Big data-driven risk decision-making and safety management in agricultural supply chains

Guanghe Han*, Xin Pan, Xin Zhang

College of Economics and Management, Heilongjiang Bayi Agricultural University, Daqing, China

***Corresponding Author:** Guanghe Han, College of Economics and Management, Heilongjiang Bayi Agricultural University, Daqing 163319, China. Email: hanguanghe1006@163.com

Received: 1 January 2024; Accepted: 4 February 2024; Published: 22 March 2024

RESEARCH ARTICLE

Abstract

In the era of digitization, the integration of big data technologies has become instrumental in advancing agricultural supply chain management and bolstering risk decision-making processes. Agricultural supply chains, critical to ensuring food security and bolstering rural economies, face vulnerabilities stemming from a myriad of internal and external elements, including natural disasters and market dynamics. Consequently, the urgency to adopt effective risk management strategies is paramount. Contemporary studies have explored the utilization of big data in decision-making processes specific to agricultural supply chain risks, predominantly concentrating on preliminary risk prediction and characterization. Nonetheless, there exists a shortfall in comprehensively analyzing the intricate interplay among risk factors and establishing a holistic risk management decision-making framework based on such analyses. This research addresses these deficiencies through two principal investigative components. First, this research explores the analysis of risk factors and their interrelationships in the agricultural supply chain based on a decision tree algorithm with a transition structure. This algorithm enhances decision-makers' understanding of risk factors and their interrelationships, and guide the implementation of effective risk mitigation measures and the formulation of contingency plans. Subsequently, the research constructs a corresponding data-driven multi-criteria decision-making method, assisting managers in balancing different risk management strategies in a volatile supply chain environment, considering costs, benefits, and feasibility to formulate the optimal strategy. The innovation of this research lies in the development of a novel risk analysis tool based on the transition decision tree algorithm. This is the first time that such advanced algorithms are applied to agricultural supply chain risk management, filling a gap in the current research. The outcomes of this study not only contribute to enhancing risk management practices within agricultural supply chains but also offer novel insights and methodological tools that are applicable in research and practices across related domains.

Keywords: big data; agricultural supply chain; risk decision-making; safety management; decision tree; multi-criteria decision-making

Introduction

In the swiftly transforming digital age, big data technology has been recognized as a crucial catalyst propelling advancements across various sectors (Awad *et al.*, 2023; Chandrasekaran *et al.*, 2021; Chennouk *et al.*, 2022; Hasan *et al.*, 2022; He *et al.*, 2022; Kusrini *et al.*, 2022; Lazarevska *et al.*, 2022; Li and Gao, 2022; Zhang *et al.*, 2023; Zhao *et al.*, 2022). The agricultural industry, quintessential for sustaining human life, relies heavily on the efficiency and security of its supply chain. This supply chain's effectiveness is intrinsically connected to food security, income generation for farmers, and societal stability (Nguyen, 2022; Xing and Zhao, 2013). It faces a myriad of risks, primarily because of the unpredictable nature of environmental conditions and market

demands, encompassing threats, such as meteorological disasters, the spread of epidemics, and market volatilities. The management and decision-making processes pertaining to these risks play a pivotal role in ensuring the safety of agricultural production and strengthening the resilience of the supply chain. In this context, big data technology emerges as a transformative tool, providing new insights and approaches (Li *et al.*, 2023; Ye, 2021; Zhai, 2023).

The burgeoning development of information technology has amplified the importance of big data in the realms of risk decision-making and safety management within the agricultural supply chains (Chen *et al.*, 2022; Dai and Liu, 2020; Wang and Wu, 2022). The comprehensive gathering, processing, and analytical examination of large-scale data empower decision-makers to more precisely pinpoint potential risks and formulate more efficacious strategies for risk mitigation (Land and Siraj, 2021; Liu, 2022; Lu *et al.*, 2022; Zhang *et al.*, 2022). Such strategic advancements not only optimize resource distribution and augment the competitive edge of agricultural products but also enhance the capacity to manage unforeseen contingencies, thereby safeguarding the continuity and security of agricultural supply chain (Cao *et al.*, 2022; Chen and Su, 2022; Cui and Gao, 2022; Hui, 2021; Lin and Hu, 2022; Nagendra *et al.*, 2022; Xu *et al.*, 2022).

While existing studies have ventured into the realm of big data within agricultural supply chains, their focus predominantly has been on preliminary risk prediction and descriptive analysis. A notable gap exists in the thorough examination of risk factors and their interconnected dynamics as well as in the development of comprehensive risk management decision support rooted in such analysis (Ge *et al.*, 2023; Liu *et al.*, 2022). Furthermore, prevalent methodologies often overlook the intricacies of supply chain management and the multifaceted nature of decision-making, thereby impacting the efficacy and precision of decision support systems in real-world settings (Krska *et al.*, 2022; Modupalli *et al.*, 2021).

Existing agricultural risk management and safety models usually cover multiple aspects, such as production risk, market risk, financial risk, technological risk, and natural disasters. They aim to mitigate the uncertainties and potential losses in agricultural production through diversified crop planting, insurance, futures contracts, disaster relief plans, and government subsidies (Le and Chu., 2023; Yu and Liang, 2022). However, the limitations of these models often lie in their difficulty in precisely predicting and quantifying risks brought about by environmental and climate changes. They have limited responsiveness to global market fluctuations and may not cover all small-scale agricultural producers, especially in resource-limited developing countries. Moreover, these models typically require extensive data support, and there may be constraints in data collection and processing, leading to inaccuracies in risk assessment and management strategies.

This research is bifurcated into two primary segments. Initially, it undertakes a detailed investigation of risk factors within agricultural supply chains and their interrelations, utilizing a decision tree algorithm enhanced with transition structures. This technique equips decision-makers with the tools for intuitive risk identification and mapping of interrelationships, thereby facilitating the crafting of more robust risk mitigation strategies and emergency response plans. Subsequently, the study embraces a data-driven multi-criteria decision-making approach. This approach is designed to assist managers in navigating the complexities of decision environments within agricultural supply chains, weighing the intricacies of costs, benefits, and feasibility to develop optimal management policies. Through a comprehensive exploration of these research components, this study not only broadens the scope of big data application in agricultural supply chain risk decision-making and safety management but also endows decision-makers with a more scientific and systematic tool for decision support, imbued with significant research merit and practical relevance.

The novelty of this research is in the direct application of advanced data analysis algorithms to risk assessment and management in the agricultural supply chain, particularly the introduction of decision tree algorithms with transition structures as the core tool, which was not common in the previous research. This approach breaks through traditional analysis models and can delve deeply into complex interactions between risk factors, providing new perspectives for risk management in critical states. In addition, by integrating multi-criteria decision-making methods, this study further enhances the comprehensiveness of strategy formulation and the ability to cope with complexity. Such a comprehensive analysis and decision-making framework is novel in the current literature.

The research tools in this paper, in terms of practical application, mainly provide scientific and precise support for risk management in agricultural supply chain. For instance, in the food industry, the decision tree algorithms with transition structures can predict and manage potential risks related to food production and supply, such as yield fluctuations, food contamination, or supply interruptions. In stages such as the harvesting, storing, and packaging of crops, these advanced algorithms can help managers assess risks due to weather, pests, or changes in market demands, and accordingly formulate optimized operational plans and response strategies to reduce losses.

Their advantage lies in the ability to process large amounts of data and extract key information, aiding in identifying and quantifying risk points across the entire supply chain. Furthermore, the integration of multi-criteria decision-making methods with cost, benefit, and feasibility analysis ensures that the formulated risk management strategies are both economical and practical. However, the application of these research tools also has certain limitations. First, they have a strong dependency on high-quality and comprehensive data, and challenges in data collection and processing may limit their practical scope. Second, the complexity of the algorithms may require specific expertise, which could affect the usability and popularity among users. Finally, the adaptability of these tools in different regions and cultural contexts may also pose challenges and require localization adjustments and testing prior to implementation.

