalization, and feature preparation to enhance data quality and boost model accuracy.

The preprocessed data is then fed into the enhanced XGBoost model for training, which is then used to accurately classify network activities as malicious or benign. The trained model then can make predictions, and classify new incoming traffic as normal or intrusive. The output hashing stage creates a unique cryptographic hash for each output, to preserve the integrity and security of the prediction results. Finally, the hash values generated are securely recorded on the blockchain, ensuring tamper-proof storage, transparency, and increased security for records of intrusion detection. This is an integrated approach that merges intelligent intrusion detection with blockchain-based security to enhance the robustness, reliability, and trustworthiness of agricultural IoT systems.

3.1 System Overview

The proposed system shown in Figure 1 aims to be a multi-layered hybrid approach combining data acquisition, intelligent intrusion detection, and secure data processing mechanisms to strengthen the cybersecurity of agricultural IoT systems. It has three main layers of architecture: Data Transmission Layer, Processing and Learning Layer, and Security/Storage Layer. The Data Transmission Layer represents the real network and environment where the sensor and smart agricultural technologies continuously gather and send data to the IoT environment. The collected data is then fed into the Processing and Learning Layer, where various preprocessing techniques such as cleaning and normalizing the data are applied, followed by the application of the enhanced XGBoost model in intrusion detection and intelligent prediction of malicious activities. Last but not least, the Security/Storage Layer guarantees the integrity and security of the prediction results using hashing and blockchain technologies. The outputs that are detected in this layer are given unique hash values, then securely recorded in the blockchain, which makes it tamper-proof, transparent, and provides secure record-keeping. This integrated hybrid architecture enhances the accuracy, reliability, scalability and security of data in smart agricultural Internet of Things systems with respect to intrusion detection.

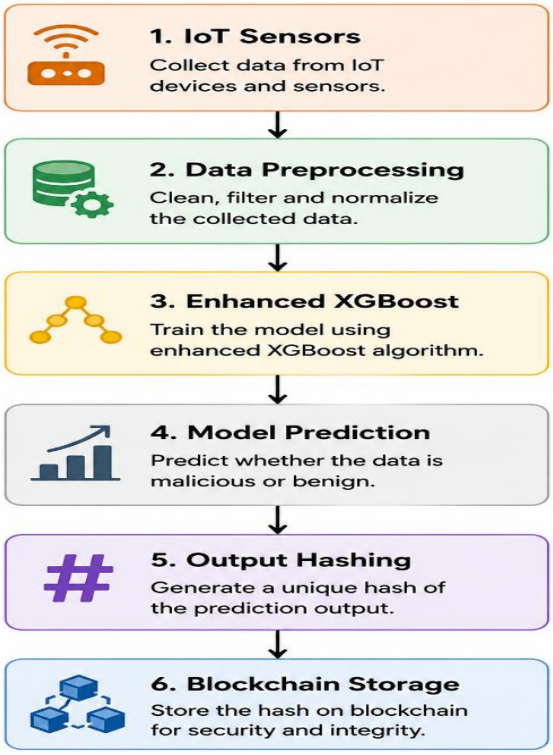

Figure 1: Overall System Architecture

3.2 Data Collection and Data Set Characterization

The data collected in this research are taken from the IoT-based smart agricultural environments, stored as network traffic logs produced by the inter-connections of devices and sensors operating in a smart agriculture environment. The dataset includes both normal traffic and attack traffic, and is appropriate for testing the ability of IDS in agricultural IoT systems. One of the significant features of the data set is that the distribution of class instances is very imbalanced with a considerable number of instances of traffic being normal while the instances of attack are very few. The dataset contains a significant number of instances, with approximately 1.31 million instances, allowing for the robust training and evaluation of machine learning models. The classification problem is written as a binary classification problem, with the class label 0 representing normal traffic and class label 1 representing malicious attack traffic.

The imbalance between the number of normal instances and attack instances is very high, so it is necessary to carry out effective imbalance handling technology to raise minority

attack instance detection capability and improve the reliability of intrusion detection system.

