

## Research Article

# Blockchain-Based Internet of Things: Machine Learning Tea Sensing Trusted Traceability System

Yuting Wu <sup>1</sup>, Xiu Jin,<sup>1</sup> Honggang Yang,<sup>1</sup> Lijing Tu,<sup>1</sup> Yong Ye,<sup>2</sup> and Shaowen Li <sup>1,2</sup>

<sup>1</sup>Anhui Province Key Laboratory of Smart Agricultural Technology and Equipment, Anhui Agriculture University, Hefei 230001, China

<sup>2</sup>School of Information and Computer Science, Anhui Agricultural University, Hefei 230001, China

Correspondence should be addressed to Shaowen Li; shaowli@ahau.edu.cn

Yuting Wu and Xiu Jin contributed equally to this work.

Received 26 September 2021; Revised 23 October 2021; Accepted 31 January 2022; Published 21 February 2022

Academic Editor: Chaolong Zhang

Copyright © 2022 Yuting Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A framework combining the Internet of Things (IoT) and blockchain can help achieve system automation and credibility, and the corresponding technologies have been applied in many industries, especially in the area of agricultural product traceability. In particular, IoT devices (radio frequency identification (RFID), geographic information system (GIS), global positioning system (GPS), etc.) can automate the collection of information pertaining to the key aspects of traceability. The data are collected and input to the blockchain system for processing, storage, and query. A distributed, decentralized, and nontamperable blockchain can ensure the security of the data entering the system. However, IoT devices may generate abnormal data in the process of data collection. In this context, it is necessary to ensure the accuracy of the source data of the traceability system. Considering the whole-process traceability chain of agricultural products, this paper analyzes the whole-process information of a tea supply chain from planting to sales, constructs the system architecture and each function, and designs and implements a machine learning- (ML-) blockchain-IoT-based tea credible traceability system (MBITTS). Based on IoT technologies such as radio frequency identification (RFID) sensors, this article proposes a new method that combines blockchain and ML to enhance the accuracy of blockchain source data. In addition, system data storage and indexing methods and scanning and recovery mechanisms are proposed. Compared with the existing agricultural product (tea) traceability system based on blockchain, the introduction of the ML data verification mechanism can ensure the accuracy (up to 99%) of information on the chain. The proposed solution provides a basis to ensure the safety, reliability, and efficiency of agricultural traceability systems.

## 1. Introduction

As a valuable branch of agriculture, the tea industry has had a long history of development. At present, tea is one of the three major beverages worldwide. China is a major tea exporter [1]. The demand from developed countries such as Europe and the United States accounts for a quarter of the total tea exports. However, since 2001, the presence of pesticide residues and heavy metals has affected the quality and safety of tea, and these problems have emerged as bottlenecks in the development of the tea industry [2]. With globalization, an increasing number of regulators are focusing on the traceability of tea safety and credibility, and the customer

expectations regarding the tea quality are rising. Traditional tea tracing systems cannot guarantee the accuracy of data, tamper-proof modifications, and efficient storage of the system. Therefore, it is necessary to establish a light-weight traceability system for the tea supply chain that is credible and traceable.

The existing supply chain of tea has a complex structure, so it is necessary to establish an efficient and credible traceability system to realize the supervision and recall of tea. Such a system can help rapidly identify the relevant links and data to promptly solve the problem [3]. To build an agricultural traceability system, it is necessary to collect a large amount of data pertaining to the agricultural industry

chain. Notably, system information collection mostly relies on manual entry, which may lead to problems such as information recording errors and inefficient use of the system. The Internet of Things (IoT) and sensor technologies can be used to effectively solve these problems [4]. In general, bar codes, quick response (QR) codes, radio frequency identification (RFID), and other technologies are used in the information collection process of tea traceability, which can promptly and effectively transmit information. The RFID technology can track and monitor the whole process “from a tea garden to a teacup.” In the case of a dispute over the origin or traceability information, the problem can be rapidly identified [5]. Wireless sensor networks (WSNs) and RFID are components of the IoT. WSNs include temperature, humidity, and pressure sensors. Through the corresponding network, the temperature, humidity, and other information in the process of tea planting can be automatically obtained and transmitted to the system. IoT devices can ensure the transparency and reliability of the tea supply chain.

The existing tea supply chain management systems involve several limitations. Although information is automatically collected through IoT equipment, the equipment may produce abnormal data in the process of data collection due to the influence of the manufacturing technology, process, and cost and network transmission. Therefore, the accuracy of the source data cannot be guaranteed [6]. Moreover, the tea traceability system involves many participants, which may lead to inefficient information sharing. Records of the pesticide content, origin and grade of tea, and tea traceability process information may be imperfect. Moreover, a whole process supply chain traceability system for tea has not been established yet [7, 8].

The problem of category imbalance is encountered in many fields such as medical diagnosis [9], outlier detection [10], software quality detection [11], credit card fraud detection [12], and text classification [13]. Research on outlier detection for agricultural traceability data is aimed at classifying unbalanced data. The outliers and normal values are known as the minority and majority classes, respectively. Compared with majority class instances, minority class instances often receive more attention. An ensemble method is an effective algorithm to enable leaning from minority samples [14]. In particular, an ensemble algorithm achieves a high accuracy by forming a set of base classifiers. Boosting is an ensemble method that can enhance the performance of weak classifiers in an iterative manner by modifying the weights of the samples after each iteration. Specifically, by iteratively assigning higher weights to misclassified samples, the subsequent models can be more accurately classified. Adaptive boosting (AdaBoost), gradient boosting, and XGBoost are all well-known boosting algorithms [15] that can enhance the performance of any weak classifier. Random forests (RFs) represent another type of ensemble method that can introduce additional randomness to the prediction of a decision tree to obtain a variety of classifiers.

To address the abovementioned challenges, we propose and design a machine learning- (ML-) blockchain-IoT-based tea credible traceability system (MBITTS). The pur-

pose of the system is to promptly collect data, verify the accuracy of the source data, and establish a new traceability system with distributed storage, data scanning, and monitoring. The main functions of the system can be summarized as follows: First, the source data are stored in a relational database by IoT devices. An ensemble learning algorithm (AdaBoost, XGBoost, gradient boosting decision tree (GBDT), or RF) is used to perform binary classification of the traceability data to filter the outliers. Moreover, several models, such as logistic regression (LR) and K nearest neighbors (KNN), are trained, and their performance is compared with that of the ensemble algorithms. After the information is validated, it is transferred to the blockchain for storage. Subsequently, a collaborative storage architecture for blockchain, relational database, and memory database is established. Based on the multitasking real-time scanning of blocks in the memory database, a fast query mechanism for data tracing is constructed. Finally, any deletion or tampering in the system is recorded and reported, and a decentralized coordinated agricultural product tracking system is established.

The remaining paper is organized as follows: Section 2 presents a review of the research on the use of IoT and blockchain in the current agricultural product industry. Section 3 describes the proposed framework, with IoT and blockchain technologies used to design the system. The system requirements are identified, and the design of the system architecture is established. Section 4 describes the design and implementation of the system. Section 5 presents the experimental analysis. Section 6 presents the concluding remarks, discusses the limitations, and recommends directions for future work.

## 2. Related Work

Owing to the discrete nature of the various functions and stakeholders in the agricultural supply chain, the government, consumers, and enterprises must establish a complete and efficient agricultural supply chain information supervision system. Pignini and Conti [16] proposed the use of near field communication (NFC) technology to achieve efficient traceability in the agricultural supply chain. To promptly collect information regarding key points in the supply chain by establishing NFC tags, Alfian et al. [17] used IoT technology and ML methods to enhance the efficiency of the fresh food traceability system. However, these technologies involve several drawbacks. Specifically, these technologies fail to solve the fundamental problem of traceability systems. For example, in the case of centralized data storage, people may easily delete or tamper with the data. In addition, the traceability information of the whole process may lead to storage redundancy and low query efficiency. Kamble et al. [18] proposed a smart agriculture system based on blockchain technology and the IoT. All users can use smartphones to enter the data into the system and collect data through IoT devices. All the traceable data can be communicated after node consensus, and digital signatures are obtained. Galvez et al. [19] highlighted how blockchain is used in the food supply chain, considering soybeans as an example, and discussed traceability-related cases in the food supply chain. Unilever [20] also makes the tea traceability system more transparent through blockchain technology, so as to facilitate consumer

query and enterprise management. Recently, blockchain technology has emerged as a promising technology. Notably, because it can solve the credibility and security problems caused by information opacity and tampering, blockchain technology is widely used in the field of agricultural food [19, 21–24]. Liao and Xu [25] proposed a tea quality and safety management system based on blockchain. The Ethereum platform and MySQL relational database constitute the data layer, which ensures the distributed storage of tea and the credibility of the data in the traceability process. This technology is often combined with other technologies. For example, Khan et al. [26] combined the IoT, blockchain, and deep learning methods to optimize the food source problem in Industry 4.0, using a regression neural network to predict the relationship between the supply and demand of food. Blockchain technology can be used to potentially overcome the limitations of traditional traceability software systems.

