IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 22, 2021, accepted February 24, 2021, date of publication February 26, 2021, date of current version March 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3062410

Effective Management for Blockchain-Based Agri-Food Supply Chains Using Deep Reinforcement Learning

HUILIN CHEN[1,2], ZHEYI CHEN[3], FEITING LIN[1,4], AND PEIFEN ZHUANG[1]
[1]College of Economics, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2]College of Economics and Trade, Fujian Jiangxia University, Fuzhou 350108, China
[3]College of Engineering, Mathematics, and Physical Sciences, University of Exeter, Exeter EX4 4QF, U.K.
[4]School of Economics and Management, Minjiang University, Fuzhou 350108, China

Corresponding author: Peifen Zhuang (peifenzhuang@fafu.edu.cn)

ABSTRACT In agri-food supply chains (ASCs), consumers pay for agri-food products produced by farmers. During this process, consumers emphasize the importance of agri-food safety while farmers expect to increase their profits. Due to the complexity and dynamics of ASCs, the effective traceability and management for agri-food products face huge challenges. However, most of the existing solutions cannot well meet the requirements of traceability and management in ASCs. To address these challenges, we first design a blockchain-based ASC framework to provide product traceability, which guarantees decentralized security for the agri-food tracing data in ASCs. Next, a Deep Reinforcement learning based Supply Chain Management (DR-SCM) method is proposed to make effective decisions on the production and storage of agri-food products for profit optimization. The extensive simulation experiments are conducted to demonstrate the effectiveness of the proposed blockchain-based framework and the DR-SCM method under different ASC environments. The results show that reliable product traceability is well guaranteed by using the proposed blockchain-based ASC framework. Moreover, the DR-SCM can achieve higher product profits than heuristic and Q-learning methods.

INDEX TERMS Agri-food supply chains, agri-food safety, product traceability, profit optimization, blockchain, deep reinforcement learning.

I. INTRODUCTION

IN recent years, the problems of agri-food safety and farmer income have received widespread attentions [1]–[4]. The issues of agri-food safety may occur in each part of agri-food supply chains (ASCs), while inefficient management strategies of ASCs would lead to low profits. However, many factors may constrain the normal work of ASCs. First, due to the complex structure of ASCs, it is hard to record the full circulation information of agri-food products while ensuring that the information will never be tampered with. Second, the shift of consumer preferences has become the main barrier of precisely determining the production and storage of agri-food products with the consideration of profit maximization. Such uncertainties and dynamics undoubtedly increase the toughness of designing an efficient ASC framework. To address these problems, the effective traceability and management for agri-food products in ASCs have become urgently necessary [5], [6].

On one hand, to guarantee the agri-food safety, the information of agri-food products in ASCs including production, processing, storage, distribution, and retail should be collected and recorded when establishing a mechanism of product traceability [7]. However, most of the traditional traceability solutions of ASCs rely on a centralized system maintained by a trusted third party, which may suffer from the potential single-node failure and security threats such as data tampering and information leakage [8]. Blockchain, a distributed, append-only, and tamper-proof ledger, offers an effective architecture for reliable transactions on the Bitcoin network [9] without the control of a centralized third party. Each information recorded in a blockchain should be verified by the majority of participants to reach a global consensus, which ensures the information source

The associate editor coordinating the review of this manuscript and approving it for publication was Shen Yin.

with auditability and transparency. Moreover, there is no need for a blockchain-based traceability solution to connect to a remote cloud data center because it only requires the stable connection among adjacent participants. Therefore, the blockchain technology can be used to realize a reliable product traceability in supply chains, which has recently become a new research direction and attracted many research interests. For example, Tian [10] proposed a traceability system for ASCs with the radio frequency identification (RFID) and blockchain technologies, where RFID-based devices and blockchains are used for collecting and storing data, respectively. Furthermore, the author [11] designed another traceability system for ASCs based on the Hazard Analysis and Critical Control Points (HACCP), blockchain and Internet-of-Things (IoT) technologies. Toyoda *et al.* [12] proposed a blockchain-based product ownership management system (POMS), which can be used to prevent counterfeit products in supply chains. Caro *et al.* [13] developed the AgriBlockIoT, a blockchain-based traceability solution, which can acquire the agri-food data of production and consumption from IoT devices along ASCs. Mao *et al.* [14] designed a blockchain-based credit evaluation system to optimize the supervision and management of food supply chains, which collected the credit via smart contracts and made analysis by using the long short-term memory (LSTM). Lin *et al.* [15] proposed an information and communications technology (ICT) based system by integrating the blockchain technology. Tse *et al.* [16] discussed the application of the blockchain technology to the food supply chain and made comparisons with traditional traceability systems. Abeyratne and Monfared [17] proposed that the application of blockchains can help raise trust levels of supply chains by using transparent and traceable transactions.

On the other hand, to enhance the farmer income (i.e. product profits), ASC systems are expected to allocate production and storage properly according to the market demands for agri-food products [18]. However, it would be a highly-challenging task to continually make decisions for profit optimization in complex and dynamic environments of ASCs [19]. Many classic solutions for ASC management are based on heuristics [20]–[24], game theory [25]–[28], and control theory [29]–[32]. For example, Dwivedi *et al.* [21] proposed a mixed integer nonlinear programming (MINLP) model to optimize ASCs with the consideration of carbon emissions, where two meta-heuristic algorithms are used to allocate vehicles and choose orders. Kocaoglu *et al.* [22] developed a heuristic-based hybrid algorithm to reduce the delivery costs and computational time in the supply chain management. Raj *et al.* [26] designed a generalized analytical model for sustainable supply chains by using a two-stage Stackelberg game-theoretic method. Halat and Hafezalkotob [27] adopted a Stackelberg game approach to optimize the carbon regulation policies in inventory decisions of a multi-stage green supply chain. Wu and Chen [30] utilized the control theory to find the optimal advertising strategy with the coordination of a supply chain under competitive environments. Zhao and Wang [31] proposed a feedback-control based strategy for controlling inventory in hybrid supply chain with uncertain orders. However, these classic solutions are commonly developed for a specific application scenario. Therefore, they might be feasible in a static environment, but it would be hard for them to fit in complex and dynamic environments of ASCs.

As a branch of machine learning (ML), reinforcement learning (RL) [33] can be used to handle the complicated problem of ASC management by an adaptive and flexible way. Although the existing RL-based methods can solve the problem of ASC management to some extent [34]–[36], most of them adopt the traditional value-based RL algorithms (e.g. Q-learning [33]). Therefore, these algorithms may not be able to efficiently deal with a high-dimensional state space because they calculate and record the Q-values of each action under all states. To address this problem, a deep Q-networks (DQN) algorithm [37] was proposed by combining deep neural networks (DNNs) [38]. In light of the DQN algorithm's advantages, such deep reinforcement learning (DRL) has great potentials to well address the problem of ASC management.

In response to the problems of agri-food safety and farmer income, we propose an effective traceability and management solution for ASCs based on the blockchain and DRL technologies. To the best of our knowledge, the integration of the blockchain and DRL technologies in ASCs is still an untapped but worthy research area. The main contributions of this work are summarized as follows.