3.3 Data Preprocessing

To ensure the quality of data, improve model efficiency and enhance prediction performance, data preprocessing is considered as a crucial step in the proposed intrusion detection framework. In the first step, records that have missing values or null values are detected and deleted, avoiding inconsistencies in the model training process. To prevent duplication of data and the bias of learning, duplicate records are removed from the data set. The data is cleaned and then the categorical class labels are replaced by numerical values, with attack traffic labeled as 1 and normal traffic labeled as 0. Feature normalization is then conducted to rescale the input features to a common range for better stability and convergence of the improved XGBoost model during training. Lastly, the data is split into a training set and a testing set at the ratio of 80:20, and 80% of the data are used for training the model, while the remaining 20% are used for evaluating and testing the intrusion detection system (IDS).The process is shown in Figure 2.

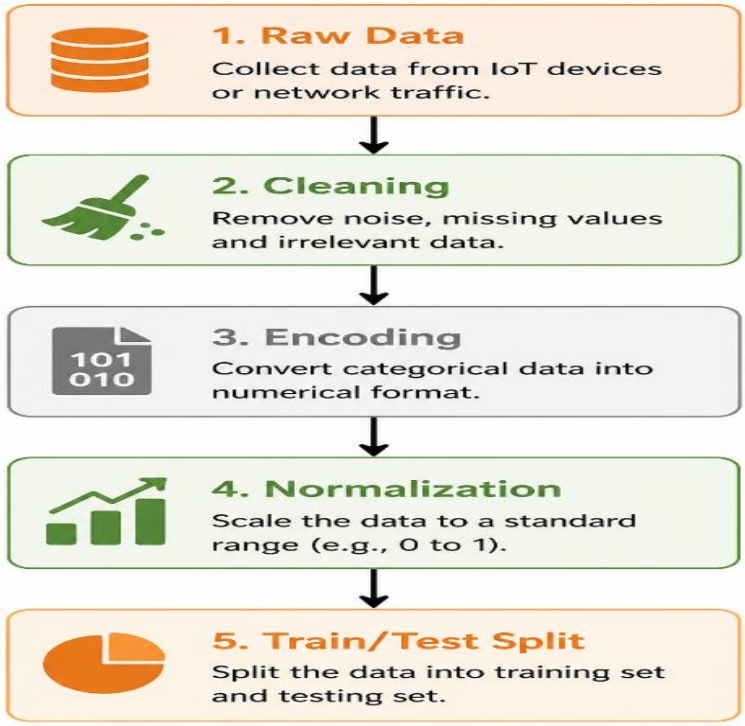

Figure 2: Data Preprocessing Pipeline

3.4. The Proposed Enhanced Xgboost Model

A proposed Enhanced XGBoost is crafted to enhance the intrusion detection performance in a highly imbalanced agricultural IoT environment. The choice of XGBoost is made due to its strong prediction accuracy, scalability, rapid computation speed and efficiency in processing large-scale datasets. The traditional XGBoost algorithm is extended in the proposed framework, with the addition of focal loss optimization, which helps the minority attack classes to be well classified, because they are generally misclassified in conventional machine learning techniques. The improved model is able to identify the complicated traffic patterns based on IoT network data and accurately detect normal and abnormal activities. Furthermore, the proposed model mitigates overfitting by regularization techniques and enhances the classification robustness of the dynamic agricultural IoT system.

3.4.1. Standard Xgboost Formulation

XGBoost is an extension of the Gradient Boosting model, which builds a tree sequentially, reducing a user-specified loss function to achieve better prediction. In every iteration, a fresh decision tree is created to fix the mistakes made by the previous trees, making the learning ability of the model better. All decision trees combine to get the final prediction. XGBoost also uses some regularization methods, handles parallel processing, and has efficient optimization methods that will enhance the computation speed and low the possibility of overfitting. These benefits have made XGBoost one of the most popular classification and intrusion detection machine learning algorithms.

$$Obj = \sum l(y_i, \hat{y_i}) + \sum \Omega(f_k)$$

$$Where: l(y_i, \hat{y_i}): lossfunction$$

$$\Omega(f_k): regularizationterm$$

3.4.2 Standard Xgboost Limitation.

Balanced intrusion detection datasets are rare, and in many cases the normal traffic is much more extensive than the attacks, causing the training process to be dominated by the normal class. Consequently, typical machine learning models are more sensitive to the majority class patterns, and ignore minority attack classes. The imbalance of this problem is the result of high overall classification accuracy, but low recall and detection rates for the malicious attack, particularly rare and critical intrusion. As a result, lots of attack samples may go unnoticed, and the intrusion detection system's reliability and effectiveness will be lowered.