Although the abovementioned studies discussed the advantages of blockchain in agricultural supply chain traceability, they did not consider specific application scenarios and system optimization issues, such as

- (i) *Information Immutability.* The blockchain ensures the originality and authenticity of records, and thus, once the original data are input to the blockchain system, they cannot be modified. However, this aspect does not ensure the authenticity of the original information, and the initial data must be assumed to be correct. To address the practical problems of agricultural supply chains, we must formulate appropriate methods to accurately screen the error value of the source data and establish effective information chains [27]
- (ii) *Data Storage.* Different types of data exist in the agricultural supply chain. Blockchain information is stored in the block body [28], and collection of the traceability information of the whole chain (including the information of the key process from production to processing to logistics) may significantly increase the size of the block storage load, thereby affecting the system performance. Therefore, it is necessary to improve the existing blockchain storage model to optimize the efficiency of the system operation

Therefore, this article constructs an ML-blockchain-IoT-based system solution that can guarantee the traceability of the agricultural supply chain while improving the data accuracy and information security management of the system. The proposed system relies on GBDT and other ensemble learning methods to analyze the source data and screen outliers. Compared with most existing classification methods, an ensemble learning method provides a considerably higher accuracy than that achieved using only LR and KNN, and it can thus help ensure the source data accuracy [29, 30]. In addition, the proposed data storage and indexing mechanisms and data scanning and recovery mechanisms can provide a theoretical foundation for the optimization of blockchain systems.

### 3. Framework

The proposed method uses the IoT and Ethereum as the underlying technology. RFID is one of the most common technologies in the IoT. Specifically, RFID is a noncontact automatic identification communication technology [31]. In the field of traceability, RFID is used to automatically obtain data in all aspects of logistics, such as production, processing, warehousing, and distribution. Consequently, the IoT technology is a key technology for intelligent agricultural traceability management. Ethereum is a typical application of blockchain, which was invented by a Russian–Canadian computer scientist named Vitalik Buterin in 2013. Ethereum is an open-source public blockchain platform with a smart contract function. A decentralized Ethereum virtual machine is used to process point-to-point contracts through the dedicated cryptocurrency, ether (ETH).

This section introduces the credible traceability system for the entire tea process based on ML blockchain-IoT, named MBITTS. Figure 1 shows the framework design of MBITTS, which is divided into three modules.

*Traceability layer.* In this layer, raw data from external systems are recorded through connected devices (such as IoT sensors, ERP system data, and RFID) and input to the system business logic layer through a unified data exchange interface (JSON-HTTPS). The external device collects and transmits information such as the geographic location, temperature, humidity, and pesticide spraying amount. RFID tags are used to automatically identify and track the tea products on the chain. These later stages communicate with the shared ledger information in the blockchain.

*Business layer.* This layer contains functional modules such as data upload, data verification, data scanning and monitoring, data rollback, and system credit evaluation. The data of the entire agricultural industry chain (including production, processing, and logistics information) are transferred to the ML network model through IoT equipment or manual entry for data verification. The verified data are transferred to the blockchain database through the smart contract to automatically execute the script code. The system simultaneously performs batch processing (day-to-day chain scanning and day-end integration) and transaction monitoring operations to integrate and manage various processes and transactions in the traceability chain. The characteristics of blockchain technology ensure that the entire process of storage, reading, and execution is transparent, traceable, and nontamperable. Moreover, a state machine system is constructed using the consensus algorithm provided with the blockchain to enable smart contracts to run efficiently. The ensemble learning method is used to classify outliers in the traceable data. Direct, high-quality data can be generated, and the feedback results can help enhance the overall credibility of the system.

*Database layer.* This layer includes relational databases and the Ethereum blockchain platform. As a ledger, the blockchain stores each transaction in a distributed P2P network and records it after it is verified by all participating nodes. The data verified by the ML model are uploaded to

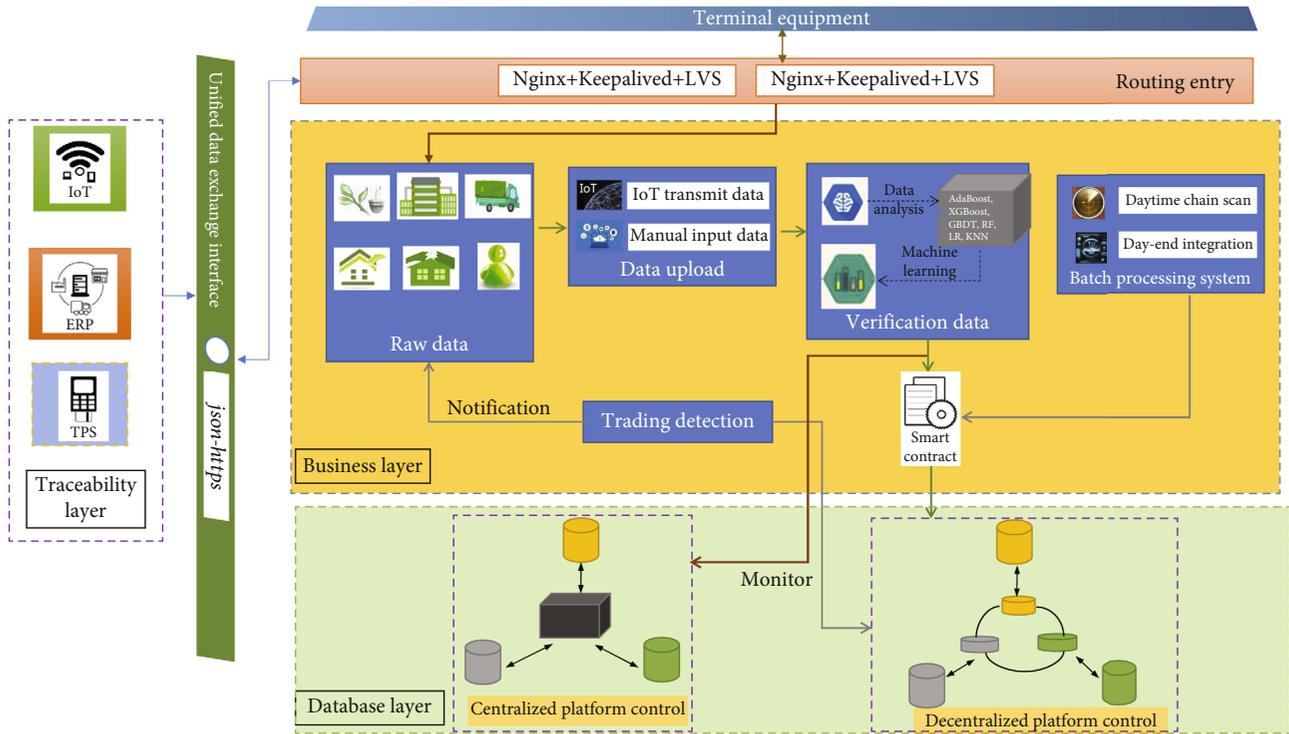


FIGURE 1: Framework of the ML-blockchain-IoT-based tea credible traceability system (MBITTS).

the Ethereum blockchain platform through the data communication protocol, and the ledger data are agreed upon at each node of the blockchain through a certain propagation mechanism. Each node in the blockchain records information, and each node reaches a consensus by calling the smart contract in the blockchain network. Ethereum uses consensus mechanisms such as proof of work (POW), proof of stake (POS), and proof of authority (POA), and the encrypted hash value after the consensus is stored in the blockchain node of the ledger.

#### 4. System Design and Implementation

According to the system architecture design method described in Section 3, the blockchain system contains the production, processing, circulation, and sales information of the tea supply chain. Information integration, on-chain data accuracy verification, fast data query, traceability data recovery, and other functions can be reflected in the platform. The following contents introduce system dataflow, system storage, data recovery mechanism, data verification algorithm, and the implementation of the system.

**4.1. Dataflow of the System.** The main members of the tea trusted traceability system include farmers, producers, distributors, transporters, product users, and managers. Information regarding a tea producer includes the company's registration information, temperature, humidity, and amount of pesticides used in the tea planting process. Tea production information includes the processing batch number, date, and output. Tea sales information includes the price, shelf life, and sales date. The key information gener-

ated in each link of the traceability system is automatically obtained through GPS or wireless sensors and other equipment. The EPC code is used for unified identification through RFID tags. The data are uploaded to the database through the data acquisition module, and a classification model is built through ML methods to filter abnormal data. The correct data are recorded in the blockchain in the form of transactions through smart contracts. A transaction in a contract includes the transaction address, content, gas, date, and signature. As shown in Figure 2, the transaction data, tea information (including location, timestamp, and tea variety), and number of tea leaves are verified on the chain with the digital signature of the tea buyer and seller. Subsequently, all the data are hashed, and the data verified by the smart contract are uploaded to the blockchain network.

#### 4.2. Storage and Data Recovery Mechanism

**4.2.1. Storage Mechanism.** The mechanism of most existing blockchain systems is to store the information of the whole supply chain process in the blockchain. With the increase in transactions, the nodes must store more data, and the load pressure of blockchain storage increases. To address this aspect, we design a model (Figure 3) known as the centralized and decentralized cooperative storage management model.