- A blockchain-based framework is designed to guarantee agri-food safety with product traceability in ASC systems, where the proof-of-work (PoW) is utilized to ensure the global consensus so that the tracing data is consistent, unique, and cannot be falsified. Therefore, the proposed framework can ensure a decentralized security for the agri-food tracing data in ASCs.
- A Deep Reinforcement learning based Supply Chain Management (DR-SCM) method is proposed to determine the amount of production and storage in order to achieve higher profits. The DR-SCM can autonomously attain the policy of ASC management by interacting with complex and dynamic environments of ASCs. Notably, the problem of high-dimensional state space is well addressed by introducing DNNs.
- The extensive simulation experiments are conducted to verify the effectiveness of the proposed DR-SCM method in blockchain-based ASC systems. The results show that the DR-SCM can achieve higher product profits compared to heuristic and Q-learning methods under different ASC environments.

The rest of this paper is organized as follows. In Section II, the proposed blockchain-based framework for secure product traceability in ASCs is introduced. Section III formulates the problem of ASC management and discusses the proposed DR-SCM method in detail. In Section IV, the proposed solution is evaluated by simulation experiments.

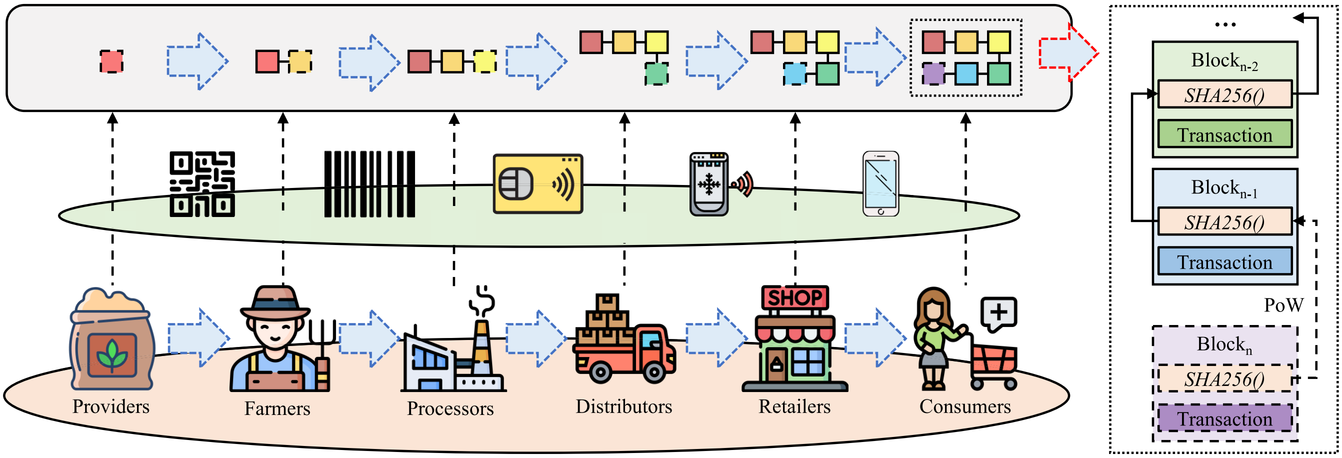


FIGURE 1. The proposed blockchain-based framework for secure agri-food tracing data in ASC systems.

Finally, Section V concludes this paper and looks for future work.

II. BLOCKCHAIN TECHNOLOGY FOR SECURE PRODUCT TRACEABILITY IN ASCs

The original concept of Bitcoin was appeared in 2008 [9], which has since developed into a highly popular decentralized digital currency. Bitcoin is built on a Peer-to-Peer (P2P) network (i.e. the Bitcoin network), which reveals three consensus issues as follows.

- **Realizing synchronization.** Transaction records on different devices will be inconsistent if some devices are disconnected or not running Bitcoin clients. Therefore, it requires synchronization among devices for maintaining the same list of transaction records.
- **Preventing falsification.** Transaction records may be falsified by hackers, which would cause the contradictions of transaction records on different devices and lead to the errors of the Bitcoin network.
- **Avoiding reuse.** A Bitcoin income may be transferred to different users simultaneously. However, due to the different broadcasting routes of a transaction on the Bitcoin network, the order in which transactions are received by different devices might be different. This will result in disagreements of devices on the validity of transaction records.

Therefore, the blockchain technology [9] was proposed to solve the above consensus issues on the Bitcoin network. As a distributed ledger, a blockchain is maintained by the participants on the Bitcoin network. More specifically, a blockchain is made up of many blocks in only one chain, which stores the blocks that have been verified by the majority of participants. A transaction is placed into a new block that is added to the blockchain after a blockchain user completes a PoW task with verification from all the other participants. Thus, the blockchain technology provides a reliable mechanism to secure transactions.

In order to guarantee the security of agri-food tracing data in ASCs, we propose a blockchain-based framework for ASC systems. As shown in Figure 1, through using digital technologies (e.g. QR/bar codes, RFID, NFC, sensors, and mobile devices), the tracing data captured during each transaction in ASCs will be added to a block. After each block is validated by the participants of ASCs and reaching a consensus, the block will be added to the blockchain they maintain and become a secure permanent record. The main components of the proposed framework are described as follows.

- **Providers.** The tracing data includes the information of agri-food raw materials (e.g. seeds, pesticides, and fertilizers), the transactions with farmers, etc.
- **Farmers.** The tracing data includes the information of farms, farming practices, cultivation process, weather conditions, the transactions with providers and processors, etc.
- **Processors.** The tracing data includes the information of factories, processing ways, the transactions with farmers and distributors, etc.
- **Distributors.** The tracing data includes transportation details, storage conditions (e.g. temperature and humidity), the transactions with processors and retailers, etc.
- **Retailers.** The tracing data includes the information of agri-food products (e.g. quality, quantity, price, and expiration dates), storage conditions, the transactions with distributors and consumers, etc.
- **Consumers.** Consumers can use mobile devices to get the detailed information of agri-food products (from providers to retailers).

In the blockchain network, each participant has an opportunity of mining blocks with the successful application of the Proof-of- Work (PoW) [9]. As shown in Figure 1, the PoW requires participants to prove their work by completing a mining task, which is a mathematical puzzle that is extremely difficult to be solved but easy to be verified. Commonly, this

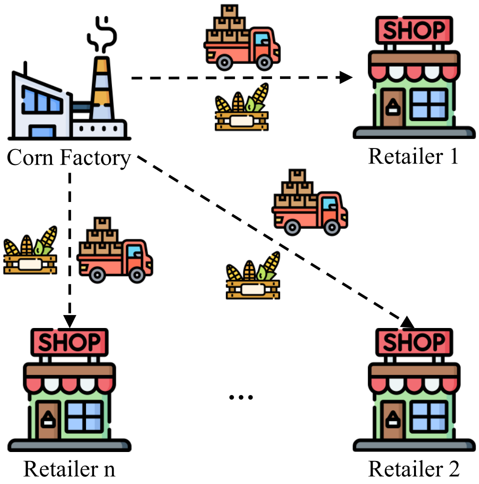


FIGURE 2. **A simplified scenario of ASC management.**

puzzle is defined as

$$Find\ n \quad s.t.\ SHA256(SHA256(h.n)) < target \tag{1}$$

where h represents the contents of a block, n is a random number, and "." is a string concatenate operator. By using the *SHA*256() function [39], a cryptographic hash function, we can get a 256-bit binary number. If this number is smaller than the set *target*, which represents the difficulty of mining, the puzzle is solved successfully.