Analysis of Risk Factors and Their Interrelationships in Agricultural Supply Chains

The agricultural supply chain exhibits complex interrelations among risk factors. For instance, climate change can intensify the frequency and severity of pest outbreaks, directly impacting crop yield and quality. The uncertainty in yield and quality can affect market prices, thereby influencing farmers' planting decisions and stability of the supply chain. Moreover, while technological advancements can mitigate some impacts of natural risks, they may also introduce new technological risks. For example, reliance on novel storage technologies could disrupt the entire supply chain in the event of malfunctions. Risks affect not just a single segment but can also propagate and amplify throughout the supply chain. For instance, a shortage of primary agricultural products may cause stagnation in the processing stage, subsequently affecting supply in the retail segment and availability to end consumers. This chain reaction is particularly evident in the food chain, as the production, processing, and distribution of food typically involve multiple interdependent stages.

Therefore, this paper advocates for a systematic perspective in the risk management of agricultural supply chains. By thoroughly analyzing the interrelationships among risk factors and adopting comprehensive management measures, the risks in each segment can be mitigated, and the resilience of the entire chain can be optimized. This involves the continuity of risk assessment, transparency across all aspects of the supply chain, timely sharing of information, and coordinated collaboration among multiple stakeholders.

The risk management of agricultural supply chains was addressed in this work by implementing a decision tree algorithm that incorporates transition structures. This approach incorporates transition structures into the conventional decision tree analysis, enabling test samples to probabilistically transition between decision paths. This approach surpasses the constraints of singular path assessments, allowing the decision-making process to integrate information from several paths, thereby reflecting a more comprehensive probability distribution. This approach effectively captures the interplay of nonlinear, complex, and constantly evolving risk elements in the supply chain, offering a reliable way for assessing and predicting risks under unpredictable circumstances. Figure 1 displays a schematic representation of a decision tree. Figure 1 contains a decision tree structure, featuring a hierarchical arrangement of nodes and leaves. The decision tree starts from the topmost root node (labeled as 1), with two values on it: 40 and 30. From the root node, two child nodes branch out, labeled 2 and 7, each respectively annotated with a pair of values: 22:18 and 13:17. This branching process continues, with some nodes further branching down into new nodes, until reaching the bottom leaf nodes (labeled as 4, 6, 9, 11, 12, and 13), which also bear a set of values. Arrows indicate the direction of decision-making, and the number at each node represents the decision or statistical value at that point. The entire structure reflects a decision-making process from root to leaf, potentially representing different steps and outcomes in data classification.

The fundamental concept behind utilizing the decision tree algorithm with transition structures for examining risk factors in agricultural supply chains and their interconnections entails merging the decision tree model with the theory of Hidden Markov Model (HMM) (Gueham and Merazka, 2024; Nagahama *et al.*, 2014). This results in a composite model that can effectively capture temporal dependencies and implicit states in sequential data. Each node in this approach serves as both decision condition or output and a hidden state in HMM. The hidden states are linked by a transition matrix, which represents the probabilities of transitioning between different risk variables. The transition structure enables decision routes to stochastically transition across nodes, capturing the unpredictable and dynamic nature of risk variables in the agricultural supply chain environment. The emission matrix quantifies the likelihood of considering a particular risk impact for each concealed state. It represents the probability distribution of different risk events occurring based on the specific state of a supply chain. By applying the technique, the probability distribution of test samples across all nodes can be derived. Then the expected values of these distributions are utilized to make predictions about the links between risk variables and impacts. The expected values represent the overall probability of risk happening across the whole supply chain. Therefore, decision-makers can acquire understanding not only of individual risk forecasts but also of the combinations

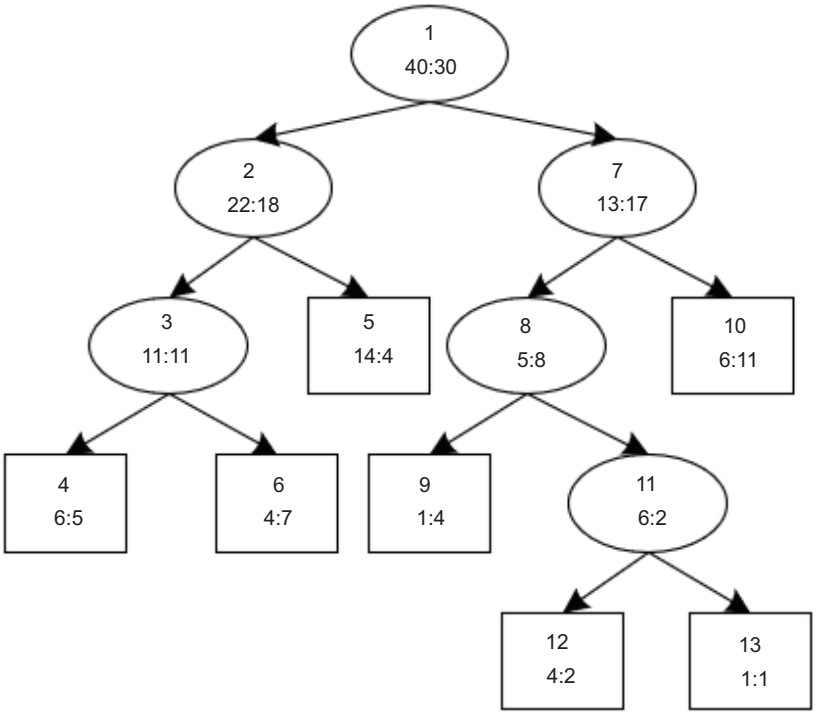

Figure 1. Schematic diagram of a decision tree.

and interconnections of risks across several conceivable routes, thereby developing more comprehensive and efficient risk management strategies.

Given that the algorithm's initial state is denoted as $\tau = (1,0,...,0)_{1*V}$, the emission matrix as $Y = (O_1,O_2,... ,O_V)^S$, and the prediction at decision tree depth u as PR_u, the probability distribution of samples in nodes after u-step transitions from the root node 1 is represented by τO^u. A formula is presented for calculation:

$$PR_u = \tau O^u Y. \quad (1)$$

To facilitate clarity in explanation, the following definitions are presented.

The transition matrix is a square matrix with dimensions equal to the square of the number of hidden states. The total of the probabilities in each row is equal to one, indicating the probability distribution of transitioning from one state to different states. The matrix in question is a fundamental notion in HMM, and precisely characterizes the probability of transitioning between any two hidden states within a decision tree. Within the framework of agricultural supply networks, this refers to the likelihood of shifting from one particular risk factor to another. The transition matrix of a sample under the given decision tree S is expressed as follows:

$$O = \left[o_{uk}\right]_{V\times V}, \sum_{k=1}^{V} o_{uk} = 1 \text{ OR } 0, \quad (2)$$

where O_{uk} represents the transition probability from node u to node k, and V represents the total number of decision tree nodes.

Within the decision tree, target nodes correspond to leaf nodes. These leaf nodes signify the outputs of a decision tree, specifically the ultimate risk assessment outcomes. Within the decision tree framework, the target nodes can be likened to the final states in HMM, as they offer precise insights into the risk states of the supply chain. Non-target nodes, on the other hand, refer to the internal nodes in the decision tree that are not leaf nodes. In the context of HMM, these intermediate hidden states serve as representations of intermediate judgments or transitions that occur before reaching the final risk assessment. These nodes direct samples from the current node to the next node, depending on branching circumstances, or simplify transitions between states. For a given node u, if a sample a is assigned to a certain child node k according to a rule, then k is designated as the target node for an inside u, while the remaining child nodes k' are classified as non-target nodes.

The allocation of sample data to different child nodes in a given decision tree is determined by the division rules of non-leaf nodes, which are based on certain traits or features. Every internal node is linked to a decision rule, which might be either a threshold or a set. Once a sample reaches a non-leaf node, it is assigned to the appropriate child node according to a predetermined decision rule. If the attribute of the sample meets the criteria of

the rule, it will go to the appropriate child node. If not, it may transition to a node on a different path based on the transition structure. This shift occurs by utilizing the probabilities in the transition matrix, enabling the algorithm to investigate unconventional paths of risk states and detect more intricate risk patterns. When dividing the sample a into the target node k at the non-leaf node u in the decision tree S, the equation $IN_{u'} = (o_{uk}, o_{uk'})$ holds if the transition probability is represented by o_{uk}. Node u is regarded as the sibling node of node k', meeting the condition $o_{uk} + \Sigma_{k'} o_{uk'} = 1$.