The proposed system mainly includes enterprises, consumers, and government regulators, among which the corporate entities include manufacturers, packers, distributors, and retailers. For example, for the tea production information, the traceability fields of the local database are shown in Table 1. The fields include teaPick.id, teaPick.landID,

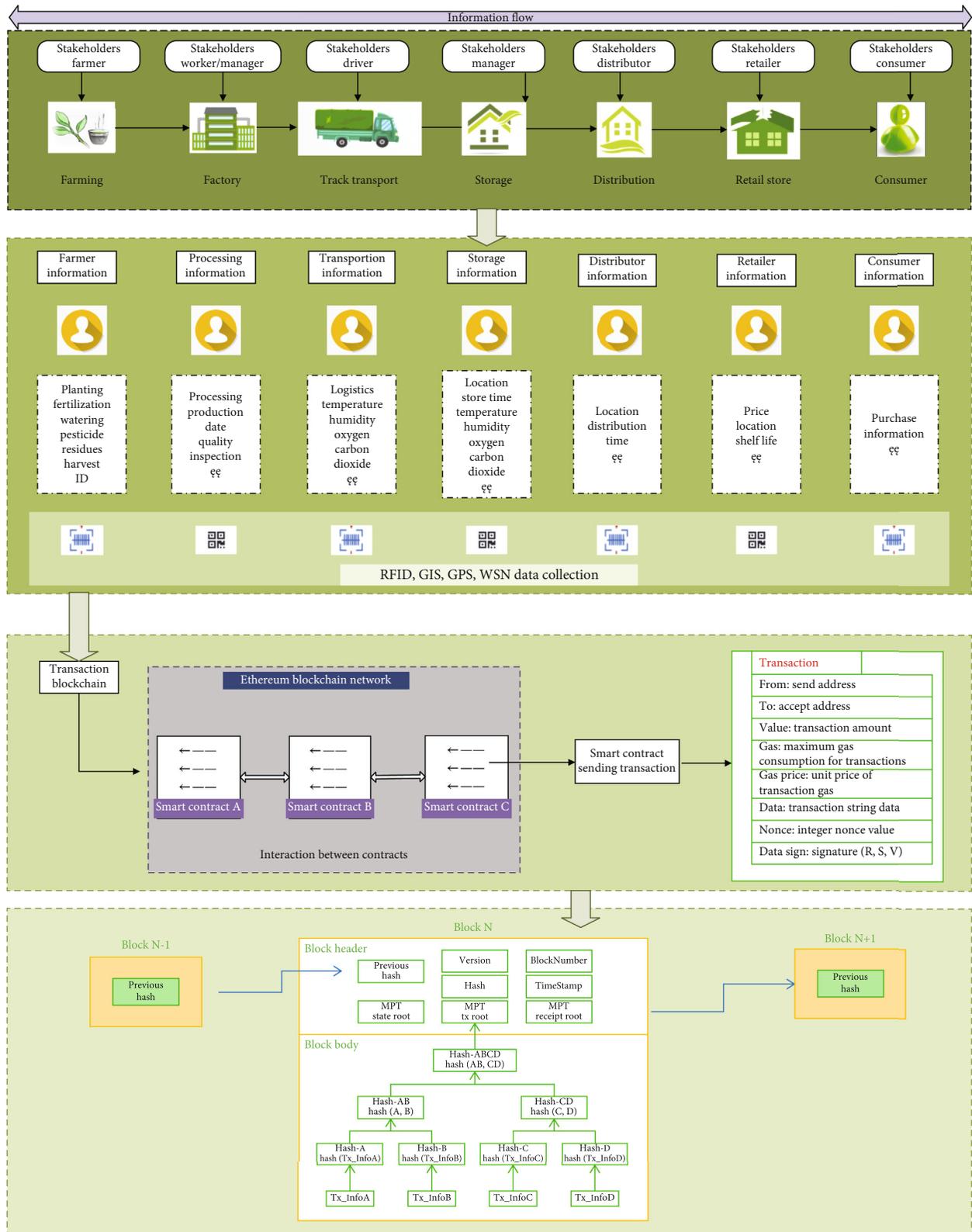


FIGURE 2: Dataflow of the MBITTS system.

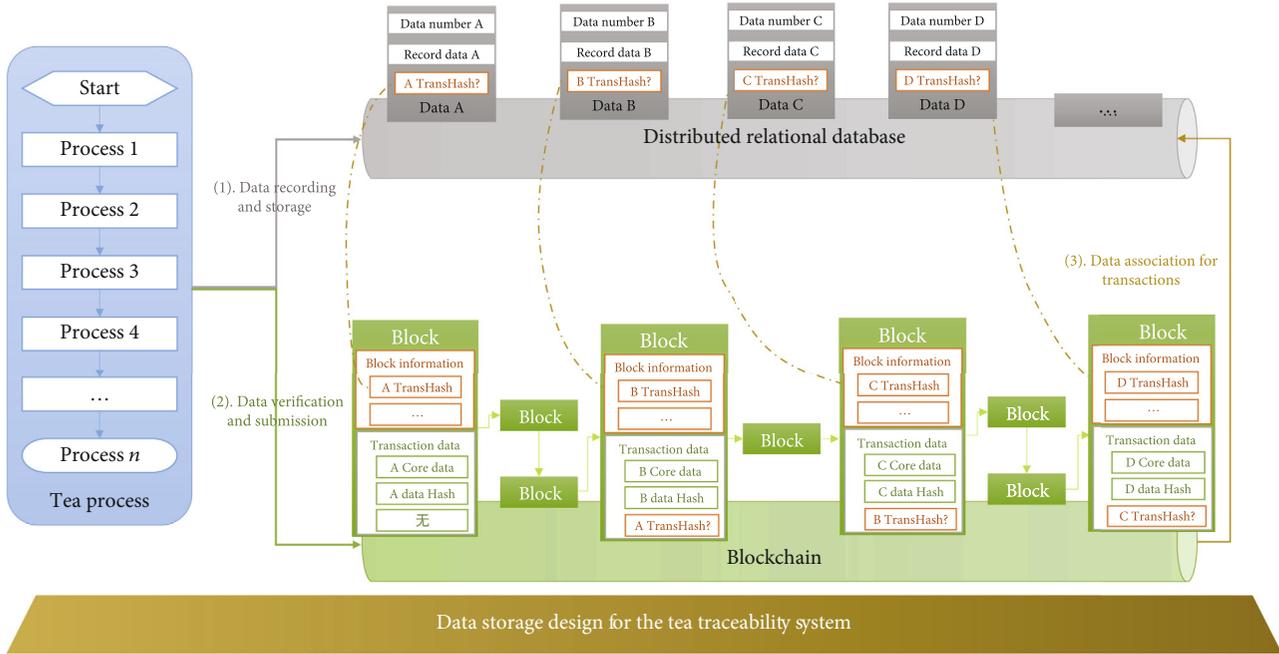


FIGURE 3: Storage mechanism.

TABLE 1: Structure of TeaPick.

No.	Field	Type	Null (true/false)	Description
1	teaPick.id	int	False	Picking number
2	teaPick.landID	String	False	Land number
3	teaPick.weather	String	False	Picking weather
4	teaPick.varieties	String	False	Tea variety
5	teaPick.num	uint	False	Picking quantity
6	teaPick.pickTime	Date	False	Picking time
7	teaPick.usrId	String	False	Operator
8	BlockNum	int	True	Block number

teaPick.weather, teaPick.varieties, teaPick.num, teaPick.pickTime, teaPick.usrId, and BlockNum. The ID number is the unique identifier of the information record. The ID and name fields must be uploaded to the chain, and other fields are uploaded to the chain according to specific roles and processes. These data models are finally converted into JavaScript Object Notation (JSON) strings for storage.

In terms of data storage, as the blockchain adopts a chain structure, it exhibits time irreversibility and nontamperability, and thus, the data stored in the blockchain are continuously accumulated. Data entry and storage consume a large amount of resources in the blockchain system, and the operating efficiency of the system is affected. Additionally, in the blockchain, only the chain structure is shown on the blockchain. The internal transactions are packaged and stored together and do not exhibit any relationship. The proposed collaborative management model stores all traceable data in the cloud server, and only the core data of the transaction and data hash value of the previous transaction are stored in

the blockchain. When new data need to be written, we identify the last transaction hash stored at the previous time and write it as a transaction in this block to find the associated parent transaction hash value in each block, thereby promptly identifying the complete traceability chain information. In this manner, the storage pressure of the blockchain system can be alleviated.

In terms of data queries, smart contracts are used in the blockchain to manage the on-chain data in the food supply chain. Usually, the traceability information is stored in the cloud database for each node, and the complete transaction information is synchronized to the blockchain. When sending a query command, all nodes must be indexed level by level to obtain complete traceability information (i.e., all nodes are traversed to find the associated transaction information). Consequently, in a complex supply chain environment, the response time of the system may significantly increase. By looking up the hash value of the previous traceability data stored in each node, we can directly trace the last transaction data to the previous transaction, thereby enhancing the query efficiency of the blockchain system.

Although two-layer or multilayer frameworks have been proposed in blockchain systems [32, 33], most of these frameworks are aimed at solving the problem of participation of regulators and privacy protection of confidential information in nodes. It is mainly the design of information storage at the business level. In their design, they considered the separation of information storage and query but did not change the internal storage structure of the blockchain. The framework proposed in this paper is different from other multilayer frameworks. This section presents a simple example to illustrate the data storage mechanism of the collaborative management system. We consider a tea supply chain as an example, in which processes 1, 2, 3... represent tea picking, processing, packaging, etc.

- (1) The user sequentially submits each tea process to the database, as shown in Figure 3. When step A is executed, tea process 1 is stored in the local server. At this point, the data are stored and not packaged in the blockchain system
- (2) When step B is implemented, ML is used to eliminate the abnormal values of traceability data, and the data are submitted to the blockchain as a transaction, including the record number, core data, transaction hash, and TransHash of the previous process. In this period, the data are about to be packed. (After each piece of data is submitted, the system verifies whether the previous process is completed. If the transaction is completed, the system executes the TransHash operation of the previous process when submitting the data point; otherwise, the submission is prohibited.)
- (3) The data association is completed by performing step C. At this time, the data in process 1 cannot be modified or deleted

Therefore, by recording the hash value of the transaction record corresponding to process 1 in the transaction generated in process 2, the previous transaction information can be promptly identified during the query without traversing all nodes. In addition, when data are deleted, data can be restored through data rollback. This model is named the collaborative management storage model.