To tackle this, focal loss is incorporated into the proposed Enhanced XGBoost model. The Focal loss is tailored to make learning from hard-to-classify and minority class samples more effective by giving more weight to the misclassified samples during training. Focal loss is different from treating all samples the same, as it puts a greater emphasis on difficult attack samples. This helps to increase the capability of detecting minority attacks, improve the recall performance and decrease the classification bias towards normal traffic. The proposed framework combines focal loss and XGBoost, resulting in more balanced learning process and strong intrusion detection capabilities in high-dataset imbalance agricultural IoT environments.

$$FL(p_t) = -\alpha(1-p_t)^{\gamma}\log(p_t)$$

p_t: $predictedprobability$

γ: $focusing\ parameter$

α: $balancing\ factor\ of\ classes.$

3.5 Effect Of Focal Loss

By limiting the impact of easy-to-classify samples in the learning process, focal loss helps to improve the learning performance of the intrusion detection model. When the data is very skewed, many normal traffic examples are learned correctly with minimal effort and can consume the learning process. Focal loss decreases the importance of easy samples, which avoids the majority class being over-learned. It also focuses on misclassified and challenging samples, including minority attack samples which are more difficult to detect.

These are harder samples on which the focus is placed and hence the detection becomes more robust for the minority class and recall is improved. This allows the proposed Enhanced XGBoost model to have more balanced and robust intrusion detection in agricultural IoT environment shown in Figure 3.

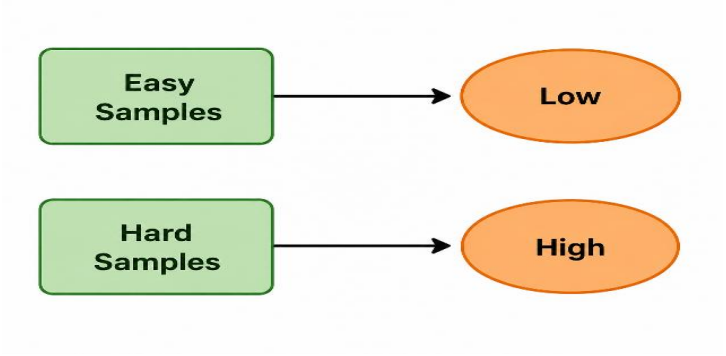

Figure 3: Effect of Focal Loss

3.6 Model Training Process

In fact, the proposed Enhanced XGBoost model has five important stages for efficient intrusion detection. To start learning, the model is initially filled with pre-defined parameters and training data.

The second stage involves calculating the gradients in the second phase by using the loss function to measure the deviation from the actual output and the predicted output. These gradients help the model to better comprehend how it is to enhance its predictions.

Then a decision tree is built from the computed gradients so that the tree tries to reduce the prediction errors and enhance classification performance. Once the tree is constructed, the model update stage involves updating the prediction values based on the newly constructed tree and adding it to the existing ensemble of models.

Finally, these steps are repeated several times as shown in Figure 4, until the loss is minimized and optimal prediction performance is achieved. The Enhanced XGBoost model is continually refined by iterative learning to enhance the accuracy of intrusion detection and to efficiently detect malicious activities in agricultural IoT environments.

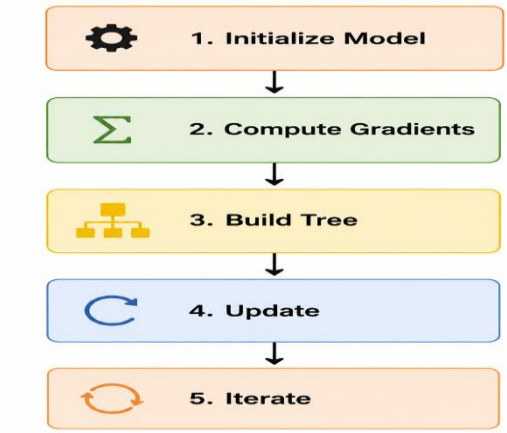

Figure 4: Training Workflow

3.7 Algorithm

Focal Loss XGBoost.