Constructing the transition probabilities between nodes is a crucial stage in the decision tree algorithm for agricultural supply chain risk analysis, which involves transition structures. This entails computing the transition probabilities to both target and non-target nodes and converting these probabilities from a fixed form to a probabilistic distance form. The subsequent information outlines the construction process.

When constructing the decision tree, the initial step is to define the probable transition relationships between each node. This is derived from the comprehension of risk factors and examination of the past data, identifying the nodes that are probable to undergo a transition. The base probability of transitioning to other nodes is determined for each individual node. One can acquire this information by either analyzing the frequency of transitions between risk events in historical data or by utilizing estimations from experts with specialized knowledge. Fuzziness and randomness are incorporated into decision-making scenarios to account for inherent uncertainties. The transition probabilities at the base level are modified using fuzzy logic or probabilistic distributions to represent accurately the uncertainties and intricacies of decision-making in real-world scenarios. It's assumed that sample a is divided into the target node k at a non-leaf node u, the function involving $a[x_u]$ and S are represented as o_{uk}, and the number of child nodes of u is represented as v. To compute the transition probabilities from u to k and k′, the following method is used:

$$o_{uk} = h(a[x_u], S), 0 \le o_{uk} \le 1, \quad (3)$$

$$o_{uk'} = \frac{1 - o_{uk}}{v - 1}. \quad (4)$$

The initial transition probabilities are commonly displayed in a constant format, which represents the ideal conditions based on specific assumptions. Nevertheless, the actual situation in agricultural supply chains tends as more intricate, requiring a conversion of these consistent probabilities into formats that encapsulate uncertainty. By applying the concept of probabilistic distance, it is possible to convert constant transition probabilities into probabilistic distributions. This entails the introduction of a probability density function for each transition probability. It is commonly believed that these probabilities follow a normal distribution or another suitable distribution. The parameters of these distributions are given by constant transition probabilities and the related risk data. Because of this transformation, each transition probability is no longer represented by a single value but by a probabilistic distribution. This allows for a more accurate representation of uncertainties and variabilities that exist in the real world. The equation that represents the constant form of O_{uk} is as follows:

$$o_{uk} = \beta, 0 \le \beta \le 1. \quad (5)$$

It's assumed that the cumulative distribution function of the x_uth feature utilized at node u is denoted by Θ_{xu}, the base transition probability by a hyperparameter β, and the distance between the feature value $a[x_u]$ of the sample and the node threshold s_u by $|\Theta_{xu}(a[x_u]-\Theta_{xu}(s_u)|$. The expression for O_{uk} in its probabilistic distance version is given by equation (6).

$$o_{uk} = \beta + (1-\beta)\,|\,\Theta_{x_u}(a[x_u]) - \Theta_{x_u}(s_u)\,|, 0 \le \beta \le 1. \quad (6)$$

Utilizing the decision tree algorithm with transition structures to analyze risk factors and their effects in agricultural supply chains offers notable benefits by describing the transition probabilities between nodes in probabilistic distance forms. First and foremost, this form presents a more precise representation of uncertainties and variations of risk factors in real-life scenarios, as it offers a probability distribution instead of a single transition probability number. This method encompasses the full spectrum of potential risk events and the attributes of their distribution. Furthermore, this approach enables decision-makers to comprehend visually a more intricate risk scenario, facilitating the assessment and comparison of anticipated results across various risk management tactics.

Incorporating transition probabilities in a probabilistic distribution form allows for the integration of information from various data sources, such as historical data, real-time data, and expert opinions. This thorough examination enhances the precision of risk forecasts and the dependability of risk mitigation choices. In essence, this establishes a strong theoretical basis for constructing a versatile and responsive system for managing risks in the agricultural supply chain. This system can effectively handle existing identified hazards as well as unforeseen risks that may arise in the future. For example, let $a \sim V(0,1)$. The values a_1, a_2, and a_3 are respectively equal to 0, 1, and 2. By calculating the distances $DI(a_1,a_2,)$ and $DI(a_2,a_3)$, it is intuitively believed that $DI(a_1,a_2,)$ is bigger than $DI(a_2,a_3)$. This is because in a finite sample set T that

follows a specified distribution, the number of elements in the range $\{a|a_1<a<a_2, a\in T\}$ is greater than the number of elements in the range $\{a|a_2<a<a_3, a\in T\}$. Nevertheless, the distance between $DI(a_1,a_2,)$ and $DI(a_2,a_3)$ is equal to 1, which deviates from the expected norm. By employing probabilistic distance, the value of $DI(a_1,a_2,)$ is approximately 0.24 and the value of $DI(a_2,a_3)$ is around 0.14, thereby resolving this matter. In order to utilize probabilistic distance, it is necessary to possess knowledge of the distribution function of variable x. When node u is a leaf node, the following is observed:

$$o_{uk}=\begin{cases}1, u=k\\ 0, else\end{cases}. \quad (7)$$

Trade-Offs in Data-Driven Decision-Making for Managing Rural Supply Chain Risks

This study employs a data-driven approach to evaluate and prioritize various options for managing risks in rural supply chains. Managers can use this technique to make decisions by considering various interconnected factors, such as cost, benefits, risk levels, resource availability, and socioeconomic implications. The data-driven method utilizes advanced technologies, such as machine learning and statistical analysis, to extract insights from historical data (Asfaw *et al.*, 2023; Jiang *et al.*, 2023). These insights are then integrated into a multi-criteria decision-making framework after identifying patterns and trends. This methodology facilitates quantitative analysis, enhancing the transparency of the decision-making process and effectively illustrating the performance of alternative decision options across several criteria and their influence on the ultimate results of risk management. Moreover, the data-driven multi-criteria decision-making method facilitates the incorporation of real-time data, hence improving the promptness and flexibility of decisions. The adaptability and agility of this approach enables managers to modify tactics in response to shifting market and environmental circumstances, guaranteeing the stability and longevity of the agricultural supply chain, thus optimizing economic and social worth. Figure 2 depicts the optimization process of decision-making for managing risks in rural supply chain. This technique is based on the dynamic selection of various classifiers.

The purpose of developing an optimization model for decision-making in rural supply chain risk management

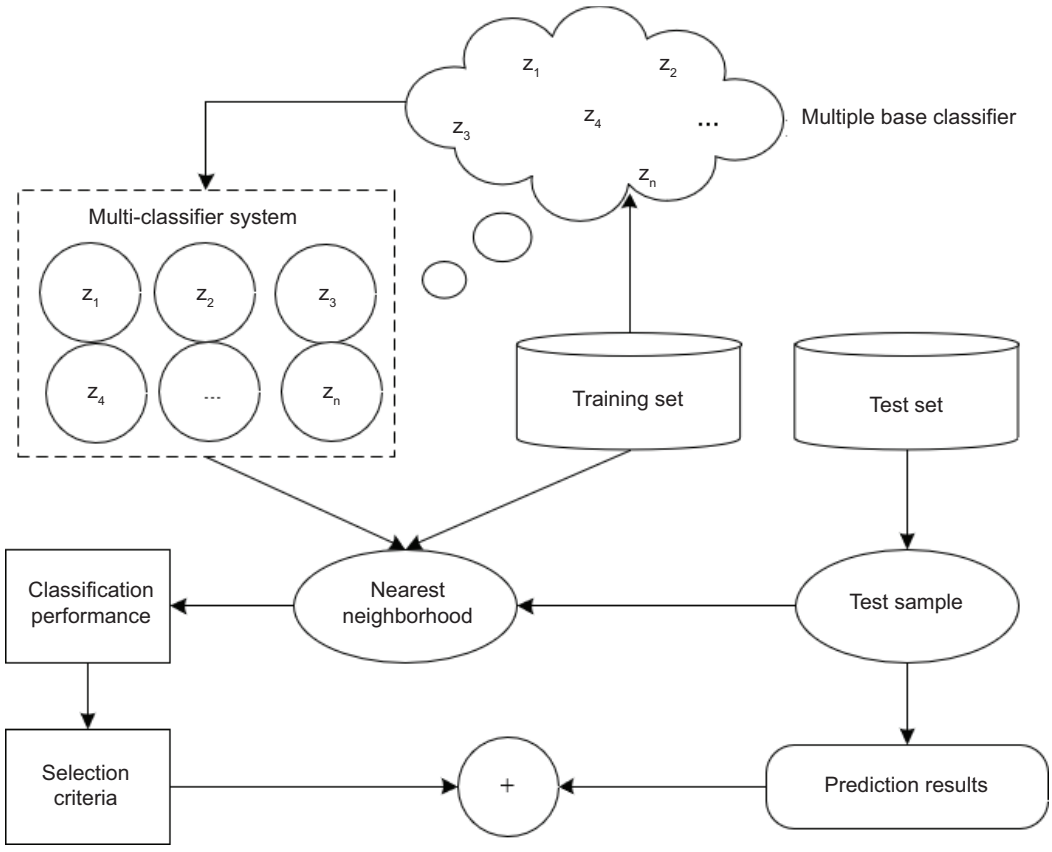

Figure 2. Optimization process of rural supply chain risk management decision-making based on dynamic selection of multiple classifiers.