**4.2.2. Data Recovery Mechanism.** At present, MySQL is usually introduced between the application platform and Ethereum blockchain. However, owing to its centralized nature, the data stored in MySQL are vulnerable to contamination. If the hash value of a historical transaction stored in the database is deleted, the user cannot retrieve the transaction data in the blockchain. Therefore, even if the data stored in the blockchain are safe, the data whose addresses are lost may not be able to be retrieved. This aspect leads to loopholes in the data query. Therefore, the proposed system uses a set of data scanning and recovery mechanisms (Figure 4). This section presents a simple example to illustrate the data recovery and correction mechanism.

The batch processing system obtains the total number of current blocks in the blockchain. If the system is required to traverse and scan the massive amount of transaction data in the blockchain, the process will be extremely time intensive. Therefore, batch scanning (scanning by block group) is performed. The width of the block to be scanned is defined and recorded as a block group (start, stop). (1) Scanning tasks are performed in a cyclic manner (for example, every day or every hour) to increase the rate of scanning the system transaction data in the blockchain. (2) The data in the database are reverse checked through the transaction hash in the blockchain. If the data cannot be checked, it is considered that the data have been maliciously deleted. This information is recorded and written into the intelligent contract, and the database data are rolled back. (3) If the data are found in the database, the database data are evaluated

through hashing. If the data are inconsistent with the hash block, which is suggestive of data tampering, the data information is recorded and reported, and the in-memory database data are rolled back for the next scan. This process ensures the accuracy of the data in the MySQL database. The data recovery process is presented as Algorithm 1. Considering the example of tea picking, Teapickservice.update (data item) is used to perform the rollback.

**4.3. Ensemble Method for Data Check.** ML can facilitate decision-making by learning the human reasoning process to judge the results in advance. A model is built by learning from a real dataset, and an unknown dataset is forecast to enhance the system accuracy [34]. Classification is a widely examined topic in the ML domain. Traditional classification methods, such as decision trees, LR, support vector machines (SVMs), and Bayes classification, impose many requirements for the distribution of the original data. In particular, when the data are unbalanced, the prediction performance of the model is inferior [35]. When the agricultural product traceability data are classified into two categories, the number of outliers is less than normal data, and thus, this dataset is unbalanced. At present, the integration method is an effective way of solving this problem. This method combines the data mode and algorithm level methods to address the problem of unbalanced data [36].

The integration method is a meta-algorithm that combines several ML technologies to form a prediction model. This approach is a commonly used classifier. Integrated learning methods are effective for different datasets [37]. Bagging and boosting algorithms are commonly used integrated learning algorithms. Representative algorithms based on bagging include RF, while common algorithms based on boosting include AdaBoost, GBDT, and XGBoost.

- (i) *Bagging*. The original training set is divided into  $M$  subsets of the same size by replacement or sampling without replacement. Each subset builds a classifier and assigns equal weights to each classifier. Finally, the results of the underlying classifier are used as votes for the final prediction
- (ii) *Boosting*. Introduced by Freund and Schapire in 1996, boosting is currently the most popular method for integrated learning. This approach constructs a strong classifier by combining multiple weak classifiers [38]. Subsequently, the new weak learner is continuously used to compensate for the “deficiency” of the previous weak learner to construct a strong learner in an iterative manner. This strong learner can ensure that the value of the objective function is sufficiently small. The learning feature of boosting is highly suitable to address class imbalance problems

This paper selects representative methods from all ensemble methods: AdaBoost, GradientBoost, XGBoost, and RF. Next, we introduce the algorithms of RF, AdaBoost, and GBDT.

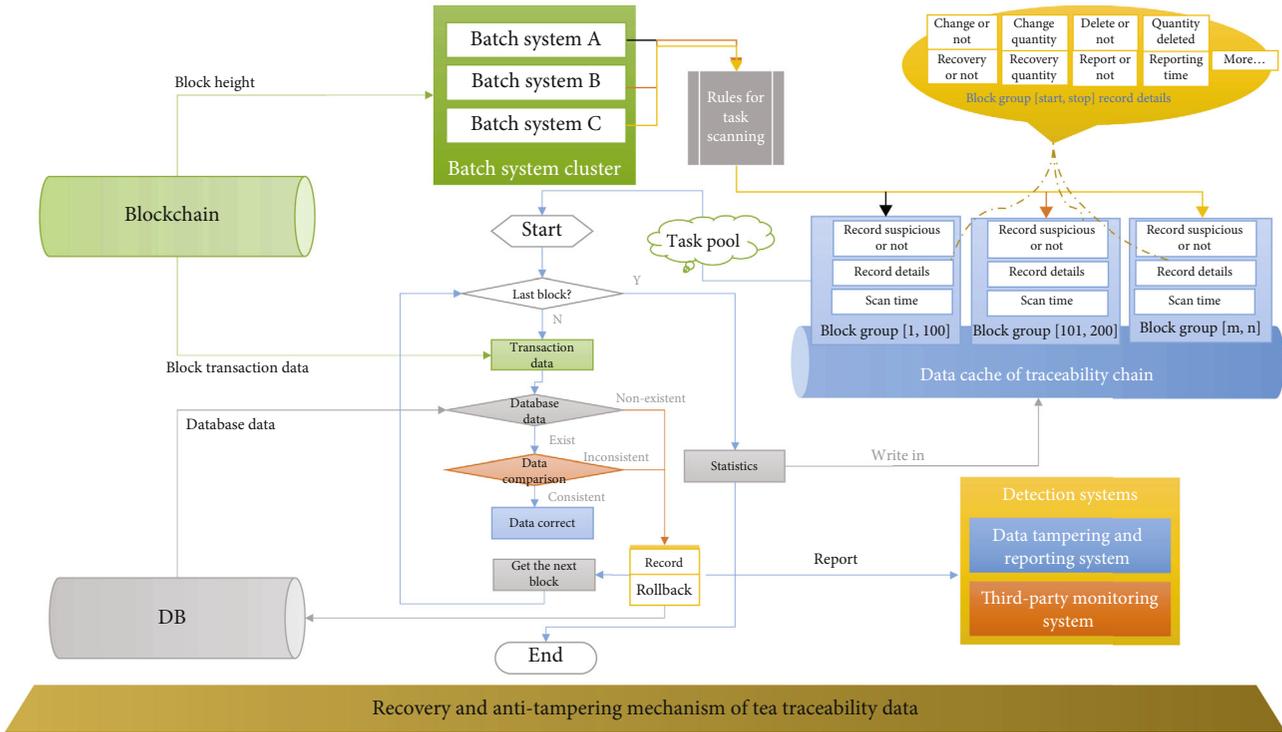


FIGURE 4: Recovery and antitampering mechanism.

```

Pair<Boolean,Pair<ITeaEntity<?>,String>>compareResult = ((TeaPick)teaWrapperEntity.getContractData()).compare(teaPickDb);
if (!compareResult.getLeft()) {
ITeaEntity<?> teaPickPrepareRepairDbData = compareResult.getRight().getLeft();
teaPickService.Update(teaPickPrepareRepairDbData);
}
    
```

ALGORITHM 1: Data recovery process (tea pick).

4.3.1. *RF*. RF is a representative bagging model, which is a new and effective classifier. An RF is composed of many decision trees. Proposed by Breiman in 2001 [39], an RF model simultaneously trains multiple trees. The majority decision regarding the number is considered the final prediction result of the model.

4.3.2. *AdaBoost*. AdaBoost is considered the most famous boosting algorithm. AdaBoost was proposed by Freund and Schaire in 1997 [40, 41] and is widely used in the field of binary classification. The bagging algorithm introduced in the previous section generates several base classifiers by bootstrapping. The sample sets may be the same or different for each learning run. AdaBoost learns by increasing the weight of the misclassified samples in the iteration of boosting. Specifically, the training dataset is fixed, and the accuracy is enhanced by learning the misclassification of the existing samples.

4.3.3. *GBDT*. The GBDT is an ensemble learning method based on a decision tree and is a kind of boosting algorithm [42]. The GBDT constructs a phased addition model. Unlike

AdaBoost, GBDT uses a Cart tree as a subclassifier. Each iteration of GBDT generates a decision tree by fitting a negative gradient [43]. The algorithm for constructing the GBDT classifier is as follows:

By using an ensemble learning method to perform binary classification of the traceability data of agricultural products (tea), the outliers of the data on the chain can be more accurately and efficiently screened out. Thus, the quality of the source data is ensured to a certain extent.

4.4. *System Interface*. Tea has a history of thousands of years in China and is an important agricultural product. Anhui is the hometown of Chinese tea and yields many famous types of tea, such as Huangshan Maofeng, Lu’an Guapian, and Taiping Houkui. This paper builds a tea blockchain traceability system by investigating the supply chain process of Huangshan tea companies. The system implements distributed storage on the Ethereum Geth platform, which provides a standard web API interface.