III. EFFECTIVE ASC MANAGEMENT BASED ON DRL

We formulate the problem of ASC management in order to enhance the farmer income (i.e. product profits). As shown in Figure 2, we consider a simplified scenario that consists of one processor (i.e. a corn factory) with multiple retailers over a fixed number of periods, where the corn is shipped from the factory to retailers through distributors (i.e. trucks). During each period, the amount of corn to be produced and stored at the factory as well as the amount of corn to be shipped to retailers would be decided. For each retailer, there is a seasonal demand for corn. If a retailer cannot meet the demand, a punishment will occur until the demand is satisfied. To make the problem closer to reality, there are limitations on production capability and storage capacity of the factory as well limitations on storage capacity of retailers. Moreover, we assume that the demand may exceed the production capability of the factory, and thus there should be enough stock of corn at retailers. In response to this problem, both the factory and retailers should be able to efficiently rebuild stock according to the demands of corn.

Specifically, there are one factory, denoted by v_0, and multiple retailers, denoted by $L = \{v_1, v_2, \ldots, v_n\}$. For the clarity of presentation, they are integrated into a set, denoted by $V = \{v_0, v_1, \ldots, v_n\}$. The major symbols used in the problem formulation are defined in Table 1.

TABLE 1. **Major symbols used in the problem formulation.**

Symbol	Definition
V	Set of a factory and multiple retailers
TO	Total turnover of selling corn at retailers
p	Unit price of corn
d_i	Corn demand of a retailer
$d_{i,t}$	Corn demand of a retailer at a time period
d_{max}	Predefined maximum demand at a retailer
$sto_{i,t}$	Stochastic factor
C_{prod}	Production cost of corn
c_{pr}	Unit cost of producing corn at the factory
a_0	Production level of the factory
C_{trans}	Transportation cost of corn
$c_{tr,i}$	Truck cost of shipping corn to a retailer
t_cap_i	Truck capacity to a retailer
a_i	Transportation volume to a retailer
C_{store}	Storage cost of corn
$c_{st,i}$	Cost of storing corn at a factory or retailer
s_i	Stock level of a factory or retailer
C_{punish}	Punishment cost of dissatisfying demands
c_{pu}	Unit punishment cost
$Profits$	Net profits of producing and selling corn

First, the total turnover obtained by selling corn at retailers is defined as

$$TO = p \cdot \sum_{i=1}^{n} d_i \tag{2}$$

where p is the unit price of corn, $d_i = \sum_{t=1}^{T} d_{i,t}$ is the corn demand of a retailer, and T is the number of time periods. Specifically, we consider a time period of 12 months for ASC management. Meanwhile, the application of vacuum packs and frozen storage promises that the corn is fresh for over 12 months. Therefore, the shelf life of the corn can be neglected under this situation. As mentioned before, there is a seasonal demand for corn at different retailers, and thus the corn demand of a retailer at a time period is defined as

$$d_{i,t} = \lfloor \frac{d_{max}}{2} + \frac{d_{max}}{2} \cdot \sin\left(\frac{(2i+t)\cdot\pi}{6}\right) + sto_{i,t} \rfloor \tag{3}$$

where d_{max} is the predefined maximum demand at a retailer. $sto_{i,t}$ is the stochastic factor that is randomly assigned a value of 0 or 1 at a retailer over different time periods, which is used to simulate the sudden increase in demand.

Next, the production cost of corn is defined as

$$C_{prod} = c_{pr} \cdot a_0 \tag{4}$$

where c_{pr} and a_0 are the unit cost of producing corn and the production level of the factory, respectively.

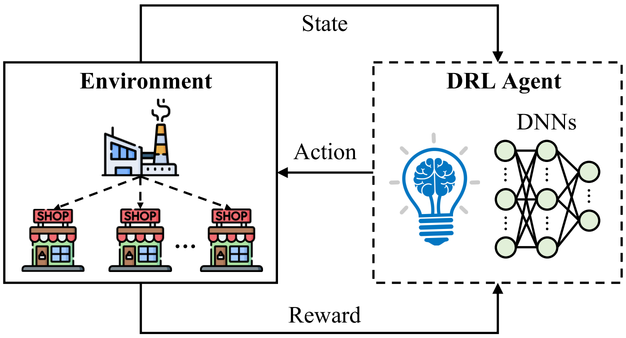


FIGURE 3. Framework of the proposed DR-SCM method.

Then, the transportation cost of corn is defined as

$$C_{trans} = \sum_{i=1}^{n}(c_{tr,i} \cdot \lceil \frac{a_i}{t\_cap_i} \rceil) \tag{5}$$

where $c_{tr,i}$ is the truck cost of shipping corn to a retailer and t_cap_i is the truck capacity to a retailer. a_i is the transportation volume to a retailer, and the total volume to all retailers should be less than the stock of the factory.

Moreover, the storage cost of corn is defined as

$$C_{store} = \sum_{i=0}^{n}(c_{st,i} \cdot s_i) \tag{6}$$

where $c_{st,i}$ and s_i are the cost of storing corn and the stock level of a factory or retailer, respectively.

Besides, the punishment cost occurs when a retailer cannot meet the demand of corn, which is defined as

$$C_{punish} = c_{pu} \cdot \sum_{i=1}^{n}(d_i - s_i) \tag{7}$$

where c_{pu} is the unit punishment cost.

Therefore, net profits of producing and selling corn in ASCs can be calculated by

$$Profits = TO - C_{prod} - C_{trans} - C_{store} - C_{punish} \tag{8}$$

After formulating the problem of ASC management, we propose a Deep Reinforcement learning based Supply Chain Management (DR-SCM) method. The DR-SCM can be used to efficiently determine the production and storage of corn for optimizing product profits. As shown in Figure 3, a scenario of ASC management is regarded as the environment, and the Deep Reinforcement Learning (DRL) agent takes actions by interacting with the environment. Moreover, we define the state space, action space, and reward function for the proposed DR-SCM method as follows.