is to offer managers a systematic approach to evaluate and select various solutions for managing risks. The underlying premise of this strategy revolves around two essential steps: First, the process involves using base classifiers to predict the evaluation values of a gold standard. This is done by training validated classifiers on historical data. The goal is to predict the performance of each risk management strategy based on criteria, such as cost efficiency, degree of risk reduction, and execution feasibility. The prediction findings act as benchmark assessment values, offering foundational data for the future optimization model. Furthermore, the process of determining criterion weights seeks to discover the optimal set of weights that accurately represents the correlation between risk management techniques and gold standard evaluation outcomes. The optimization model evaluates the relative relevance of each criterion by analyzing the influence of each criterion on the gold standard assessment results. This is commonly accomplished using optimization algorithms, such as genetic algorithms or gradient descent methods, with the objective of minimizing the discrepancy between predictive evaluation values and the gold standard. In the end, the model utilizes the acquired optimal criterion weights to combine the assessment values of each strategy across several criteria, resulting in a composite evaluation score. This score can be used as the benchmark evaluation outcome for recommending strategies. These outcomes not only demonstrate the comparative advantages and disadvantages of various risk management strategies when evaluated using many criteria but also offer concrete and measurable evidence for decision-makers. This enhances the transparency, rationality, and feasibility of risk management decisions. Figure 3 illustrates the procedure for generating a base classifier.

When making decisions about managing risks in rural supply chains, managers must make optimal choices considering several factors that impact risk management, including cost, time, resource utilization, environmental consequences, and social accountability. The assessment of each criterion is represented by an interval number, which indicates the range of uncertainty in evaluation results. This interval number represents the upper and lower limits of the best and worst potential outcomes. In the context of managing risks in rural supply chain, it is presumed that there are V different techniques for managing these risks, denoted as $\bar{a}_v (v = 1,..., V)$. The gold standard evaluation values for these V techniques are derived by experts in the relevant field who have chosen M decision criteria, denoted as $\{r_1, ..., r_M\}$. At first, experts must assign specific evaluation values to these V risk management strategies based on M criteria. The evaluation vector of the assessed rural supply chain risk management strategy a_v under M criteria is denoted as A^-_v. The evaluation value of interval number a^-_v under choice criterion $r_u (u = 1,..., M)$ is represented as $A^-_{v,u} = [A^{-\,-}_{v,u}, A^{-\,+}_{v,u}]$.

The responsibility of decision-makers is to discern the most advantageous approach from all the available possibilities, taking into account all relevant criteria. Hence, it is necessary to devise a technique to synthesize or compare various interval numbers in order to ascertain the combined interval number that signifies the most ideal performance of a strategy. Typically, this process entails the comparison and ranking of interval numbers as well

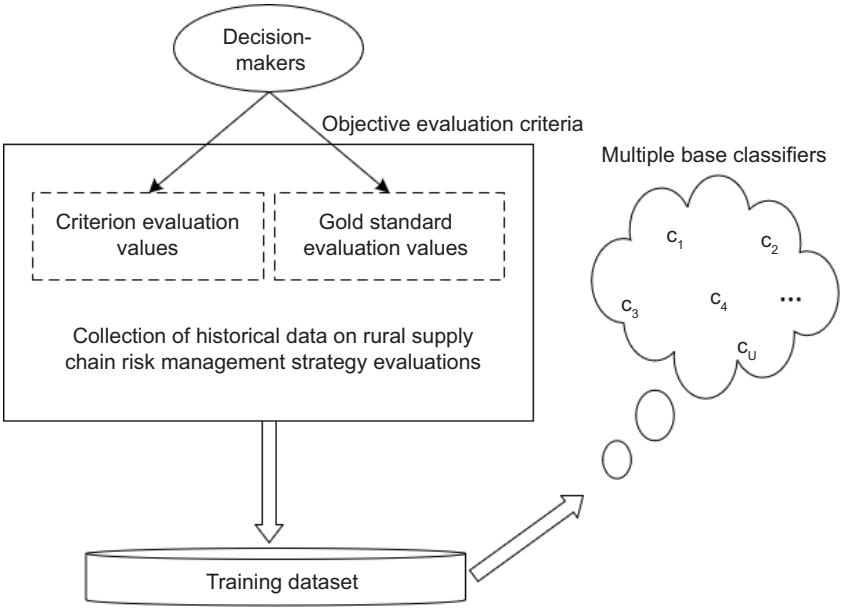

Figure 3. Process of generating a base classifier.

as the determination of the most favorable criterion weights. The process of determining criterion weights can be acquired through a data-driven method, which quantifies the relative significance of various criteria in the ultimate decision-making process. First, a collection of historical risk management strategies with established evaluation values, denoted by $\{a_m\}_{m=1}^{L}$, is obtained. The evaluation values under M criteria are denoted as $\{A_m\}_{m=1}^{L}$, while the matching gold standard evaluation values are denoted as $\{O_m\}_{m=1}^{L}$. These values together form a training set represented by $\{(A_m, O_m)\}_{m=1}^{L}$. The evaluation value of the historical risk management strategy a_m is represented by the gold standard O_m, which is defined as the interval $[O_m^-, O_m^+]$. The multi-criteria decision issue under consideration involves Y types of evaluation values, denoted as $\{O^{\to y}\}_{y=1}^{Y} = \{[O^{\to y-}, O^{\to y+}]\}_{y=1}^{Y}$, where $O_m \in \{O^{\to y}\}_{y=1}^{Y}$.

A dynamic classifier selection approach is given as a solution to the described problem. This method does not depend on a fixed K value but instead selects classifiers dynamically by considering data similarity and the predictive quality of base classifiers. The set of base classifiers derived from the decision evaluation set $\{(A_m,O_m)\}_{m=1}^{L}$ of historical risk management strategies $\{a_m\}_{m=1}^{L}$ is denoted as $z = \{z_1, ..., z_I\}$.

The study initially establishes the concept of individual evaluation vector similarity, which pertains to the extent of similarity between the evaluation vector of a new strategy and the evaluation vectors of past strategies. The degree of similarity can be measured by computing the distance between vectors, such as with Euclidean distance, cosine similarity, or other comparable metrics. The individual evaluation vector similarity is employed to locate instances in the historical data that closely resemble the new approach. These instances are then utilized to forecast the gold standard evaluation value of the new method. The training dataset, denoted as $\{(A_m,O_m)\}^{L}_{m=1}$, and the rural supply chain risk management strategy to be assessed, denoted as a_v^-, could be used to represent the individual evaluation vector of a_v^- under M criteria, which is denoted as $A_v^- = \{A_{v,u}^-\}_{u=1}^{M}$. If $z_i(A_m)$ and $z_i(A_v^-)$ are the gold standard evaluation predictions received by a_m and $A\text{-}_v$ respectively from z_i(i = 1,..., I), and if $z_i(A_m)$ is equal to $z_i(A_v^-)$, then z_i determines that a_m and A_v^- are similar.