According to the system architecture, the system has two main modules: (1) an on-chain information verification module and (2) a blockchain information query and

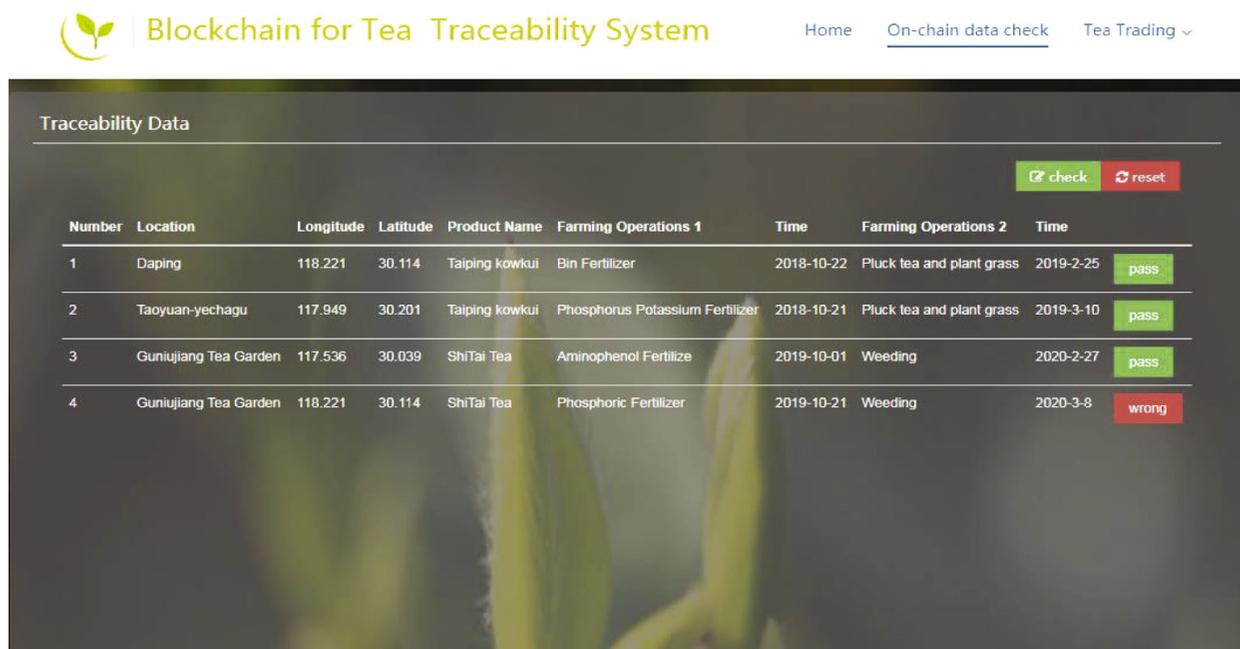


FIGURE 5: On-chain data verification page.

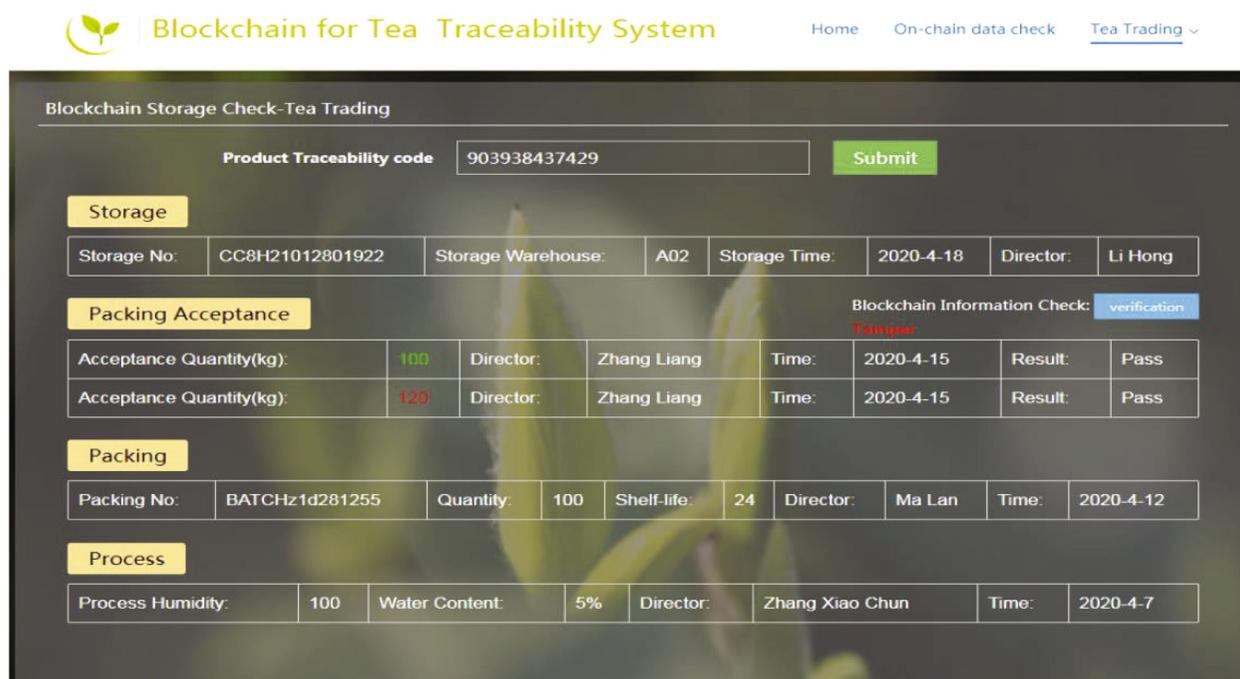


FIGURE 6: Blockchain information verification page.

verification module. Figure 5 shows the source data verification page. Using the ensemble learning method to filter the outliers from the traceability data, users can intuitively identify incorrect data on the chain. Figure 6 shows the blockchain information verification interface. All the tea traceability information is stored in the block, with the corresponding timestamp and hash value. This information cannot be tampered with. For a problematic product, the government and enterprises can compare the corresponding

hash index value with the information stored in the database. The tampered data are marked in red to complete the verification of the traceability information. This framework ensures the authenticity and reliability of the information. In addition, the system provides a blockchain traceability query interface. Users can view information regarding product traceability through a unique traceability code, and the displayed hash value is the blockchain information inquiry proof.

TABLE 2: Experimental environment.

Environment	Details
Development platform	Eclipse Luna 4.4.2; VS code 1.49.1
Operating system	Ubuntu 16.04
Blockchain module	Ethereum Geth v1.9.9; nodejs v8.10.0; solidity 0.5.7
Running environment	Java 1.8.0; Tomcat 8.5.35, MySQL 5.7.31
Language	Java; HTML, Javascript, JQuery, CSS; Go
Development framework	s2sh

1: **initialize:**  $F_0(x) = \text{argmix}_{\rho} \sum_{i=1}^N L(y, \rho)$ , sample size  $N$ .  
2: For  $m = 1$  to  $M$  do:  
3: end For  
(1):  $\tilde{y}_i = 2y_i / (1 + \exp(2y_i F_{m-1}(x_i)))$ ,  $i = 1, 2, \dots, N$   
(2): According to  $\tilde{y}_i$ , the  $M$ th regression tree is obtained, and the corresponding leaf node region is  $\{R_j\}_1^J$ .  
(3):  $\gamma_{m,j} = \sum_{x \in R_{m,j}} \tilde{y}_i / \sum_{x \in R_{m,j}} |y_i| (2 - |\tilde{y}_i|)$ .  
(4):  $F_m(x) = F_{m-1}(x) + \sum_{j=1}^J \gamma_{m,j} I(x \in R_{m,j})$   
4: **Output:** Final classifier  
 $F_M(x) = F_0(x) + \sum_{m=1}^M \sum_{j=1}^J C_{m,j} I(x \in R_{m,j})$

ALGORITHM 2: GBDT algorithm.

## 5. Experiments and Numerical Analysis

An MBITTS system exhibits an efficient data upload mechanism and data recovery mechanism and can ensure the reliability of source data and efficiency of storage and query. The system is of practical significance in a tea tracing system. The main advantages can be summarized as follows: (1) The IoT technology guarantees the transparency of the traceable data, (2) the distributed storage improves the security of the tea tracing system, (3) the ML-based filtering mechanism enhances the system data quality, (4) the lightweight design of the storage and query mechanisms increases the system efficiency, and (5) the data recovery mechanism improves the integrity of the traceability information. This section introduces the experience platform, design of smart contracts, and evaluation of the source data filtering model. Finally, the proposed system is compared with other traceability systems.

*5.1. Experiment Platform.* The proposed framework is developed based on Ethereum Geth v1.9.9. All the experiments involving ML are implemented with Python 3.8. The hardware configuration of the computer used in the experiments is as follows: Intel(R) Xeon(R) CPU X5670 @2.93 GHz \* 6, 16 GB memory, and a 120 GB hard disk. The whole system is developed in Java on the Eclipse Luna platform of Windows 7, and other details of the basic environment are presented in Table 2.

*5.2. Smart Contract Design.* Smart contracts represent a key technology in the traceability management system of an agricultural product supply chain driven by blockchain. These contracts can enable the automatic screening of the

entire process information of the agricultural product supply chain and preservation of blockchain transactions. A smart contract can be understood as an event-driven executable program that contains status information in the blockchain [28, 38, 44]. The two parties involved in the transaction agree on the content of the contract, execution conditions, and default conditions, and the smart contract is deployed on the blockchain. Each contract generates an address, and both parties involved in the transaction use the address to call the contract. Considering tea picking information storage as an example, a smart contract execution plan is designed.

Algorithm 3 describes the data upload process in the tea picking process. The data are input through the application platform pertaining to the tea picking process. After collecting the tea picking data, it is necessary to verify whether the previous stage of tea picking was successfully recorded in the blockchain. After the verification, the tea picking data are encapsulated into the DefaultContractEntityWrapper object, and the tea picking contract method is associated with the data through the TeaPickFunctionProvider. The data enter the chain through RawTransactionManager.sendTransaction, and the transaction hash is returned.