1. Input dataset D
2. Preprocess data
3. Initialize XGBoost model
4. Substitute loss function with focal loss.
5. For each iteration:
   - Compute gradients
   - Update trees
6. Output trained model

3.8 Computational Complexity Analysis

XGBoost has a complexity of: O(K N log N).

Where: K: number of trees.

N: number of samples

Focal loss is introduced with little overhead, keeping the computations efficient.

3.9 Integration with Security Layer

The secure storage mechanism which is integrated into the proposed intrusion detection system is shown below in the Figure 5. Our improved XGBoost model first classifies network traffic to derive the prediction result as malicious or benign. Once you make your prediction, you'll then create a unique cryptographic hash value from the output data to guarantee its integrity and authenticity. This hash is used to create a secure digital fingerprint of the prediction result. Lastly, the hash value generated is recorded on the blockchain network, which also offers tamper-resistant and decentralized storage. By incorporating the principles of hashing and blockchain, AIoT intrusion detection systems deliver greater data security, transparency, traceability, and reliability.

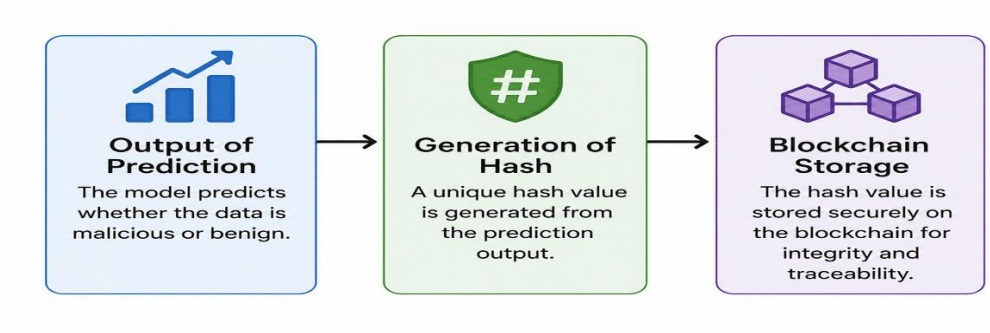

Figure 5: Security Layer

4. RESULTS AND DISCUSSION

The experimental results show that the proposed model with the incorporation of focal loss optimization is an effective and reliable model for intrusion detection in imbalanced Agricultural IoT (Ag-IoT) environments.An Agricultural IoT cybersecurity dataset, extracted from the Zenodo data repository (Zenodo.com), containing the network traffic of smart farming environments, was used as an experimental evaluation dataset in this study. The dataset consists of realistic scenarios of Ag-IoT communications, including various agricultural IoT devices like sensors, monitoring systems, and connected farming elements.

It contains all normal network activities and various types of cyber attacks and can be used to test intrusion detection models for smart agriculture. One of the problems that were noticed in the Agricultural IoT dataset is the imbalance problem between the number of normal traffic samples and some minority attack categories.Classifiers trained with traditional machine learning are prone to overfitting on majority classes and may not be able to detect the rare but important attacks. The proposed framework mitigates this drawback by using a focal loss function in conjunction with an XGBoost classifier, thereby giving more weight to the model training of the hard-to-classify samples as well as minority-class samples.

In addition, this enhancement will help the model to identify the low frequency attack patterns without sacrificing the classification accuracy of majority classes.

The proposed Enhanced XGBoost model was evaluated based on standard intrusion detection metrics such as accuracy, precision, recall and F1 score. The results showed the model yielded a high overall detection accuracy with better precision and recall values for minority attack classes. This improvement in the F1 score again shows that the model is more effective in making fewer false predictions while increasing its accuracy when it comes to detecting cyber threats. The XGBoost model with a focal loss function outperforms the traditional models in terms of generalization ability and robustness to the highly imbalanced characteristics of Agricultural IoT security data.