- **State space.** The state at the time period t is defined as $s_t = [s_0, s_1, s_2, \ldots, s_n, d_t]$, where s_0 and $(s_1, s_2, \ldots, s_n)$ are the stock levels of the factory and retailers, respectively, which are within the maximum stock capacity s_cap_i. $d_t = [d_{1,t}, d_{2,t}, \ldots, d_{n,t}]$, defined in Eq. (3), represents the demands of different retailers at the time period t.
- **Action space.** The action taken by the DRL agent is to make decisions for the production of the factory and the amount of products shipped to retailers. Therefore, the action at the time period t is defined as $a_t = [a_0, a_1, a_2, \ldots, a_n]$, where $a_0 = \{0, 1, \ldots, \beta | \beta \in \mathcal{N}\}$ is the production level of the factory, which is limited by the predefined maximum production level β. $(a_1, a_2, \ldots, a_n)$ is the amount of products shipped to each retailer, which is less than the stock of the factory, denoted by $\sum_{i=1}^{n} a_i \leq s_0$.
- **Reward function.** The reward function is used to guide the DRL agent to learn the optimized policy of ASC management with higher rewards, aiming to enhance product profits. Therefore, the reward function at the time period t should follow a positive increment of net profits (defined in Eg. (8)) as $r_t = Profits$.

A. Q-LEARNING ALGORITHM

As a classic RL algorithm, the Q-learning [33] learns policies by recording and utilizing the rewards of each state-action pair, denoted by $Q(s, a)$. For each timestep, the RL agent calculates and records each $Q(s, a)$ in a Q-table. $Q(s, a)$ can be used as a long-term reward, and it is defined as

$$Q'(s, a) = r + \gamma \cdot max_{a'} Q(s', a') \tag{9}$$

where s and a are the current state and action. s' and a' are the next state and action. γ $(0 \leq \gamma \leq 1)$ is the discount factor. When γ is close to 0, the RL agent tends to focus on the immediate reward. By contrast, when γ is close to 1, the RL agent tends to consider the future reward.

In the Q-Table, $Q(s, a)$ is updated by

$$Q(s, a) = Q(s, a) + \alpha \cdot (Q'(s, a) - Q(s, a)) \tag{10}$$

where α is the learning rate.

Next, the actions will be chosen by using the ϵ-greedy algorithm. That is, there is a probability of ϵ that the action with the greatest value of $Q(s, a)$ in the current Q-Table will be chosen. Otherwise, a random strategy will be taken for choosing actions. Thus, the Q-learning may be able to gradually find the optimal policy of ASC management. The main steps of the Q-learning are shown in Algorithm 1.

B. DR-SCM METHOD

Although the Q-learning can solve the problem of ASC management to some extent, it may not be able to efficiently deal with a high-dimensional state space. This is because the Q-learning calculates and records the Q-values of each action under all states in a Q-table, and thus the matrix of $Q(s, a)$ would become large, which may cause the algorithm crashed due to memory overflow. To address this problem, a DQN algorithm was proposed in [37] that utilizes DNNs [38] to estimate $Q(s, a)$ rather than calculates a Q-table. In light of the DQN algorithm's advantages, we propose the DR-SCM

Algorithm 1 The Q-Learning for ASC Management

Input: the state space $\mathcal{S}$, action space $\mathcal{A}$, discount factor γ, and learning rate α.

Output: the policy of ASC management, denoted by $\pi(s) = argmax_{a\in|\mathcal{A}|}Q(s,a)$.

```
Initialize $Q(s,a)$ randomly;
for each episode do
    Initialize a state $s$;
    for each timestep of episode do
        Choose an action $a$ under the state $s$ by $\epsilon$-greedy;
        Execute the action $a$, receive the reward $r$, and
          observe a new state $s'$:
          $Q(s,a) \leftarrow Q(s,a) + \alpha \cdot [r + \gamma \cdot max_{a'}Q(s',a')$
          $-Q(s,a)]$;
        Update state: $s \leftarrow s'$;
        if the state $s$ is terminal then
            Break;
        end if
    end for
    if $\forall s, a, Q(s,a)$ converges then
        Break;
    end if
end for
```

method to autonomously obtain the policy of ASC management by interacting with complex and dynamic environments of ASCs. The main steps of the DR-SCM method are shown in Algorithm 2.

Compared to the Q-Learning, the improvement of the DR-SCM is introducing the mechanism of experience replay, which stores the acquired transitions in memory. Commonly, the transitions constructed in chronological order are highly-correlated and non-stationary, which may cause great difficulties in training convergence. Through the random sampling in experience replay and introducing the target network, the sample correlation and model fluctuation can be removed to a certain extent, which makes the algorithm more stable and easier to converge.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed DR-SCM method in different blockchain-based ASC environments and make comparisons with the other two classic methods.

A. SIMULATION SETTINGS

The simulation environment is built on Windows 10 64-bit with Intel® Core^{TM} i7 CPU @2.30 GHz and RAM 8.00 GB DDR4. We simulate three different ASC scenarios based on Python 3.6 referring to real-world data settings [40]. In general, there is one factory in different ASC scenarios with the fixed maximum production level $\beta = 20$, the unit cost of producing corn $c_{pr} = 1k$ RMB/ton, the storage capacity $s_0 = 100$ tons, and no storage cost. From the perspectives of turnover, we set the unit price of corn as $p = 2.3k$ RMB/ton.

Algorithm 2 The Proposed DR-SCM Method

Input: the state space $\mathcal{S}$, action space $\mathcal{A}$, discount factor γ, and learning rate α.

Output: the Q-networks $Q_\phi(s,a)$ of ASC management.

```
Initialize the experience replay $\mathcal{D}$ with the capacity $N$;
Initialize the parameters $\phi$ of the Q-networks randomly;
Initialize the parameters $\hat{\phi} = \phi$ of the target Q-networks
  randomly;
for each episode do
    Initialize a state $s$;
    for each timestep of episode do
        Choose an action $a$ under the state $s$ by $\epsilon$-greedy;
        Execute the action $a$, receive the reward $r$, and
          observe a new state $s'$:
        Store the transition $(s,a,r,s')$ in $\mathcal{D}$;
        Randomly sample a mini-batch of transitions
          $(s^*,a^*,r^*,s^{*\prime})$ from $\mathcal{D}$;
        $y = \begin{cases} r^*, & terminal \\ r^* + \gamma \cdot max_{a'}Q_{\hat{\phi}}(s^{*\prime},a'), & otherwise \end{cases}$;
        Perform gradient descent on $(y - Q_{\hat{\phi}}(s^*,a^*))^2$;
        Update state: $s \leftarrow s'$;
        Update parameters every $C$ timesteps: $\hat{\phi} \leftarrow \phi$;
        if the state $s$ is terminal then
            Break;
        end if
    end for
    if $\forall s, a, Q_\phi(s,a)$ converges then
        Break;
    end if
end for
```

TABLE 2. Scenario 1: A simple scenario with a single retailer.

Retailers	Retailer 1	Retailer 2	Retailer 3
	$c_{st} = 0.25k$ RMB/ton, s=25 tons	×	×
Trucks	Truck 1	Truck 2	Truck 3
	$c_{tr} = 0.5k$ RMB/truck	×	×

As for each retailer, we set the maximum corn demand as $d_{max} = 10$ tons (defined in Eq. (3), the demand changes seasonally), the number of time periods as $T = 12$, and the unit punishment cost as $c_{pu} = 1k$ RMB/ton (defined in Eq. (7), the punishment occurs when the demand cannot be satisfied). As for the capacity of a truck t_{cap}, we fix it as 5 tons. Other detailed settings of different ASC scenarios are described as follows.