Algorithm 4 describes the query contract for traceable information in the tea supply chain. Each role inputs query information (tracking source code) through the traceability interface and calls the data query contract. Blockchain data are obtained through transaction hashes and query database data based on blockchain data, and the data are compared to determine whether they are consistent. When data are received, the server divides the data according to the information type, obtains the operation

```

// Verify the correctness of picking data and judge whether the last link is completed.
Results results= isValid(teaPick);
if (!results.isSuccess()){
return results;
}
// Package picking data.
IContractEntityWrappercontractEntityWrapper=new DefaultContractEntityWrapper(TBL_TEAPICK,TeaPick.class.getName(),tea-
Pick,SIGN_SHA256);
// Generate associated contract data and generate contract call methods.
Function teaPickFunction =TeaPickFunctionProvider.getInstance().run(new
Object[] {contractEntityWrapper.getContractData().getTid(),JsonHandler.getInstance().get(contractEntityWrapper)
,TEAPICK});
IProviderHandler functionProvider = new DefaultProviderHandler(teaPickFunction);
Credentials credentials = ContractsMap.getInstance().get(contractAddress);
// Trading store.
return EthTrans.send(credentials, contractAddress, functionProvider);

```

ALGORITHM 3: Data acquisition contract (tea picking).

```

String transactionPreHash = transactionHash;
do{
IContractEntityWrapper ethContractEntityWrapper = (IContractEntityWrapper)queryTranInfo(transactionPreHash).getData();
ITeaEntity<?> dbContractData = selectDbEntityService(ethContractEntityWrapper);
DefaultContractEntityWrapper dbContractEntityWrapper = new DefaultContractEntityWrapper(ethContractEntityWrapper.getTable-
Name(), dbContractData.getClass().getName(), dbContractData, ethContractEntityWrapper.getSignMethod());
Pair<Boolean, Pair<ITeaEntity<?>, String>> compareResult = ethContractEntityWrapper.compare(dbContractEntityWrapper);
list.add(compareResult);
transactionPreHash = ethContractEntityWrapper.getContractData().getPreTransHash();//Move to the previous data
}while(!StringMan.isNull(transactionPreHash)); // If the previous number transHash is empty, the traceability ends
return new Results().OK(new Object[] {list, calculateScore(scoreList)});

```

ALGORITHM 4: Data query contract.

record information of each link by selecting the mark of each traceable link, and returns the recorded traceable information in chronological order.

**5.3. Implementation of the Ensemble Model.** The classification process of agricultural product (tea) traceability data is shown in Figure 7. The experiment is designed according to the process flow to test the performance of different ensemble algorithms for binary classification.

**5.3.1. Experimental Dataset.** The dataset used in this experiment pertains to the traceable information provided by a tea company in Huangshan and has 500 data points, including 26 characteristic values: longitude, dimension, product specification, shelf life, product grade, appearance characteristics, planting date, packaging date, etc. The dataset contains 450 correct values and 50 abnormal values. The dataset is expanded to 1000 points via the smoothing method, including 900 correct values and 100 outliers.

The results of this experimental classification model are traceable normal values and abnormal values. This experiment focuses on determining the outliers of the traceable data. The outliers are marked as 0 to represent positive samples; that is, abnormal eigenvalues such as longitude or

dimension appear in traceability. Marking accurate data as 1 means negative sample; that is, no error occurred in each characteristic value of traceability. Subsequently, the dataset is randomly divided into a training set and test set with a ratio of 7:3. The abovementioned procedure corresponds to the preprocessing stage for the tea traceability data.

Next, four ensemble learning algorithms are used to establish the traceability data classification model. Moreover, LR and KNN, which are commonly used data analysis algorithms that can classify traceable data, are considered. We compare the six algorithms and evaluate their classification performance. In this paper, the system learns and generates classifiers from the dataset through machine learning method. After the generated classifiers are loaded into the traceability system, the judgment results of traceability data are given. The data inspection method based on ML blockchain can help realize quality assurance for the source data of the blockchain system.

**5.3.2. Performance Metrics.** To evaluate a model, the accuracy and error rate are the most widely used indexes. To verify the effectiveness of the model and evaluate its performance in several aspects, this article uses five metrics: accuracy, precision, recall, F1-score, and area under the receiver operating characteristic (ROC) curve (AUC).

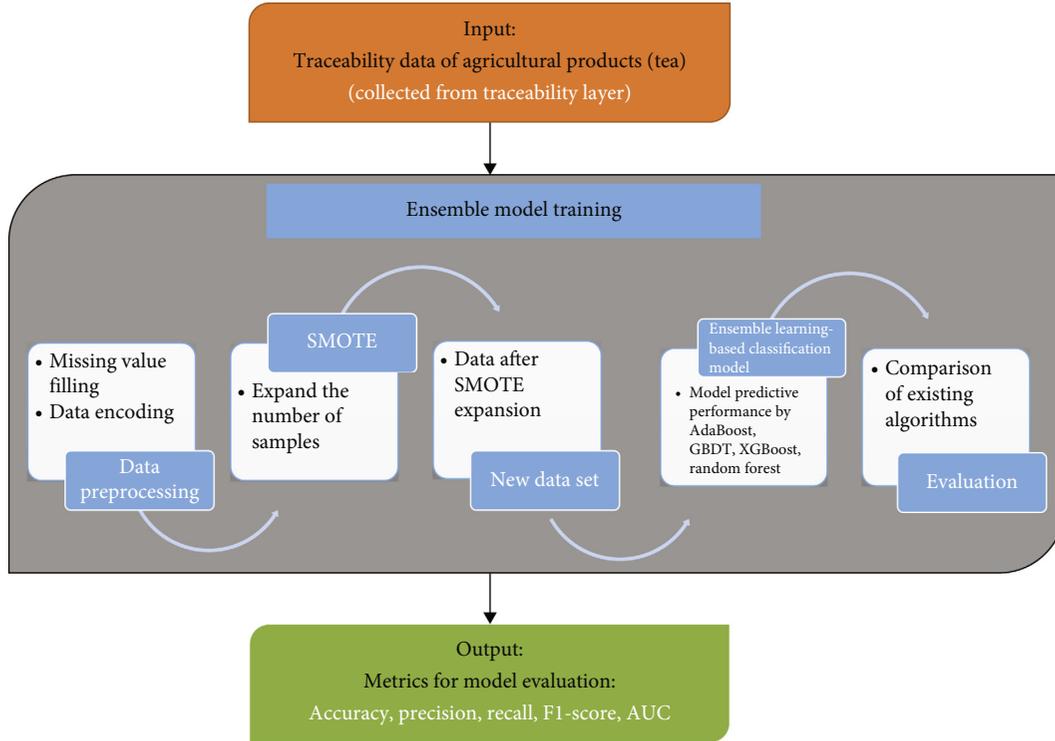


FIGURE 7: Traceability data classification process for agricultural products (tea).

TABLE 3: Confusion matrix for binary classification.

		Predicted	
		Positive	Negative
Actual	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

This is more suitable for the evaluation of unbalanced datasets than using a single index alone. All these indicators are calculated from the confusion matrix. A confusion matrix is a two-dimensional matrix, and its structure is shown in Table 3. Researchers usually regard a few categories as positive, while most categories are regarded as negative. Table 4 shows the measured performance metrics for binary classification based on accuracy, precision, recall, F1-score, and AUC.

There are many indicators for evaluating learners dealing with unbalanced datasets, and the ROC curve is one of the most extensive evaluation methods [45]. It represents the relationship between the true positive rate on the  $y$ -axis and the false positive rate on the  $x$ -axis. The AUC represents classifiers that randomly select positive samples as being ranked higher than the probability of randomly selecting negative samples. For unbalanced datasets, we pay more attention to the accuracy of the minority class (positive class). Therefore, we want the TPR to be as high as possible and the corresponding FPR to be as low as possible. Generally, when  $AUC > 0.8$ , we consider the model to perform well in classification [46].

**5.3.3. Experimental Results.** This section describes the results of the experiment conducted using AdaBoost, GBDT, XGBoost, and RFs. Moreover, results obtained using the LR and KNN are also presented for comparison. Table 5 presents the details of the dataset.

Tables 6–9 present the classification results of the RF, XGBoost, AdaBoost, and GBDT, respectively. Table 6 shows that the accuracy of the RF model is 0.96. The recall rate is 0.61; in other words, 61% of outliers are determined to be positive samples by the model. The false negative rate of the model is 0.39; in other words, 39% of the outliers are misjudged as negative. The false positive rate is 0, and thus, no normal values are identified as outliers by the model. The specificity is 100%; in other words, all normal values are correctly classified.

The recall rates of XGBoost, AdaBoost, and GBDT are 0.54, 0.71, and 0.86, respectively. In other words, 54%, 71%, and 89% of the outliers are detected. The false negative rates of the models are 0.46, 0.29, and 0.11. In other words, 46%, 29%, and 11% of the outliers are misclassified as negative.

In this experiment, we focus on the precision and recall of outliers. According to the recall index, the RF and GBDT exhibit the lowest and highest performance among the four algorithms, respectively.

Figure 8 shows the ROC curves of the six models used in this paper. According to Figure 8, the AUC values of the RF, XGBoost, AdaBoost, GBDT, LR, and KNN models are 0.99, 0.86, 0.91, 0.98, 0.82, and 0.76, respectively. The results show that the AUC of the boosting model is more than 0.8, corresponding to a superior distinguishing ability. In addition, the

TABLE 4: Performance metrics for binary classification.