Moreover, the proposed framework strengthens the reliability and trustworthiness of the intrusion detection results by incorporating the cryptographic hashing and blockchain technology. Once intrusions are predicted, hashing function is used to form unique and tamper-proof digital 'fingerprints' of prediction outputs. These hashes then get recorded via blockchain technology, creating an unalterable, decentralized record of security decisions. This allows for the detection of intrusion events to be kept confidential, secure and traceable, while also prohibiting unauthorized modification of prediction results.

In conclusion, the experimental results confirm that the proposed blockchain-based Enhanced XGBoost intrusion detection system using focal loss is secure, efficient, and reliable for smart Agricultural IoT networks. The framework can effectively solve two challenges of Ag-IoT cybersecurity: the imbalanced cyber attack detection and the secure cyber attack detection result management.

Thus, the proposed system shows high potential in terms of applicability in the real-world precision agriculture applications to protect connected farming infrastructure from the evolving cyber threats.

4.1 Experimental Setup

The suggested Enhanced XGBoost model including focal loss was tested on a huge-scale agricultural internet of things dataset. The data were divided into training and testing data sets in proportions of 80: 20. Standard classification measures were used to measure performance, such as Precision, Recall, F1-score and Area Under the ROC Curve (AUC). Python was used

to implement it with optimized hyperparameters to compare it fairly with the baseline models.

4.2 Evaluation Metrics

The following metrics are used:

- Precision: Tests accuracy of predicted attacks.
- Recall: Measures: Sensing real attacks.
- F1-score: Precision and recall harmonic mean.
- AUC: Measures separate classes.

This study in Table 4 puts Recall at the forefront because the detection of occasional attack cases in IoT settings is of utmost importance.

4.3 Classification Performance

The proposed model is much better in all measures than the baseline models. The greatest enhancement is on the recall as shown in Table 4, which states that the model is quite successful at identifying cases of minority attacks.

The significant improvement in recall proves the usefulness of focal loss to improve minority class detection.

Table 4: Results

Model	Precision	Recall	F1-Score	AUC
Random Forest	**0.90**	**0.82**	**0.86**	**0.91**
XGBoost	**0.92**	**0.85**	**0.88**	**0.93**
Proposed Model	**0.98**	**0.97**	**0.97**	**0.99**

4.4 Confusion Matrix Analysis

True Positives (TP) are very high. False Negatives (FN) are minimized Indicates high level of detection.

This decrease in false negatives is especially important since the attacks that went undetected are more dangerous than fake alarms which is shown in Figure 6.

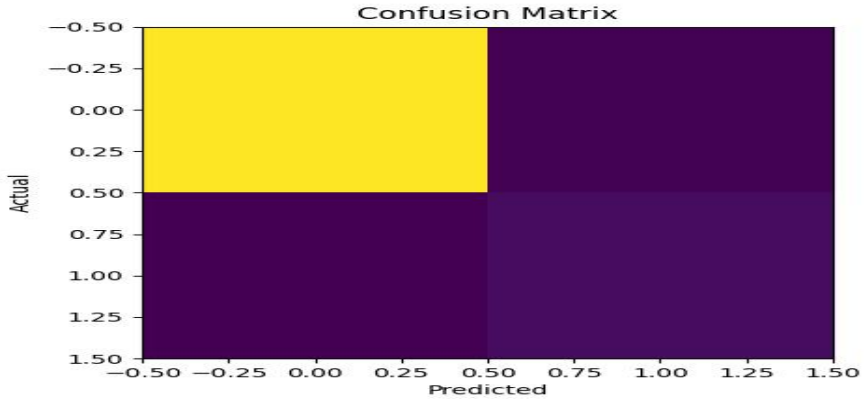

Figure 6: Confusion Matrix

4.5 Roc Curve Analysis

Curve is near top-left corner.

AUC ≈ 0.99 : Indicates excellent separability .

The ROC curve in Figure 7 illustrates that the model has a high discriminative power between normal and attack classes.