- **Scenario 1:** A simple scenario with a single retailer. As shown in Table 2, there is only one retailer (i.e Retailer 1) with the cost of storing corn $c_{st} = 0.25k$ RMB/ton and the storage capacity $s = 25$ tons. Moreover, the truck cost of shipping corn from the factory to Retailer 1 by Truck 1 is set as $c_{tr} = 0.5k$ RMB/truck.
- **Scenario 2:** A typical scenario with three retailers and same settings. As shown in Table 3, there are three retailers (i.e Retailer 1$\sim$3) with the same cost of storing corn

$c_{st} = 0.25k$ RMB/ton and the same storage capacity $s =$ 25 tons. Moreover, the truck cost of shipping corn from the factory to Retailer 1∼3 by Truck 1∼3 is identically set as $c_{tr} = 0.5k$ RMB/truck.

- **Scenario 3:** A complex scenario with three retailers and different settings. As shown in Table 4, there are three retailers (i.e Retailer 1∼3) with different costs of storing corn $c_{st} = (0.20, 0.25, 0.30)k$ RMB/ton and different storage capacities $s = (20, 25, 30)$ tons. Moreover, the truck cost of shipping corn from the factory to Retailer 1∼3 by Truck 1∼3 is differently set as $c_{tr} = (0.4, 0.5, 0.6)k$ RMB/truck.

Based on Scenario 3, RaspberryPi micro computers are used to play the roles of retailers and trucks for verifying the effectiveness of the proposed blockchain-based framework for ASCs. The hashing operations in PoW are executed based on Python's library of *hashlib.sha*256(), where the setting of PoW difficulty is referred to [41]. Moreover, the tracing data used is randomly generated, and the number of transactions per block is randomly distributed in [50, 500].

Moreover, the proposed DR-SCM method is implemented based on TensorFlow 1.4.0 [42], and NumPy is used to offer mathematical functions for matrix calculations. In the DR-SCM, two hidden layers are used in DNNs with 200 and 100 neurons, respectively. Furthermore, we set the number of episodes as 10000, the mini-batch size as 128, the discount factor as $\gamma = 0.95$, the learning rate as $\alpha = 0.001$, and the capacity of experience replay as $N = 500$.

Besides, we evaluate the performance of the other two classic methods for ASC management, including heuristic [23] and Q-learning [34] methods, to conduct comparative experiments. In the heuristic method, the stock of retailers will be replenished if the current stock is less than a predefined threshold (i.e. half of storage capacity at retailers). The Q-learning, described in Section III-A, records the Q-values of each state-action pair into a Q-table for maximizing long-term rewards, but it might suffer from the problem of high-dimensional state space.

B. EXPERIMENTAL RESULTS

Based on the simulation settings, we first evaluate the effectiveness of the proposed blockchain-based framework for secure tracing data in ASCs. Figure 4 illustrates the average computational cost of hashes with different numbers of blocks in the blockchain. As we can see from the results, the PoW tasks in the blockchain network consumes large computational resources when executing hashing operations. This mechanism ensures the global consensus so that the tracing data is consistent, unique and cannot be falsified. Therefore, the proposed blockchain-based framework can guarantee a decentralized security for the agri-food tracing data in ASCs.

Next, we evaluate the performance of the proposed DR-SCM method for ASC management, where the DR-SCM is compared with two classic methods, including heuristic and Q-learning methods. Figure 5 presents the rewards (i.e. profits) of different methods for ASC management in various scenarios. In general, the rewards of different methods increase as the number of episodes grows except the heuristic method. This is because the heuristic method always uses a predefined threshold to control the operations of ASC management, and thus it cannot learn more optimized policy of ASC management during the training process. Compared to Scenario 2 and Scenario 3, the rewards achieved in Scenario 1 is much lower. This is because there is only one retailer in Scenario 1 but there are three retailers in Scenario 2 and Scenario 3, which leads to the difference in turnover. The performance of the Q-learning always seems good in various scenarios, but it might fall into the local optimum and occur the fluctuation of rewards during the training process. By contrast, the DR-SCM always outperforms other methods for ASC management and presents a more stable training process in various scenarios. More specifically, compared to Scenario 2, Scenario 3 become more complex with changeable situations of retailers and trucks. Under this condition, the DR-SCM can still achieve better rewards and stability than the heuristic and Q-learning methods, which also demonstrates the high adaptiveness of the proposed method.

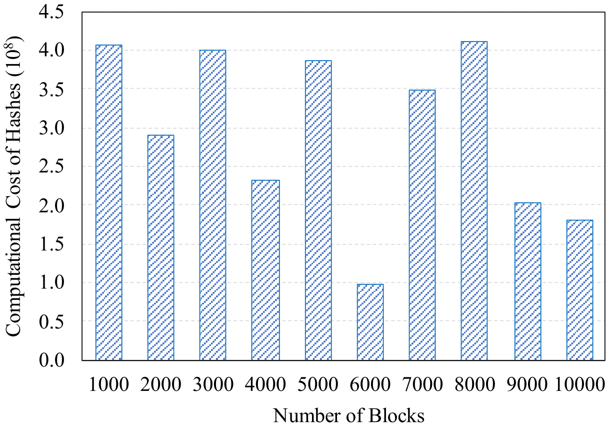


FIGURE 4. Computational cost of hashes with different numbers of blocks.