The predictive performance of base classifiers pertains to the precision and dependability of classifiers in assessing risk management techniques on past datasets. Accurate prediction entails classifiers being able to effectively assess the gold standard assessment values of risk management systems. Consider past risk management practices $\{a^i_{(v,m)}\}^{Liv}_{m=1}$ that differ from the rural supply chain risk management strategy to be assessed A_v^-, where L_v^i denotes the number of items in this set. The evaluation vector and related gold standard evaluation values of risk management strategy $a^i_{(v,m)}$ are represented by $a^i_{(v,m)}$ and $O^i_{(v,m)}$, respectively. Next, the equations $z_{i(}A^i_{(v,m))} = z_{i(}A^-_{v)}$, where $\{_{(}A^i_{(v,m))},_{(}O^i_{(v,m))}\}^{Liv}_{m=1} \subseteq \{(A_m,O_m)\}^{L}_{m=1}$ and $L^i_v <= L$, are established. The predictive quality of z_i is believed to increase as the degree of similarity between $\{_{(}A^i_{(v,m))}\}^{Liv}_{m=1}$ and a^-_v increases.

It's assumed that historical risk management strategies are denoted as a_m (where m ranges from 1 to L), and the rural supply chain risk management strategy for evaluation is denoted as a^-_v. The related gold standard prediction results for the historical strategies are represented as $z_i(A_m)$ (where m ranges from 1 to L), and the gold standard prediction result for the strategy for evaluation is represented as $z_i(A^-_v)$. It's assumed that the historical risk management strategies, which are comparable to a^-_v determined by $z_i(A^-_v)$, are represented by $a^i_{(v,m)}$. The individual evaluation values of risk management strategies, $A^i_{(v,m)}$, and A^-_v under the criterion r_u, are represented by $A^i_{(v,m),u}$ and $A^-_{v,u}$, respectively. The calculation $f(A^i_{(v,m),u},A^-_{v,u})$ represents the distance between the interval numbers $A^i_{(v,m),u}$ and $A^-_{v,u}$. The predictive quality of z_i can be determined using the following equation:

$$W_v^i = 1 - \frac{\sum_{m=1}^{L_v^i}\left(\frac{\sum_{u=1}^{M} f\left(A^i_{(v,m),u}, \overline{A}_{v,u}\right)}{M}\right)}{L_v^i}. \quad (8)$$

The predictive accuracy of basic classifiers pertains to the degree of agreement between the classifier's predictions on a given evaluation vector and the true evaluation values determined as the gold standard. Predictive accuracy is a crucial measure in the dynamic selection process, employed to ascertain which classifiers are more inclined to yield precise outcomes when forecasting the gold standard evaluation values of novel techniques. Classifiers that demonstrate a high level of accuracy in making predictions are given greater importance by assigning them larger weights. Let $\{(A^y_m,O^{\to y})\}^{Ly}_{m=1}$ represent the decision dataset of historical risk management strategies with gold standard evaluation values $O^{\to y}$ in the training set. In this dataset, L represents the total number of strategies, and $L_1 + L_2 + ... + L_Y = L$ is established. The evaluation vector sets for each type of risk management strategy under M criteria are denoted as $\{A^y_m\}^{Ly}_{m=1}$(y = 1,..., Y). These sets are then inputted into z_i to obtain the prediction result set for different types of historical risk management strategies, represented as $\{z_i(A^y_m)^{Ly}_{m=1}$. The $O^{\to y}$ distance between interval number $z_i(A^y_m)$ is denoted as $f(z_i(A^{y\text{-}}_m),O^{\to y})$, while the prediction accuracy of z_i for the *y*th type of historical risk management strategy is denoted as

X^i_y(y = 1, ..., Y). The equation below presents the calculating formula for z_i's predicted accuracy across several risk management strategies:

$$X^i_y = 1 - \frac{\sum_{m=1}^{L_y} f\left(z_i\left(A^y_m\right), \vec{O}^y\right)}{L_y}\left(y = 1, ..., Y\right). \quad (9)$$

To anticipate the gold standard evaluation value of the rural supply chain risk management strategy to be evaluated a^-_v, the $z^{v/i}$ selection can be determined by combining the above-mentioned two equations:

$$\bar{i} = \text{ARGMAX}\left\{i \mid W^i_v + X^i_y, i = 1, ..., I\right\}. \quad (10)$$

The acquisition of criterion weights is a fundamental component in the data-driven multi-criteria decision-making method. The objective is to ascertain the comparative significance of various risk management criteria and thereafter assess novel techniques in accordance with this evaluation. The initial step involves defining the similarity between individual assessment values and the expected gold standard evaluation outcomes, which is used to analyze their similarity. This is accomplished by computing the disparities between the evaluations of criteria and the established level of excellence.

It's assumed that the historical risk management strategies identical to a^-_v are denoted as $a^{-i}_{(v,m)}$. These strategies have individual evaluation values on criterion r_u, expressed as $a^{-i}_{(v,m),u}$. The gold standard evaluation prediction results generated from $z^v_{/i}$ are represented as $z^v_{/i}(A^{-i}_{(v,m)})$. The similarity between $a^{-i}_{(v,m),u}$ and $z^v_{/i}(A^{-i}_{(v,m)}$ for $a^{-i}_{(v,m)}$ can be defined as

$$TUL^{\bar{i}}_{(v,m),u} = 1 - f\left(A^{\bar{i}}_{(v,m),u}, z^v_{\bar{i}}\left(A^{\bar{i}}_{(v,m)}\right)\right). \quad (11)$$

The function $f(a^{-i}_{(v,m),u}, z^v_{/i}(A^{-i}_{(v,m)})$ represents the distance between $A^{-i}_{(v,m)}$ and $z^v_{/i}(A^{-i}_{(v,m)})$. To calculate the criterion weights, the decision variable $q^{-i*}_{(v,m),u}$ and the values of $TUL^{-i}_{(v,m),u}(u = 1, ..., M)$ are used.

$$q^{\bar{i}}_{(v,m),u} = \frac{TUL^{\bar{i}}_{(v,m),u}}{\sum_{u=1}^{M} TUL^{\bar{i}}_{(v,m),u}}, \left(u = 1, ..., M\right). \quad (12)$$

Let $q^{-i}_{v,u}(u = 1, ... M)$ be a set of the most representative criteria weights. Next, a mathematical model may be created to calculate the values of $q^{-i}_{v,u}$ (where u ranges from 1 to M) based on the given values of $q^{-i*}_{(v,m),u}$ (where u ranges from 1 to M and m ranges from 1 to L^{-i}_v) in the following manner:

$$LUV \sum_{m=1}^{L^{\bar{i}}_v} \sum_{u=1}^{M} \left(q^{\bar{i}}_{(v,m),u} - q^{\bar{i}^*}_{(v,m),u}\right)^2, \quad (13)$$

$$\text{s.t.} \sum_{u=1}^{M} q^{\bar{i}^*}_{(v,m),u} = 1, \quad (14)$$

$$0 \le q^{\bar{i}^*}_{(v,m),u} \le 1. \quad (15)$$

The optimization model has a distinct optimal solution.

$$\left(\bar{q}^{\bar{i}}_{v,1}, ..., \bar{q}^{\bar{i}}_{v,M}\right) = \left(\frac{\sum_{m=1}^{L^{\bar{i}}_v} q^{\bar{i}}_{(v,m),1}}{L^{\bar{i}}_v}, ..., \frac{\sum_{m=1}^{L^{\bar{i}}_v} q^{\bar{i}}_{(v,m),M}}{L^{\bar{i}}_v}\right). \quad (16)$$

After establishing the criterion weights, the evaluation values $\{A^-_{v,1}, ..., A^-_{v,M}\}$ for the rural supply chain risk management strategy a^-_v can be calculated across M criteria. This is done by using $\{q^{-i}_{v,u}\}^M_{u=1}$ obtained from the aforementioned process. The result is an interpretable gold standard evaluation value, denoted as $\Sigma^M_{u=1} q^{-i}_{v,u} \times A^-_{v,u}$. This process entails assigning weights to the evaluation findings of each criterion based on their importance in order to calculate a composite evaluation value that considers all relevant criteria.