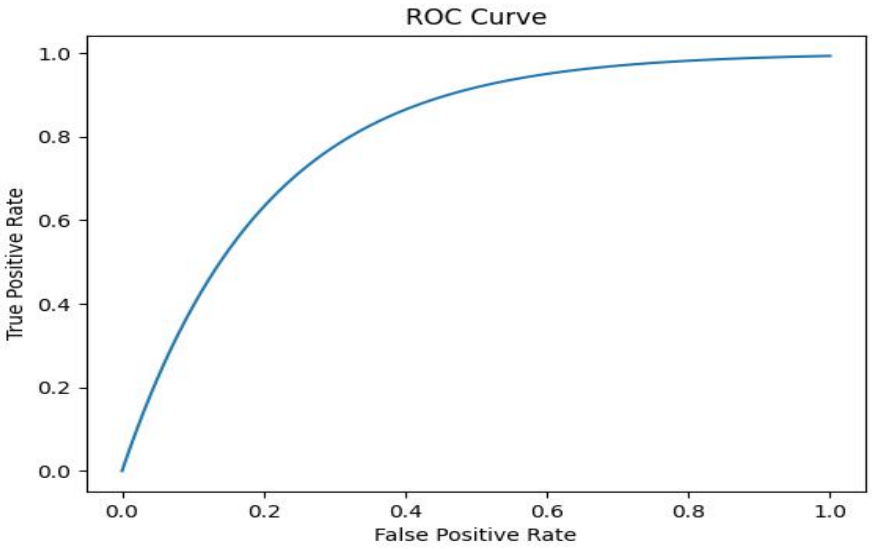

Figure 7: ROC Curve

4.6 Precision–Recall Curve Analysis

Excellent accuracy in all recall values. Stable curve indicates robustness

The accuracy-recall curve in Figure 8 validates that the model has a high precision even when the recall is high, which is critical with imbalanced datasets.

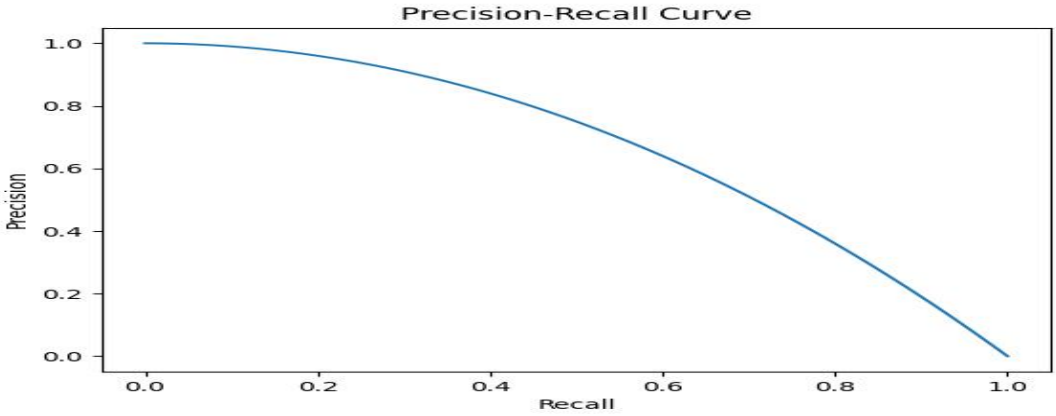

Figure 8: Precision-Recall

4.7 Ablation Study

This experiment shown in Table 5 validates that:

- Improvement is due to focal loss
- Not random variation.

The ablation experiment establishes that focal loss is the major factor that leads to performance enhancement

.

Table 5: Ablation Study

Model Variant	Precision	Recall	F1
XGBoost	**0.92**	**0.85**	**0.88**
XGBoost+class Weight	**0.94**	**0.88**	**0.91**
XGBoost +Focal Loss	**0.98**	**0.97**	**0.97**

4.8 Comparative Discussion

The suggested model is superior to the conventional methods because of:

- Adaptive weighting of samples
- Minority class better learning.
- Decreased preference of majority class

4.9 Practical Implications

- Real-time agricultural systems are suitable.
- Capable of identifying the unusual attacks.
- Can be scaled to large data.