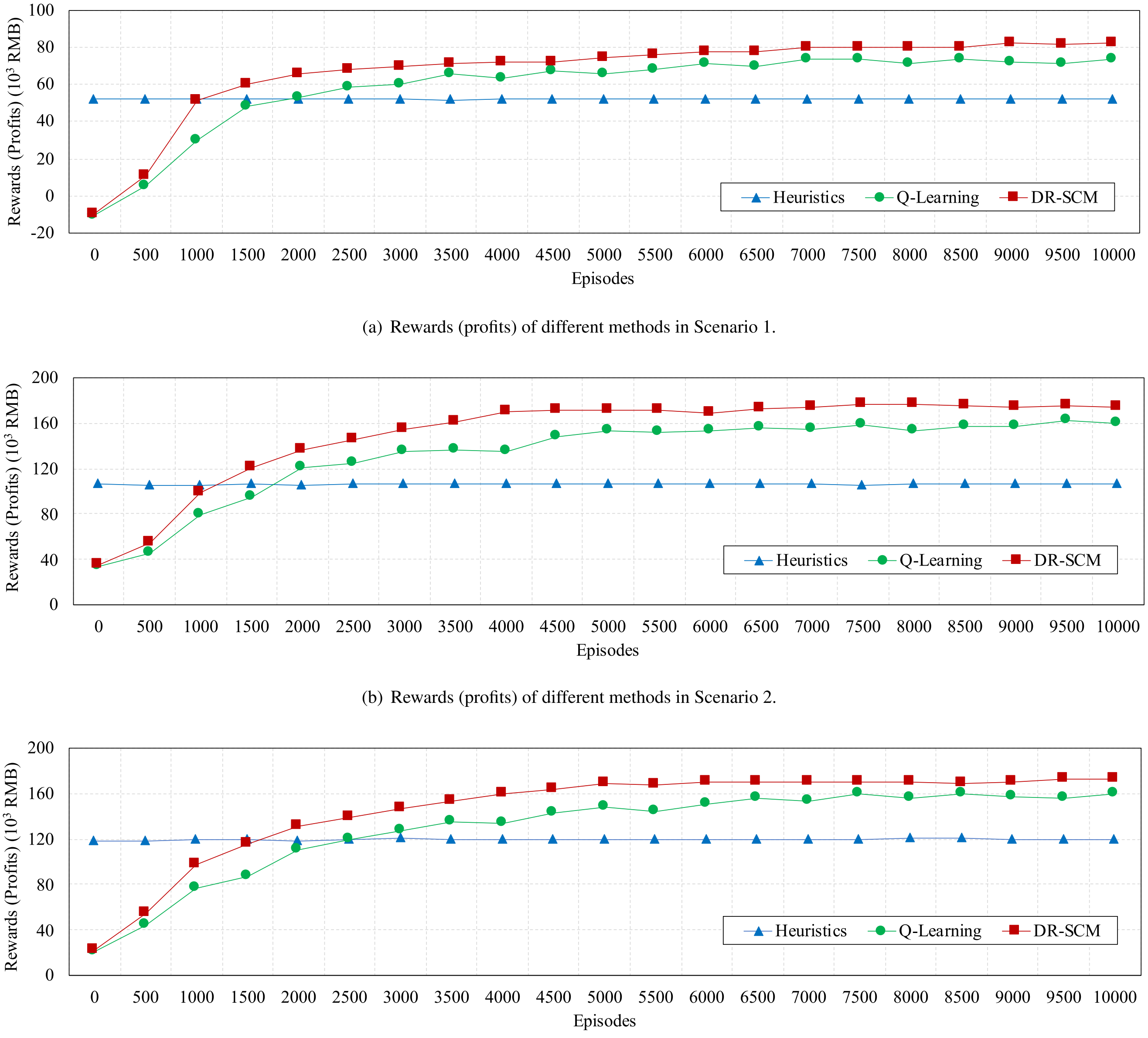
Also, we compare the stock of factory and retailers by using the proposed DR-SCM and the heuristic method in Scenario 3 over different time periods. As shown in Figure 6(a), the DR-SCM can keep the stock of the factory available for retailers during most of the time periods, and thus the stock of retailers can remain positive basically. This because the DR-SCM can flexibly adjust the production and storage levels at factory and retailers in response to the seasonally changeable demands over different time periods. In contrast, as shown in Figure 6(b), the heuristic method always keep the stock of the factory at a stable level, which may meet the demand from retailers in some cases but cannot work well when demands increase. This flaw leads to the negative stock of retailers, which means that demands cannot be satisfied and thus the profits will be reduced. The results highlight the

TABLE 3. Scenario 2: A typical scenario with three retailers and same settings.

Retailers	Retailer 1	Retailer 2	Retailer 3
	$c_{st} = 0.25k$ RMB/ton, s=25 tons	$c_{st} = 0.25k$ RMB/ton, s=25 tons	$c_{st} = 0.25k$ RMB/ton, s=25 tons
Trucks	Truck 1	Truck 2	Truck 3
	$c_{tr} = 0.5k$ RMB/truck	$c_{tr} = 0.5k$ RMB/truck	$c_{tr} = 0.5k$ RMB/truck

TABLE 4. Scenario 3: A complex scenario with three retailers and different settings.

Retailers	Retailer 1	Retailer 2	Retailer 3
	$c_{st} = 0.20k$ RMB/ton, s=20 tons	$c_{st} = 0.25k$ RMB/ton, s=25 tons	$c_{st} = 0.30k$ RMB/ton, s=30 tons
Trucks	Truck 1	Truck 2	Truck 3
	$c_{tr} = 0.4k$ RMB/truck	$c_{tr} = 0.5k$ RMB/truck	$c_{tr} = 0.6k$ RMB/truck

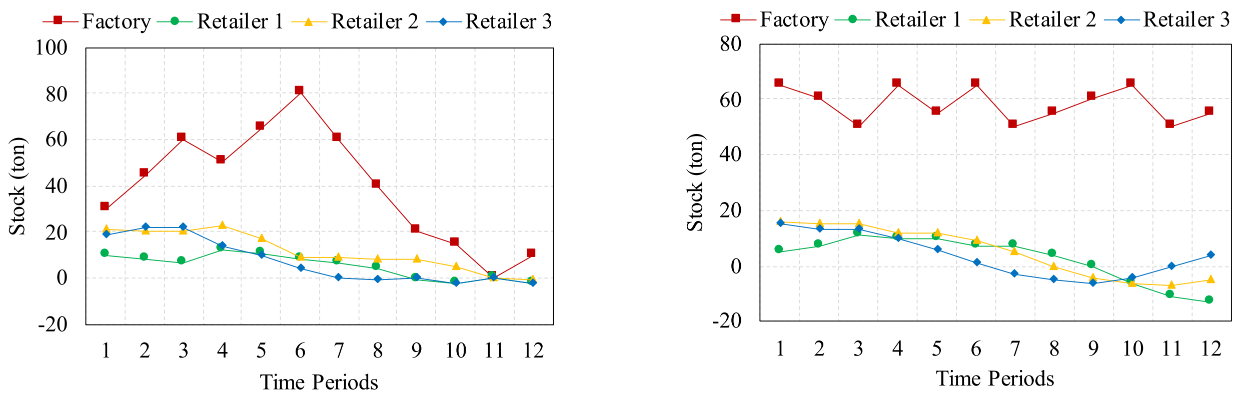


(a) Rewards (profits) of different methods in Scenario 1.

(b) Rewards (profits) of different methods in Scenario 2.

(c) Rewards (profits) of different methods in Scenario 3.

FIGURE 5. Performance comparison among different methods for ASC management in various scenarios.



(a) Stock of factory and retailers by using the DR-SCM.

(b) Stock of factory and retailers by using the heuristic method.

FIGURE 6. Stock comparison between the DR-SCM and heuristic methods in Scenario 3 over different time periods.

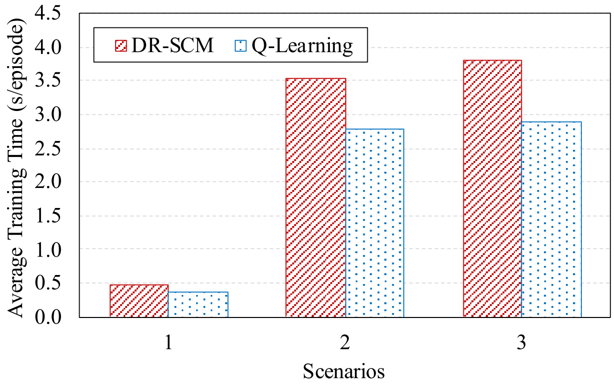


FIGURE 7. Average training time of the DR-SCM and Q-learning in different scenarios.

advantages of the DR-SCM in handling the complex problem of ASC management in dynamic environments.