Ultimately, the composite evaluation value must be converted into a format that enables effective communication with decision-makers and can be readily utilized in the decision-making process. For example, the composite evaluation value can be transformed into a classification, indicating high, medium, or low risk. It can also be converted into a numerical score or probability value, allowing decision-makers to comprehend easily the amount of risk associated with each option. The transformation performed to $\Sigma^L_{i=1} w^{-u}_{n,i} * X^-_{n,i} \Sigma^M_{u=1} q^{-i}_{v,u} * A^-_{v,u}$, resulting in $\tilde{O}_v$, assumes that $\Sigma^M_{u=1} q^{-i}_{v,u} * A^-_{v,u}$ contains Y distinct types. The function $f(O^{\to y}, \Sigma^M_{u=1} q^{-i}_{v,u} * A^-_{v,u})$ represents the distance between the interval number $O^{\to y}$ and $\Sigma^M_{u=1} q^{-i}_{v,u} * A^-_{v,u}$,

$$\tilde{o}_v = \vec{o}^y,$$

$$IF = \text{ARGMIN}\left\{y \mid f\left(\vec{o}^y, \sum_{u=1}^{M} \bar{q}^{\bar{i}}_{v,u} \cdot \bar{A}_{v,u}\right), y = 1, ..., Y\right\}. \quad (17)$$

Figure 4 depicts the decision-making trade-off process for managing risk in the rural supply chain using the suggested method. This text provides a detailed explanation of the procedure that relies on the data-driven multi-criteria decision-making method for managing trade-offs in decision-making for rural supply chain risk management.

Step 1. *Expert examination*: Decision-makers utilize their profound comprehension and wide expertise of the rural supply chain environment to evaluate newly suggested risk management techniques. The assessment criteria

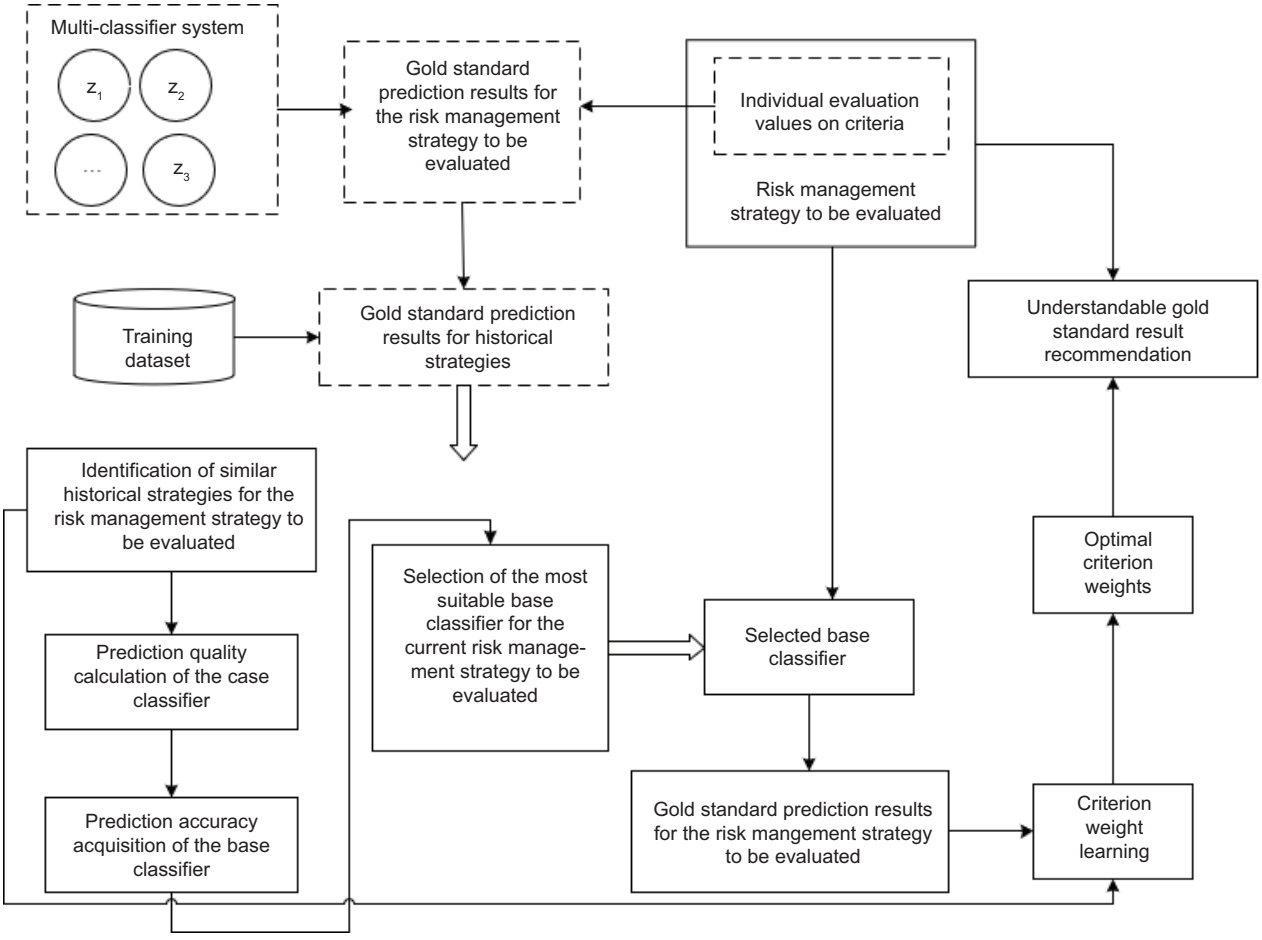

Figure 4. Decision-making trade-off process for managing rural supply chain risks based on the proposed method.

encompass cost-effectiveness of the approach, anticipated risk reduction effects, feasibility of execution, and social and environmental repercussions. The assessment result of each criterion is quantified into an individual evaluation value, which can be either quantitative figures or qualitative descriptions. These evaluation values serve as initial data for further analysis.

Step 2. *Construction of the training set*: This entails gathering historical data, which include implemented risk management techniques, their evaluation values based on different criteria, and the related gold standard evaluation values. The objective is to create an extensive dataset that serves as the basis for training machine learning models to capture accurately the connections and interactions between different criteria.

Step 3. *Construction of the base classifier set*: In this stage, a sequence of fundamental classifiers is learned using the previously specified training set. Every classifier utilizes a distinct algorithm to enhance resilience against diverse data types and risk patterns. The goal is to develop a set of predictive models with varied attributes and capabilities for dynamic selection in following stages.

Step 4. *Prediction evaluation using gold standard*: After acquiring the collection of basic classifiers, the criterion evaluation vectors of the strategies being assessed are fed into these classifiers. The classifiers utilize the acquired knowledge from the training set to forecast the prospective gold standard assessment values of novel strategies. This technique entails converting the evaluation values provided by experts into a collection of features that can be analyzed by classifiers. These features are then utilized to forecast the overall efficacy of the tactics.

Step 5. *Determining the nearest neighborhood*: The objective of this stage is to discover past risk management techniques that closely resemble the strategy being evaluated based on the evaluation criteria vectors. The result of this stage is a subset consisting of historical strategy cases that closely resemble the examined strategy in terms of their characteristics and are deemed as valuable references for decision-making.

Step 6. *Evaluation of classifier prediction accuracy*: This entails evaluating the level of concordance between the predictions made by the basic classifier on past cases within the closest proximity and the benchmark evaluation values.

Step 7. *Partitioning the training set*: In order to improve the precision and applicability of forecasts, the training dataset is separated into smaller groups based on the specific type or attributes of the approach. The categorization can be determined by variables, such as the level of risk associated with the approach, the scope of its execution, and the expected influence it would have.

Step 8. *Acquiring the prediction result set*: The criterion evaluation vectors of various historical risk management procedures are inputted into each base classifier, based on the subset division indicated before. The base classifiers produce prediction result sets for each subgroup, providing a first assessment of the potential gold standard evaluation values for various risk management strategies. The result sets are heterogeneous due to the incorporation of several base classifiers and types of risk management procedures.

Step 9. *Calculating the accuracy of classifier's predictions*: During this stage, the precision of each individual base classifier's predictions is evaluated. In addition, macro averaging or micro averaging can be employed to evaluate completely all categories in multi-class problems.

Step 10. *Strategic forecasting determination of accuracy*: The prediction accuracy of each base classifier for the gold standard evaluation values of the evaluated method is examined based on the accuracy gained in the previous phase. This process entails inserting the evaluation vector of the technique being evaluated into each base classifier and comparing its output to the gold standard evaluation values of similar cases in historical data to assess consistency.