Performance metric	Formula
Accuracy	$(TP + TN)/(TP + FP + FN + TN)$
Precision	$(TP)/(TP + FP)$
Recall	$(TP)/(TP + FN)$
F1-score	$(2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$
AUC	$\text{TPRate} = (TP)/(TP + FN)$ ; $\text{FPRate} = (FP)/(FP + TN)$

TABLE 5: Imbalanced dataset description.

ID	Dataset	Instances	Features	Class values	Imbalance ratio
1	Tea traceability data	1000	26	2	9

TABLE 6: Classification report for RF.

RF	Precision	Recall	F1-score
0	1.00	0.61	0.76
1	0.96	1.00	0.98
Accuracy			0.96

TABLE 7: Classification report for XGBoost.

XGBoost	Precision	Recall	F1-score
0	0.83	0.54	0.65
1	0.95	0.99	0.97
Accuracy			0.97

TABLE 8: Classification report for AdaBoost.

AdaBoost	Precision	Recall	F1-score
0	1.00	0.71	0.90
1	0.97	1.00	0.99
Accuracy			0.98

TABLE 9: Classification report for GBDT.

GBDT	Precision	Recall	F1-score
0	1	0.86	0.92
1	0.99	1.00	0.99
Accuracy			0.99

RF and GBDT models exhibit the highest accuracy, followed by AdaBoost and XGBoost, and the KNN model exhibits the worst performance.

Table 10 presents a comparison of experimental indicators for different algorithms. The final experimental results show that in terms of the tea traceability data modeling, the boosting algorithms exhibit a reasonable model accuracy, recall, F1-score, and AUC. The GBDT model outper-

forms the other algorithms in terms of the recall, F1-score, and accuracy. The AUC area is only 0.01 less than that of the best RF model, which shows that the AUC result for the GBDT is excellent. The XGBoost model has the shortest runtime.

Overall, (1) ensemble learning is superior to nonensemble learning (LR and KNN). Compared with other methods, LR requires the largest amount of modeling time. (2) The boosting algorithms outperform the bagging (RF) model. In particular, for the recall index, the probability of the GBDT algorithm detecting outliers is 16% higher than that of the RF. The results show that the boosting algorithm exhibits an excellent performance in modeling tea traceability data classification (outlier detection). In particular, the GBDT model is superior to the other methods.

**5.4. Comparison.** The proposed scheme is compared with the traceability scheme of the mainstream system, and the results are shown in Table 11.

The system features and advantages can be summarized as follows:

- (i) *IoT.* Owing to the cost and system design, not all systems use IoT technology
- (ii) *Credibility and Decentralization.* Due to its inherent characteristics, compared with the traditional traceability system, the blockchain system can enhance the degree of decentralization and reliability of the system
- (iii) *Authentic Source Data.* In this paper, the integrated algorithm is applied to the binary classification model of the source data of the blockchain system to filter out traceability outliers. Compared with the existing system, this scheme is feasible and effective
- (iv) *Efficiency of Storage and Query.* The proposed storage and query mechanism can theoretically reduce the storage pressure on the blockchain system due to the continuous increase in data and increase the query efficiency
- (v) *Data Protection Mechanism.* Data rollback and supervision mechanisms are unique functions of the system

According to the table, the proposed solution exhibits several advantages over the systems proposed in [47–49].

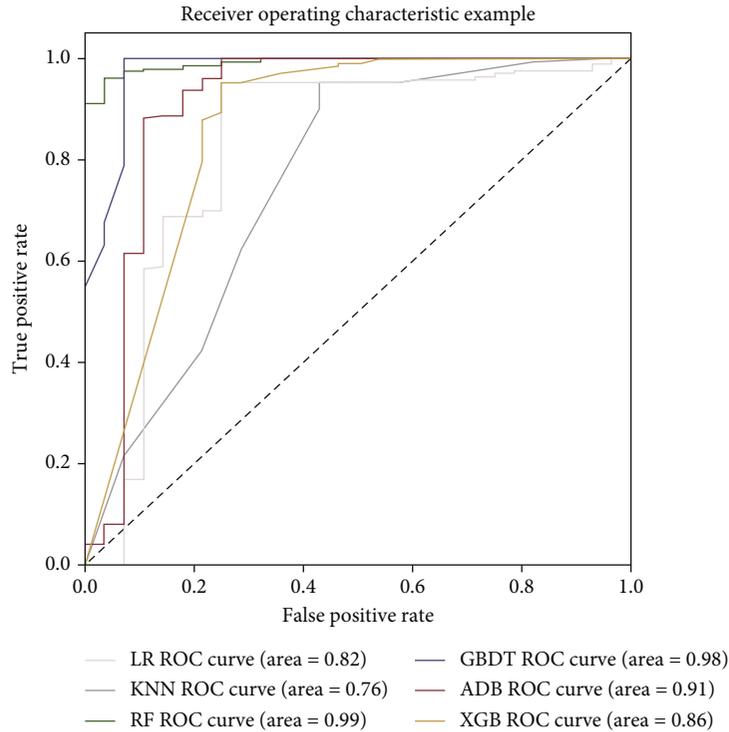


FIGURE 8: ROC curves for the different models considered in this study.

TABLE 10: Classification performance of different algorithms.

Classification models	Modeling time	Accuracy	Precision	Recall	F1-score	AUC
GBDT	0.61	0.99	1.00	0.86	0.92	0.98
AdaBoost	0.21	0.98	1.00	0.71	0.90	0.91
XGBoost	0.11	0.97	0.83	0.54	0.65	0.86
RF	0.73	0.96	1.00	0.61	0.76	0.99
LR	19.33	0.91	0.52	0.50	0.51	0.82
KNN	0.49	0.92	0.71	0.18	0.29	0.76

TABLE 11: Comparison of the proposed system with other systems.

Items	Traditional centralized system [47]	Soybean traceability system [48]	BMLFTS-based system [49]	Our scheme
IoT	No	No	Yes	Yes
Credibility	No	No	Yes	Yes
Authenticity of source data	No	No	No	Yes
Efficiency of storage and query	Low	Low	High	High
Decentralization	None	Low	High	High
Data protection mechanism	No	No	No	Yes

Our system demonstrates a superior source data validation capability than the systems proposed in [47, 48] because we use an integrated learning model to filter abnormal data. In addition, our scheme exhibits advantages associated with data storage queries and data protection. Compared with ordinary blockchain systems, the proposed mechanism can better solve the problems of data storage redundancy and data integrity.

## 6. Conclusions

This paper first highlights the challenges encountered in tea traceability systems. Subsequently, we propose a traceability data management system for agricultural products (tea) based on ML-blockchain-IoT (MBITTS) to achieve efficient storage and management of traceability data. To enhance the quality of the source data, we introduce IoT and ML

methods. First, to improve the quality of the source data, we use IoT technology to format the data uploads and reduce the possibility of incorrect data entry. Based on real tea traceability data, we propose a data verification method with ML and construct a data classification model based on ensemble learning. To increase the efficiency of blockchain storage and query, a collaborative storage management mechanism is designed. To enable data integrity supervision, a data tracking rollback and antitampering mechanism is proposed. In addition, the system architecture and implementation details of smart contracts are introduced. Finally, a prototype system is implemented based on Ethereum Geth. The proposed outlier screening model is evaluated considering a sample case. The comparative analysis of the proposed system and existing traceability system demonstrates the advantages of the novel system.

This research can promote information sharing and exchange in the agricultural product supply chain. The proposed method ensures the transmission and storage of traceability information, improves the quality of the source data, and prevents data tampering. Moreover, the approach involves a reliable data tamper-proofing mechanism for participants, consumers, and government agencies in the supply chain to reduce the occurrence of agricultural (tea) safety incidents. Future research goals can be summarized as follows: (1) In this study, one-shot encoding is implemented to encode the text and numbers in the traceability information. Consequently, it is necessary to explore different encoding methods for different types of data (text and numbers, time series, and nontime series) to increase the accuracy of outlier filtering to enhance the validity of the source data. (2) Furthermore, in order to quantitatively evaluate the operation efficiency of the system, it is necessary to further verify the mechanism proposed in this paper. (3) A privacy data protection mechanism for a traceable system involving multiple functionalities can be established to realize the safe sharing of traceable information.

## Data Availability

The data used to support the findings of this study have not been made available because of commercial confidentiality.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the design of this study, analyses, and writing of this manuscript.

## Authors' Contributions

Yuting Wu and Xiu Jin contributed equally to this work. They are co-first authors of the paper.

## Acknowledgments

This research was funded by the Primary Research & Development Plan of Anhui Province, China (201904a06020020), the Project of Anhui Provincial Key Laboratory of Smart Agricultural Technology and Equipment (APKLSA-

TE2019X009), the Key Laboratory of Agricultural Electronic Commerce, Ministry of Agriculture and Rural Affairs (AEC2018011), Primary Research & Development Plan of Anhui Province, China (201904a06020020), the Higher Education Quality Engineering Project of Anhui Province (2020jxtd089 and 2020sjjd040), the First-class Undergraduate Major Construction Project of Anhui Agricultural University (No. 2019aulyzy13), and the Major Science and Technology Projects in Anhui Province (202103b06020013).