Finally, we compare the training efficiency of the proposed DR-SCM method and the Q-learning in different scenarios. As shown in Figure 7, the values of average training time of the above methods in different scenarios are recorded. In Scenario 1, both the DR-SCM and Q-learning consume low average training time due to the simple settings of this scenario. When it comes to more complex environments (i.e. Scenario 2 and Scenario 3), these two methods consume much higher average training time because it would be harder for them to find the optimized policy of ASC management. Although the average training time of the Q-learning is slightly less than the DR-SCM in each scenario (because there is no DNN structure in the Q-learning), the rewards obtained by using the Q-learning is obviously lower than the DR-SCM, as analyzed in the aforementioned experiments. Therefore, the DR-SCM can achieve a better trade-off between profit optimization and learning efficiency than the Q-learning in different scenarios of ASC management.

V. CONCLUSION AND FUTURE WORK

In this paper, we first design a blockchain-based framework to guarantee the agri-food safety with product traceability in ASC systems. Next, we propose a DR-SCM method to make decisions on the production and storage of agri-food products for optimizing product profits in ASCs. The extensive simulation experiments verify the effectiveness of the proposed blockchain-based framework and the DR-SCM method for ASC optimization. More specifically, the results show that the proposed blockchain-based ASC framework can well guarantee reliable product traceability. Moreover, the DR-SCM outperforms common heuristic and Q-learning methods in terms of rewards (i.e. product profits) while achieving high learning efficiency in different scenarios of ASC management. Meanwhile, the DR-SCM has higher flexibility than others in arranging production and storage. In the real-world ASC environment, the demands from consumers are changing during different time periods. Based on the simulation experiments conducted by using the DR-SCM, the macro-control for the production and storage of agricultural products can be effectively performed. Thus, according to demands and costs, the production of factories can maintain available for retailers in a cost-effective way while the stock of retailers can well satisfy the demands from consumers.

The DQN algorithm utilizes a mechanism of experience replay to facilitate convergence, but the experience data in the playback memory reveals strong relevance, which may cause the low efficiency of training for achieving the optimal performance. To address this problem, in the future, we will continue our research by applying other advanced DRL-based algorithms (e.g. asynchronous advantage actor-critic) in more complex scenarios of ASC management with the demands constructed by using real-world data. Meanwhile, we will evaluate the robustness and potential improvements by using these algorithms and explore their feasibility in real-world ASC environments.

REFERENCES

[1] D.-Y. Lin, C.-J. Juan, and C.-C. Chang, "Managing food safety with pricing, contracts and coordination in supply chains," *IEEE Access*, vol. 7, pp. 150892–150909, 2019.

[2] H. Fan, "Theoretical basis and system establishment of China food safety intelligent supervision in the perspective of Internet of Things," *IEEE Access*, vol. 7, pp. 71686–71695, 2019.

[3] M. Toledo-Hernández, T. Tscharntke, A. Tjoa, A. Anshary, B. Cyio, and T. C. Wanger, "Hand pollination, not pesticides or fertilizers, increases cocoa yields and farmer income," *Agricult., Ecosyst. Environ.*, vol. 304, Dec. 2020, Art. no. 107160.

[4] J. Himmelstein, A. Ares, D. Gallagher, and J. Myers, "A meta-analysis of intercropping in Africa: Impacts on crop yield, farmer income, and integrated pest management effects," *Int. J. Agricult. Sustainability*, vol. 15, no. 1, pp. 1–10, Jan. 2017.

[5] Y. Dong, Z. Fu, S. Stankovski, S. Wang, and X. Li, "Nutritional quality and safety traceability system for China's leafy vegetable supply chain based on fault tree analysis and QR code," *IEEE Access*, vol. 8, pp. 161261–161275, 2020.

[6] C. Ganeshkumar, M. Pachayappan, and G. Madanmohan, "Agri-food supply chain management: Literature review," *Intell. Inf. Manage.*, vol. 9, no. 2, pp. 68–96, 2017.

[7] Q. Lin, H. Wang, X. Pei, and J. Wang, "Food safety traceability system based on blockchain and EPCIS," *IEEE Access*, vol. 7, pp. 20698–20707, 2019.

[8] H. Feng, X. Wang, Y. Duan, J. Zhang, and X. Zhang, "Applying blockchain technology to improve agri-food traceability: A review of development methods, benefits and challenges," *J. Cleaner Prod.*, vol. 260, Jul. 2020, Art. no. 121031.

[9] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," White Paper, 2008. Accessed: Jun. 26, 2020. [Online]. Available: https://bitcoin.org/bitcoin.pdf

[10] F. Tian, "An agri-food supply chain traceability system for China based on RFID & blockchain technology," in *Proc. 13th Int. Conf. Service Syst. Service Manage. (ICSSSM)*, Jun. 2016, pp. 1–6.

[11] F. Tian, "A supply chain traceability system for food safety based on HACCP, blockchain & Internet of Things," in *Proc. 14th Int. Conf. Service Syst. Service Manage. (ICSSSM)*, Jun. 2017, pp. 1–6.

[12] K. Toyoda, P. T. Mathiopoulos, I. Sasase, and T. Ohtsuki, "A novel blockchain-based product ownership management system (POMS) for anti-counterfeits in the post supply chain," *IEEE Access*, vol. 5, pp. 17465–17477, 2017.

[13] M. P. Caro, M. S. Ali, M. Vecchio, and R. Giaffreda, "Blockchain-based traceability in agri-food supply chain management: A practical implementation," in *Proc. IoT Vertical Topical Summit Agricult. Tuscany (IOT Tuscany)*, May 2018, pp. 1–4.

[14] D. Mao, F. Wang, Z. Hao, and H. Li, "Credit evaluation system based on blockchain for multiple stakeholders in the food supply chain," *Int. J. Environ. Res. Public Health*, vol. 15, no. 8, p. 1627, Aug. 2018.

[15] Y.-P. Lin, J. Petway, J. Anthony, H. Mukhtar, S.-W. Liao, C.-F. Chou, and Y.-F. Ho, "Blockchain: The evolutionary next step for ICT e-agriculture," *Environments*, vol. 4, no. 3, p. 50, Jul. 2017.

[16] D. Tse, B. Zhang, Y. Yang, C. Cheng, and H. Mu, "Blockchain application in food supply information security," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Dec. 2017, pp. 1357–1361.

[17] S. A. Abeyratne and R. P. Monfared, "Blockchain ready manufacturing supply chain using distributed ledger," *Int. J. Res. Eng. Technol.*, vol. 5, no. 9, pp. 1–10, Sep. 2016.

[18] A. M. Ableeva, G. A. Salimova, N. T. Rafikova, I. I. Fazrahmanov, Z. A. Zalilova, T. N. Lubova, G. R. Nigmatullina, I. N. Girfanova, F. F. Farrakhova, and A. M. Hazieva, "Economic evaluation of the efficiency of supply chain management in agricultural production based on multidimensional research methods," *Int. J. Supply Chain Manage.*, vol. 8, no. 1, p. 328, 2019.