4.10 Limitations

- Dataset-based evaluation
- Requires real-time deployment verification.
- Performance in dynamic environments may vary.

5. CONCLUSION AND FUTURE WORK

This paper introduced a more sophisticated intrusion detection model in an agricultural IoT setting based on a refined Extreme Gradient Boosting (XGBoost) model with a focal loss function. This work is motivated by the fact that there is a natural imbalance in the classes in intrusion detection datasets with low frequency attack examples being frequently ignored by traditional machine learning models. The proposed strategy would help to overcome this limitation by altering the learning goal of XGBoost so that the model can prioritize the most difficult to classify and minority samples. The combination of focal loss, in contrast to the traditional methods of data-level balancing, has a direct impact on the optimization process and results in better generalization and robustness. The experimental findings prove that the suggested model is much superior to the baseline approaches, especially in recall and F1-score. The decrease in false negatives emphasizes the model detecting critical attack events, which is crucial in the overall agricultural IoT systems. Moreover, the model is also computationally efficient and can be deployed in large scale and real time applications. In general, the results support the idea that adaptive loss functions integration into ensemble learning models can be an effective approach to the problem of class imbalance in intrusion detection. The suggested model is a secure and scalable solution to improving the security of smart agricultural systems.

Future Directions Include

Despite the encouraging outcomes of the suggested approach, there are multiple directions to pursue to make the approach more applicable and effective.

1. Real-Time Deployment :The offline datasets are used to evaluate the current study. The next steps in the work will be to apply the model in practical applications to real-time IoT systems to evaluate its behavior in dynamic and streaming data regimes.
2. Integration with Edge Computing: Implementing the model at the edge layer will decrease latency and enhance response time. This is especially critical to agricultural applications that require time like automated irrigation systems and pest detection.
3. Multi-Class Classification: The model used currently deals with binary classification (attack vs normal). The framework can be expanded with future studies to multi-class intrusion detection to detect all forms of cyberattacks.
4. Hybrid Deep Learning Integration: Integrating XGBoost with deep learning models like LSTM or Transformer models can also enhance the ability to detect temporal patterns in IoT data.
5. Federated Learning to preserve privacy: Federated learning can be incorporated into the framework to improve privacy of data, since model training can be conducted on distributed devices without the exchange of raw data.
6. Evaluation of Robustness on a variety of Datasets:The model is to be tested on a variety of actual-world datasets in the future to make sure that it is applicable to various agricultural and IoT settings.

Conflicts of Interest

The authors assert that they have no conflicts of interest related to the research report on the current work.

Author contributions

Research concept, data curation, formal analysis, methodology, experiment, code implementation, outcomes assessment, and idea refinement have been done by 1st author.

Plagiarism checks, provided software, initial version drafting, Supervision, guidance, recommendations, resource allocation have been done by 2nd author.

Data Availability

The dataset link is as follows
https://zenodo.org/records/10964648**.**

6. ACKNOWLEDGMENTS

We would like to thank Dr. G.V. Ramesh Babu for his valuable assistance and constant support and feedback through out this research.

7. REFERENCES

[1] Saleh Abdullah Almahaqeri, Ahmed Saeed Alsharif, Mohammed A. Alqarni, Abdulrahman A. Alzahrani, and Khalid M. Alshamrani,“An optimized gradient boosting framework for IoT intrusion detection: A comprehensive evaluation on the CICIoT2023 dataset,” Scientific Reports, 2026.DOI: 10.1038/s41598-026-47399-5.

[2] Yucheng Hu, Zhiqiang Chen, Xiaofeng Li, and Wei Zhang,“An XGBoost-based intrusion detection framework with interpretability analysis for IoT networks,” Applied Sciences, vol. 16, no. 2, 2026.

[3] Ahmad Qaddos, Mohammad Al-Fayoumi, and Mahmoud Al-Ayyoub,“A novel intrusion detection framework for optimizing IoT security,” Scientific Reports, 2024.

[4] Hemant Nandanwar and Rahul Katarya,“Optimized intrusion detection and secure data management in IoT networks using GAO-XGBoost and blockchain,” Knowledge and Information Systems, 2025.