Step 11. *Selection of the most effective base classifier*: After evaluating the predicted accuracy of each base classifier, the subsequent task is to select the most appropriate base classifier for the present risk management approach. The selection criteria are not exclusively reliant on accuracy but may also take into account other elements, such as the model's stability, interpretability, and computing cost. The selection of a base classifier that is most suitable for the given application scenario is determined by comparing these characteristics.

Step 12. *Obtaining similar strategy and evaluation vectors*. By utilizing the chosen optimal base classifier, comparable historical risk management methods and their criterion evaluation vectors to the strategy being evaluated are acquired. This stage entails examining the prediction outcomes of the base classifier in order to determine the previous cases that bear the closest resemblance to the new method.

Step 13. *Calculation of similarity*: The similarity is calculated between the individual evaluation values of the historical risk management strategies, which are similar to the assessed strategy across each criterion, and the gold standard evaluation values predicted by the base classifier.

Step 14. *Calculation of criterion weights*: This step tries to calculate the weights for different criteria based on the results of the similarity calculation in the previous step.

Step 15. *Calculation of the most efficient solution*: In this stage, the optimization model is solved by utilizing the calculated criterion weights from the previous steps. The optimization model can be constructed by defining a set of goal functions and constraints, which may include minimizing risk and maximizing benefits. The ideal solution offers decision-makers a comprehensive evaluation of the success of the new strategy, taking into account all criteria and their respective weights.

Step 16. *Recommendation for gold standard evaluation*: Ultimately, the optimization model produces a gold standard evaluation suggestion for the risk management approach based on the optimal solution. The advice is transformed into comprehensible decision support information, such as risk ratings, strategy priorities, or implementation ideas. This step involves the visualization of the recommendation outcomes, allowing decision-makers to gain a clear and intuitive understanding of the potential advantages and disadvantages of each option.

Experimental Results and Analysis

There are actual cases where agricultural producers in the Midwest of the United States face risks from extreme weather events, and in certain regions of Africa, crop yields are unstable due to drought and pestilence. Scholars have used decision tree algorithms to identify key climatic variables affecting harvests (such as precipitation, temperature fluctuations, etc.) and to predict harvest risks based on these. However, in practical applications, the effectiveness of these tools and methods largely depends on the quality and completeness of data, the cooperation of relevant stakeholders, and the decision-makers' understanding and acceptance of the decision models. Moreover, while these tools can help identify and mitigate risks, they cannot eliminate them entirely. Therefore, in practice, it's essential to combine these tools with experience and real-time information for flexible adjustment.

Table 1 provides necessary data to make comparative analysis of the accuracy and Area under the Curve (AUC) values of several methods on nine real datasets. The further examination and deductions are made as follows:

First, it is evident that the algorithm suggested in this study demonstrates superior accuracy across all datasets (B1–B9), compared to the other four algorithms, namely C4.5, Classification and Regression Tree (CART), Random Forest (RF), and Gradient Boosting Decision Tree (GBDT). The suggested method consistently achieves a higher AUC value, compared to the existing algorithms, suggesting its ability to maintain a reduced false positive rate while achieving a higher true positive rate. The subsequent inferences can be derived: The suggested technique consistently outperforms the other four algorithms, not only in terms of accuracy but also in AUC performance, across nine real datasets. Improvements in accuracy and AUC values indicate that the suggested method has enhanced its predicted accuracy and overall effectiveness as a classifier. Increase in AUC values indicates that the suggested algorithm possesses enhanced discriminatory power among various categories, which is particularly crucial in domains that prioritize risk assessment.

Figure 5 provides an opportunity to analyze the influence of decision tree depth on accuracy (ACC) and AUC values of various methods. For all algorithms, when the depth of the decision tree increases, the accuracy of C4.5, CART, and GBDT improves initially and eventually reaches a point of stability or mild fall. This suggests that as the complexity of the model increases, its ability to fit the training data improves, but there is also a risk of overfitting. The algorithm presented in this study demonstrates a consistent improvement in accuracy as the depth increases, reaching its highest point at a depth of 5, and subsequently exhibiting minor fluctuations. The proposed algorithm has superior accuracy, compared to previous algorithms across all depths, particularly exhibiting notable advantages at depths of 5 and beyond.

Concerning the AUC values, as the depth increases, all algorithms demonstrate a positive trend in AUC. However, the rate of increase eventually diminishes, and certain algorithms discover a decline in AUC after reaching a specific depth, possibly because of overfitting. The AUC value of RF algorithm demonstrates a deceleration in growth after reaching a depth of 6, displaying a pattern similar to ACC. Conversely, the AUC value of the GBDT remains very consistent even when the depth increases to 5, indicating strong stability. The suggested approach exhibits an increasing AUC value with depth, surpassing or matching the AUC values of the previous algorithms at all depths. The suggested approach consistently maintains a high AUC value, especially at depths of 5 and above, indicating exceptional classification ability.

The suggested approach outperforms or matches the accuracy and AUC values of other reference algorithms at different depths, demonstrating its efficacy and superiority. At a depth of 5, the suggested algorithm achieves its highest level of accuracy and AUC values, demonstrating a favorable equilibrium between model complexity and generalization ability at this depth. While all algorithms tend to overfit as their depth increases, the proposed approach has greater resilience, retaining excellent performance even in deeper trees. To summarize, the proposed method exhibits robust performance and versatility across several domains. It notably exhibits excellent stability and resilience against overfitting as depth of the decision tree increases, essential for real-world applications.

Table 1. Depth on accuracy and AUC of five different algorithms across nine actual datasets.

	Accuracy					AUC				
Dataset	**C4.5**	**CART**	**RF**	**GBDT**	**Proposed algorithm**	**C4.5**	**CART**	**RF**	**GBDT**	**Proposed algorithm**
B_1	0.7325	0.7356	0.7325	0.7458	0.7526	0.7548	0.7654	0.7741	0.7789	0.7895
B_2	0.9368	0.9315	0.9369	0.9369	0.9458	0.9426	0.9412	0.9784	0.9784	0.9884
B_3	0.6458	0.6587	0.6514	0.6541	0.6621	0.6589	0.6639	0.6859	0.6852	0.7142
B_4	0.7153	0.7123	0.7147	0.7147	0.7348	0.7214	0.7423	0.7321	0.7321	0.7662
B_5	0.7189	0.7852	0.7856	0.8125	0.8223	0.8236	0.8321	0.8369	0.8895	0.8992
B_6	0.7148	0.7123	0.7123	0.7236	0.7347	0.7256	0.7214	0.7145	0.7321	0.7621
B_7	0.7267	0.7236	0.7256	0.7236	0.7336	0.7485	0.7458	0.7895	0.7514	0.7992
B_8	0.9178	0.9189	0.9123	0.9178	0.9247	0.9236	0.9256	0.9632	0.9562	0.9784
B_9	0.7489	0.7658	0.7789	0.7562	0.7989	0.6321	0.6358	0.7214	0.6326	0.7536

AUC = Area Under the Curve; CART = Classification and Regression Tree; GBDT = Gradient Boosting Decision Tree; RF = Random Forest.

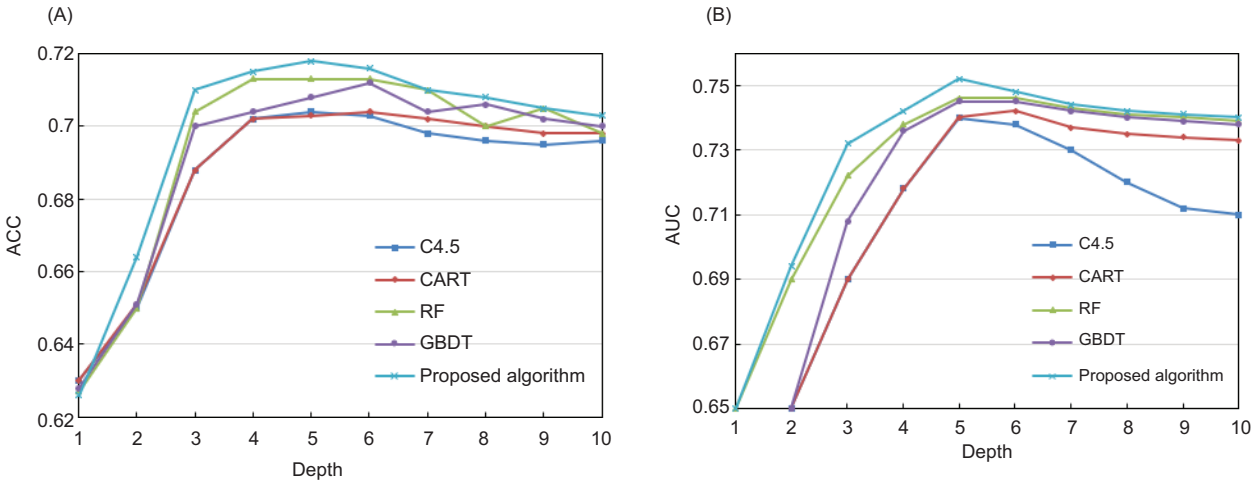

Figure 5. Impact of depth on accuracy (A) and AUC of different algorithms (B).