[19] J. A. O. Castro and W. A. Jaimes, "Dynamic impact of the structure of the supply chain of perishable foods on logistics performance and food security," *J. Ind. Eng. Manage.*, vol. 10, no. 4, pp. 687–710, 2017.

[20] M. I. Shongwe and C. N. Bezuidenhout, "A heuristic for the selection of appropriate diagnostic tools in large-scale sugarcane supply systems," *AIMS Agricult. Food*, vol. 4, no. 1, pp. 1–26, 2019.

[21] A. Dwivedi, A. Jha, D. Prajapati, N. Sreenu, and S. Pratap, "Meta-heuristic algorithms for solving the sustainable agro-food grain supply chain network design problem," *Modern Supply Chain Res. Appl.*, vol. 2, no. 3, pp. 161–177, Nov. 2020.

[22] Y. Kocaoglu, E. Cakmak, B. Kocaoglu, and A. T. Gumus, "A novel approach for optimizing the supply chain: A heuristic-based hybrid algorithm," *Math. Problems Eng.*, vol. 2020, pp. 1–24, Feb. 2020.

[23] D. Battini, A. Gunasekaran, M. Faccio, A. Persona, and F. Sgarbossa, "Consignment stock inventory model in an integrated supply chain," *Int. J. Prod. Res.*, vol. 48, no. 2, pp. 477–500, Jan. 2010.

[24] A. Samadi, N. Mehranfar, A. M. F. Fard, and M. Hajiaghaei-Keshteli, "Heuristic-based metaheuristics to address a sustainable supply chain network design problem," *J. Ind. Prod. Eng.*, vol. 35, no. 2, pp. 102–117, Feb. 2018.

[25] M. A. N. Agi, S. Faramarzi-Oghani, and Ö. Hazır, "Game theory-based models in green supply chain management: A review of the literature," *Int. J. Prod. Res.*, pp. 1–20, Jun. 2020, doi: 10.1080/00207543.2020.1770893.

[26] A. Raj, I. Biswas, and S. K. Srivastava, "Designing supply contracts for the sustainable supply chain using game theory," *J. Cleaner Prod.*, vol. 185, pp. 275–284, Jun. 2018.

[27] K. Halat and A. Hafezalkotob, "Modeling carbon regulation policies in inventory decisions of a multi-stage green supply chain: A game theory approach," *Comput. Ind. Eng.*, vol. 128, pp. 807–830, Feb. 2019.

[28] N. N. Vasnani, F. L. S. Chua, L. A. Ocampo, and L. B. M. Pacio, "Game theory in supply chain management: Current trends and applications," *Int. J. Appl. Decis. Sci.*, vol. 12, no. 1, pp. 56–97, 2019.

[29] D. Ivanov, S. Sethi, A. Dolgui, and B. Sokolov, "A survey on control theory applications to operational systems, supply chain management, and industry 4.0," *Annu. Rev. Control*, vol. 46, pp. 134–147, Jan. 2018.

[30] Z. Wu and D. Chen, "New optimal-control-based advertising strategies and coordination of a supply chain with differentiated products under consignment contract," *IEEE Access*, vol. 7, pp. 170703–170714, 2019.

[31] W. Zhao and D. Wang, "Simulation-based optimization on control strategies of three-echelon inventory in hybrid supply chain with order uncertainty," *IEEE Access*, vol. 6, pp. 54215–54223, 2018.

[32] V. L. M. Spiegler, A. T. Potter, M. M. Naim, and D. R. Towill, "The value of nonlinear control theory in investigating the underlying dynamics and resilience of a grocery supply chain," *Int. J. Prod. Res.*, vol. 54, no. 1, pp. 265–286, Jan. 2016.

[33] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[34] P. Jinqi, P. Taiyang, and R. Lei, "The supply chain network on cloud manufacturing environment based on COIN model with Q-learning algorithm," in *Proc. 5th Int. Conf. Enterprise Syst. (ES)*, Sep. 2017, pp. 52–57.

[35] L. Kemmer, H. von Kleist, D. de Rochebouët, N. Tziortziotis, and J. Read, "Reinforcement learning for supply chain optimization," in *Proc. Eur. Workshop Reinforcement Learn.*, 2018, pp. 1–9.

[36] A. Habib, M. I. Khan, and J. Uddin, "Optimal route selection in complex multi-stage supply chain networks using SARSA(λ)," in *Proc. 19th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2016, pp. 170–175.

[37] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness, M. Bellemare, A. Graves, M. Riedmiller, A. Fidjeland, G. Ostrovski, and S. Petersen, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, p. 529, 2015.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[39] H. Gilbert and H. Handschuh, "Security analysis of SHA-256 and sisters," in *Proc. 10th Int. Workshop Sel. Areas Cryptogr. (SAC)*. Berlin, Germany: Springer, 2003, pp. 175–193.

[40] *National Bureau of Statistics of China, China Statistical Yearbook*, China Statist. Press, Beijing, China, 2019.

[41] K. J. O'Dwyer and D. Malone, "Bitcoin mining and its energy footprint," in *Proc. 25th IET Irish Signals Syst. Conf. China-Ireland Int. Conf. Inf. Communities Technol. (ISSC /CIICT)*. Edison, NJ, USA: IET, 2014, pp. 280–285.

[42] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, and M. Kudlur, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, vol. 16, 2016, pp. 265–283.

HUILIN CHEN received the B.E. degree from Shanxi University, China, in 2015, and the M.E. degree from Fuzhou University, China, in 2018, both in international economics and trade. She is currently pursuing the Ph.D. degree in agricultural economics and management with Fujian Agriculture and Forestry University, China. She is a Lecturer with the College of Economics and Trade, Fujian Jiangxia University, China. Her research interests include international trade of agricultural products and circulation of agricultural products.

ZHEYI CHEN received the B.Sc. degree from Shanxi University, China, in 2014, and the M.Sc. degree from Tsinghua University, China, in 2017, both in computer science. He is currently pursuing the Ph.D. degree in computer science with the University of Exeter, U.K. He has published over ten research papers in reputable international journals and conferences, such as IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE TRANSACTIONS ON CLOUD COMPUTING, IEEE ACCESS, *Concurrency and Computation: Practice and Experience*, and the IEEE International Conference on Communications. His research interests include cloud/edge computing, resource optimization, deep learning, and reinforcement learning.

FEITING LIN received the B.E. degree in economics from Fujian Normal University, China, in 2004, and the M.E. degree in international economics from Xiamen University, China, in 2007. She is currently pursuing the Ph.D. degree in agricultural economics and management with Fujian Agriculture and Forestry University, China. She is an Associate Professor with the School of Economics and Management, Minjiang Univesity, China. Her research interests include management of agricultural economy, trade of agricultural products, and big data analysis.

PEIFEN ZHUANG received the B.E. degree in international economic cooperation from the Shanghai University of International Business and Economics, China, in 1992, and the M.M. and Ph.D. degrees in agricultural economics and management from Fujian Agriculture and Forestry University, China, in 1999 and 2007, respectively. She is currently a Professor with the College of Economics, Fujian Agriculture and Forestry University. Her research interests include international trade of agricultural products and global value chains.