## References

- [1] X. Liu, L. Xu, D. Zhu, and L. Wu, "Consumers' WTP for certified traceable tea in China," *British Food Journal*, vol. 117, pp. 1440–1452, 2015.
- [2] M. C. Wang and C. Y. Yang, "Analyzing organic tea certification and traceability system within the Taiwanese tea industry," *Journal of the Science of Food and Agriculture*, vol. 95, pp. 1252–1259, 2015.
- [3] M. M. Aung and Y. S. Chang, "Traceability in a food supply chain: safety and quality perspectives," *Food Control*, vol. 39, pp. 172–184, 2014.
- [4] K. Demestichas, N. Peppes, T. Alexakis, and E. Adamopoulou, "Blockchain in agriculture traceability systems: a review," *Applied Sciences*, vol. 10, 2020.
- [5] F. Tian, "An agri-food supply chain traceability system for China based on RFID & blockchain technology," in *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6, Kunming, 2016.
- [6] R. K. Lomotey, J. C. Pry, and C. Chai, "Traceability and visual analytics for the Internet-of-Things (IoT) architecture," *World Wide Web*, vol. 21, pp. 7–32, 2018.
- [7] J. Lin, Z. Shen, and C. Miao, "Using blockchain technology to build trust in sharing LoRaWAN IoT," in *Proceedings of the 2nd international conference on crowd science and engineering - ICCSE'17*, pp. 38–43, Beijing, China, 2017.
- [8] M. Petersen, N. Hackius, and B. von See, "Mapping the sea of opportunities: blockchain in supply chain and logistics," *IT - Information Technology*, vol. 60, no. 5-6, pp. 263–271, 2018.
- [9] J.-H. Ma, Z. Feng, J.-Y. Wu, Y. Zhang, and W. di, "Learning from imbalanced fetal outcomes of systemic lupus erythematosus in artificial neural networks," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, pp. 1–11, 2021.
- [10] M. Nardi, L. Valerio, and A. Passarella, "Centralised vs decentralised anomaly detection: when local and imbalanced data are beneficial," in *Proceedings of the Third International Workshop on Learning with Imbalanced Domains: Theory and Applications*, PMLR, pp. 7–20, Bilbao (Basque Country, Spain), 2021.
- [11] S. Boutaib, S. Bechikh, F. Palomba, M. Elarbi, M. Makhlof, and L. B. Said, "Code smell detection and identification in imbalanced environments," *Expert Systems with Applications*, vol. 166, article 114076, 2021.
- [12] A. Singh, R. K. Ranjan, and A. Tiwari, "Credit card fraud detection under extreme imbalanced data: a comparative study of data-level algorithms," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 33, pp. 1–28, 2021.
- [13] D. Kim, J. Koo, and U.-M. Kim, "EnvBERT: multi-label text classification for imbalanced, noisy environmental news data," in *2021 15th International Conference on Ubiquitous*

- Information Management and Communication (IMCOM)*, pp. 1–8, Seoul, South Korea, 2021.
- [14] A. Sen, M. M. Islam, K. Murase, and X. Yao, “Binarization With Boosting and Oversampling for Multiclass Classification,” *IEEE Trans Cybern*, vol. 46, pp. 1078–1091, 2016.
- [15] Y. Sun, Z. Li, X. Li, and J. Zhang, “Classifier selection and ensemble model for multi-class imbalance learning in education grants prediction,” *Applied Artificial Intelligence*, vol. 35, no. 4, pp. 290–303, 2021.
- [16] D. Pignini and M. J. S. Conti, “NFC-based traceability in the food chain,” *Sustainability*, vol. 9, no. 10, p. 1910, 2017.
- [17] G. Alfian, M. Syafrudin, U. Farooq et al., “Improving efficiency of RFID-based traceability system for perishable food by utilizing IoT sensors and machine learning model,” *Food Control*, vol. 110, article 107016, 2020.
- [18] S. S. Kamble, A. Gunasekaran, and R. Sharma, “Modeling the blockchain enabled traceability in agriculture supply chain,” *International Journal of Information Management*, vol. 52, article 101967, 2020.
- [19] J. F. Galvez, J. C. Mejuto, and J. Simal-Gandara, “Future challenges on the use of blockchain for food traceability analysis,” *Trac-Trends in Analytical Chemistry*, vol. 107, pp. 222–232, 2018.
- [20] M. H. Miraz and A. Mjapa, “Applications of blockchain technology beyond cryptocurrency,” 2018, <https://arxiv.org/abs/1801.03528>.
- [21] S. Ølnes, J. Ubacht, and M. Janssen, “Blockchain in government: benefits and implications of distributed ledger technology for information sharing,” *Government Information Quarterly*, vol. 34, pp. 355–364, 2017.
- [22] M. Andoni, V. Robu, D. Flynn et al., “Blockchain technology in the energy sector: a systematic review of challenges and opportunities,” *Renewable & Sustainable Energy Reviews*, vol. 100, pp. 143–174, 2019.
- [23] J. J. Sikorski, J. Houghton, and M. Kraft, “Blockchain technology in the chemical industry: machine-to-machine electricity market,” *Applied Energy*, vol. 195, pp. 234–246, 2017.
- [24] B. Yong, J. Shen, X. Liu, F. Li, H. Chen, and Q. Zhou, “An intelligent blockchain-based system for safe vaccine supply and supervision,” *International Journal of Information Management*, vol. 52, article 102024, 2020.
- [25] Y. Liao and K. Xu, *Traceability System of Agricultural Product Based on Block-Chain and Application in Tea Quality Safety Management*, IOP Publishing, 2019.
- [26] P. W. Khan, Y. C. Byun, and N. Park, “IoT-blockchain enabled optimized provenance system for food Industry 4.0 using advanced deep learning,” *Sensors (Basel)*, vol. 20, 2020.
- [27] G. Zhao, S. Liu, C. Lopez et al., “Blockchain technology in agri-food value chain management: a synthesis of applications, challenges and future research directions,” *Computers in Industry*, vol. 109, pp. 83–99, 2019.
- [28] X. Zhang, P. Sun, J. Xu et al., “Blockchain-based safety management system for the grain supply chain,” *IEEE Access*, vol. 8, pp. 36398–36410, 2020.
- [29] Y. Liu, Y. Gu, J. C. Nguyen et al., “Symptom severity classification with gradient tree boosting,” *Journal of biomedical informatics*, vol. 75, pp. S105–S111, 2017.
- [30] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, “Boosting methods for multi-class imbalanced data classification: an experimental review,” *Journal of Big Data*, vol. 7, 2020.
- [31] Y. Cao, F. Jia, and G. Manogaran, “Efficient traceability systems of steel products using blockchain-based industrial Internet of Things,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6004–6012, 2020.
- [32] Q. Ding, S. Gao, J. Zhu, and C. Yuan, “Permissioned blockchain-based double-layer framework for product traceability system,” *IEEE Access*, vol. 8, pp. 6209–6225, 2020.
- [33] Q. Ding, S. Gao, J. Zhu, and C. Yuan, “Food safety traceability system based on blockchain and EPCIS,” *IEEE Access*, vol. 7, pp. 20698–20707, 2019.
- [34] J. Song, X. Lu, and X. Wu, “An improved AdaBoost algorithm for unbalanced classification data,” in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 109–113, Tianjin, China, 2009.
- [35] P. Chen and C. Pan, “Diabetes classification model based on boosting algorithms,” *BMC Bioinformatics*, vol. 19, p. 109, 2018.
- [36] J. Błaszczyński and J. Stefanowski, “Neighbourhood sampling in bagging for imbalanced data,” *Neurocomputing*, vol. 150, pp. 529–542, 2015.
- [37] L. Rokach, “Ensemble-based classifiers,” *Artificial Intelligence Review*, vol. 33, no. 1–2, pp. 1–39, 2010.
- [38] C. Zhang and Y. Ma, *Ensemble Machine Learning: Methods and Applications*, Springer, 2012.
- [39] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [40] Y. Freund and R. E. Schapire, *Experiments with a new boosting algorithm*, Citeseer, 1996.
- [41] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [42] W. Wang, T. Li, W. Wang, and Z. Tu, “Multiple fingerprints-based indoor localization via GBDT: subspace and RSSI,” *IEEE Access*, vol. 7, pp. 80519–80529, 2019.
- [43] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” *Annals of statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- [44] A. Singh, R. M. Parizi, Q. Zhang, K. K. Choo, and A. Dehghantanha, “Blockchain smart contracts formalization: approaches and challenges to address vulnerabilities,” *Computers & Security*, vol. 88, article 101654, 2020.
- [45] F. Provost and T. J. M. L. Fawcett, “Robust classification for imprecise,” *Environments*, vol. 42, no. 3, pp. 203–231, 2001.
- [46] X.-H. Zhou, D. K. McClish, and N. A. Obuchowski, *Statistical Methods in Diagnostic Medicine*, John Wiley & Sons, 2009.
- [47] F. D. Ribeiro, F. Caliva, M. Swainson, K. Gudmundsson, G. Leontidis, and S. Kollias, “An adaptable deep learning system for optical character verification in retail food packaging,” in *2018 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, pp. 1–8, Rhodes, Greece, 2018.
- [48] K. Salah, N. Nizamuddin, R. Jayaraman, and M. Omar, “Blockchain-based soybean traceability in agricultural supply chain,” *IEEE Access*, vol. 7, pp. 73295–73305, 2019.
- [49] Z. Shahbazi and Y.-C. Byun, “A procedure for tracing supply chains for perishable food based on blockchain, machine learning and fuzzy logic,” *Electronics*, vol. 10, no. 1, 2021.