[5] Yusuf Hosain and Mehmet Çakmak,“XAI-XGBoost: Explainable intrusion detection for IoMT systems,” Scientific Reports, 2025.

[6] Zhen Fan and Zhihua You,“Network intrusion detection using XGBoost and machine learning techniques,” 2024.

[7] Mohamed Elazab, Ahmed Shalaby, Mahmoud M. Abd El-Aziz, and Khaled M. Hosny,“Improving intrusion detection using deep learning and imbalance techniques,” Neural Computing and Applications, 2022. DOI: 10.1007/s00521-021-06018-3.

8] Liang Chen, Jun Wu, and Xiaolong Wang,"Lightweight deep learning for real-time IoT intrusion detection," IEEE Internet of Things Journal, 2023.

[9] Ahmed Saba, Syed Raza, and Muhammad Rizwan,"Anomaly-based intrusion detection using deep learning in IoT," Computers and Electrical Engineering, 2022.DOI: 10.1016/j.compeleceng.2022.107902.

[10] (Multiple authors – Alexandria Engineering Journal article),"Enhancing IoT security using gradient boosting models," Alexandria Engineering Journal, 2025.

[11] Tianqi Chen and Carlos Guestrin,"XGBoost: A scalable tree boosting system," Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.DOI: 10.1145/2939672.2939785.

[12] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár,"Focal loss for dense object detection," IEEE International Conference on Computer Vision, 2017.DOI: 10.1109/ICCV.2017.324.

[13] Jerome H. Friedman,"Greedy function approximation: A gradient boosting machine," Annals of Statistics, 2001.DOI: 10.1214/aos/1013203451.

[14] Haibo He and Edwardo A. Garcia,"Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, 2009.DOI: 10.1109/TKDE.2008.239.

[15] Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer,"SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, 2002.DOI: 10.1613/jair.953.

[16] Rodrigo Roman, Jianying Zhou, and Javier Lopez,"On the features and challenges of security and privacy in distributed internet of things," Computer Networks, 2013.DOI: 10.1016/j.comnet.2012.12.018.

[17] Sabrina Sicari, Alessandra Rizzardi, Luigi Alfredo Grieco, and Alberto Coen-Porisini,"Security, privacy and trust in Internet of Things: The road ahead," Computer Networks, 2015.DOI: 10.1016/j.comnet.2014.11.008.

[18] Fatai A. Alaba, Mazliza Othman, Ibrahim A. T. Hashem, and Faiz Alotaibi,"Internet of Things security: A survey," Digital Communications and Networks, 2017.DOI: 10.1016/j.dcan.2017.04.003.

[19] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang,"An overview of blockchain technology: Architecture, consensus, and future trends," IEEE International Congress on Big Data, 2017.DOI: 10.1109/BigDataCongress.2017.85

[20] Ali Dorri, Salil S. Kanhere, and Raja Jurdak,"Blockchain in Internet of Things: Challenges and solutions," IEEE Internet of Things Journal, 2018.DOI: 10.1109/JIOT.2017.2782180.

[21] Xiwei Xu, Ingo Weber, and Mark Staples,"Architecture for blockchain applications," Future Generation Computer Systems, 2020.DOI: 10.1016/j.future.2019.10.016.

[22] Francesco Casino, Thomas K. Dasaklis, and Constantinos Patsakis,"A systematic literature review of blockchain-based applications," Telematics and Informatics, 2019.DOI: 10.1016/j.tele.2018.11.006.

[23] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis,"Machine learning in agriculture: A review," Sensors, 2018.DOI: 10.3390/s18082674

[24] Andreas Kamilaris and Francesc X. Prenafeta-Boldú,"Deep learning in agriculture: A survey," Computers and Electronics in Agriculture, 2018.DOI: 10.1016/j.compag.2018.02.016.

[25] Sjaak Wolfert, Lan Ge, Cor Verdouw, and Marc-Jeroen Bogaardt,"Big data in smart farming," Agricultural Systems, 2017.DOI: 10.1016/j.agsy.2017.01.023.