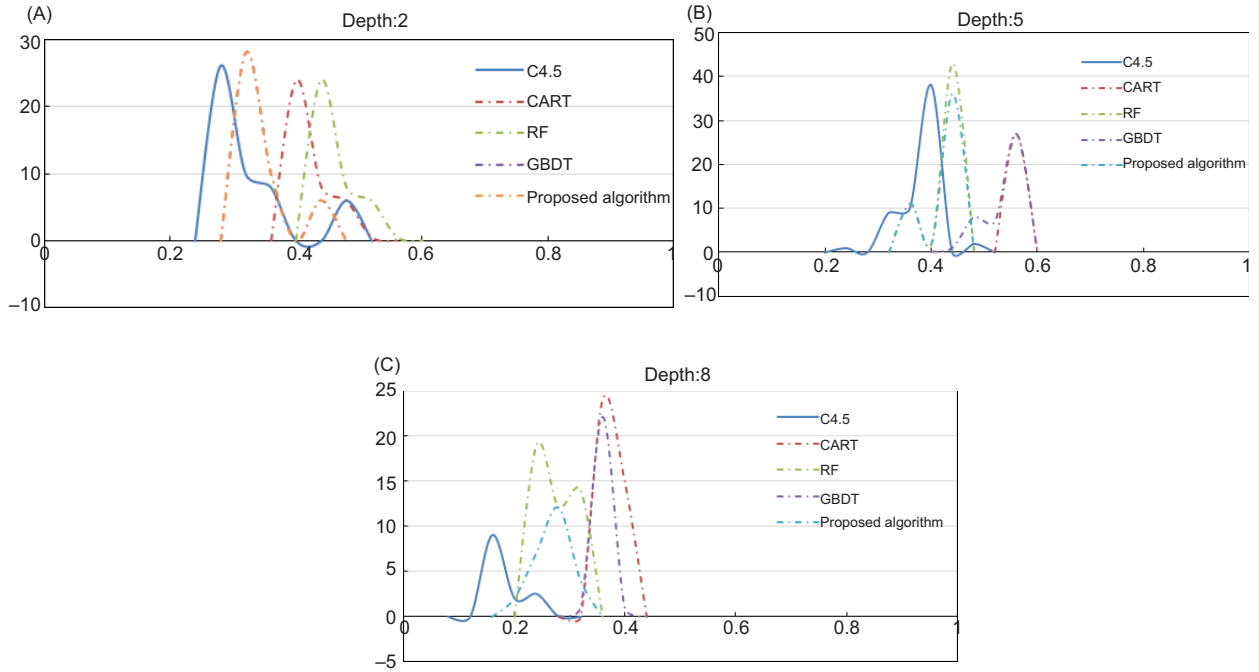

Figure 6. Probability density distribution graphs of several algorithms at various decision tree depths. (A) Depth of 2, (B) Depth of 5, (C) Depth of 8.

Figure 6 clearly shows that when the decision tree depth is set to 2, the C4.5, CART, and RF algorithms have a smaller number of samples in the high-probability region (0.6). This suggests that these algorithms have lower prediction confidence for most samples. The GBDT and the method suggested in this research allocate a greater number of samples in the region with a probability density of 0.4. This indicates that these algorithms have a higher level of prediction confidence for a larger number of samples at this particular depth, compared to other algorithms. When the C4.5 algorithm is applied with a tree depth of 5, it exhibits a more even distribution of prediction probability density. This indicates that the prediction confidence is reasonably spread out.

The CART algorithm exhibits 27 samples inside the 0.4 probability density zone while having minimal representation in other locations. This observation suggests a reasonably elevated although indeterminate level of prediction confidence for the majority of samples at this particular depth. The RF, GBDT, and proposed algorithm prioritize sampling in regions with high probability density, with RF and the proposed algorithm exhibiting a notable concentration of samples in the 0.6 region. This

indicates a high level of prediction confidence for these two methods at a depth of 5. When the tree depth is increased to 8 or beyond in the C4.5 method, the algorithm tends to focus on samples that are located in regions with low probability density. This may suggest a decline in the algorithm's confidence in making accurate predictions. The CART method has a greater number of samples in regions with medium probability density and none in the highest confidence interval, indicating a higher level of uncertainty in its predictions. The random forest model has a rather consistent probability density for predictions in the intermediate range. There are no samples in the extreme probability density regions, suggesting a notable level of confidence in predictions for the majority of samples. The GBDT and the proposed method exhibit a more evenly distributed set of samples, with the proposed algorithm specifically containing samples within the probability density range of 0.2–0.6. This suggests that the proposed algorithm is capable of making predictions with varying levels of confidence at this particular depth.

The analysis reveals that the suggested algorithm exhibits a high level of prediction certainty across different decision tree depths. Notably, at a depth of 5, the algorithm's prediction probability density in the high confidence interval resembles closely that of RF, indicating a robust predictive capability. As the depth of the decision tree increases, the suggested algorithm exhibits a more consistent and evenly distributed prediction probability density. This suggests that the algorithm is stable and less prone to overfitting. By examining the probability density distributions for different tree depths, the proposed algorithm demonstrates superior overall performance, especially in delivering predictions with high levels of confidence.

The performance measures (accuracy, recall, F1-score, and AUC) of the proposed technique are compared to three other methods, namely Dynamic Classifier Selection (DCS), Dynamic Ensemble Selection (DES), and Dynamic Weighted Majority (DWM), on five distinct datasets, as shown in Table 2. The table clearly demonstrates that the suggested technique surpasses DCS, DES, and DWM in terms of accuracy across all datasets. Remarkably, the suggested technique achieves an accuracy of 0.7851 on the B1 dataset, which is much greater than the 0.7451 accuracy of DCS. The suggested technique consistently achieves the greatest accuracy across datasets B2–B5, showcasing its unwavering accuracy across diverse datasets. The recall, which measures the proportion of correctly identified positive samples out of all real positive samples, is particularly notable for the proposed technique on the B1 dataset, with a value of 0.8542. This value is much greater than that of other

Table 2. Performance of different methods in gold standard evaluation recommendations for risk management strategies.

Method	Dataset	Accuracy	Recall	F1-score	AUC
DCS	B_1	0.7451	0.6658	0.7456	0.8456
	B_2	0.8321	0.7214	0.7894	0.9127
	B_3	0.8124	0.7831	0.7789	0.8794
	B_4	0.7286	0.4682	0.6231	0.8632
	B_5	0.7258	0.5124	0.6324	0.8692
DES	B_1	0.7214	0.6239	0.7248	0.8431
	B_2	0.8176	0.6587	0.7546	0.9123
	B_3	0.8216	0.7214	0.7548	0.8743
	B_4	0.7143	0.4689	0.6231	0.8569
	B_5	0.7321	0.5213	0.6589	0.8756
DWM	B_1	0.7389	0.6487	0.7348	0.8451
	B_2	0.8174	0.6523	0.7289	0.8632
	B_3	0.8215	0.7215	0.7541	0.8895
	B_4	0.7369	0.5123	0.6458	0.9123
	B_5	0.7123	0.5489	0.6698	0.8678
Proposed method	B_1	0.7851	0.8542	0.8689	0.9321
	B_2	0.8742	0.7238	0.8241	0.9546
	B_3	0.8863	0.7487	0.8236	0.9143
	B_4	0.8871	0.6123	0.7489	0.9057
	B_5	0.8864	0.6487	0.7874	0.9162

DCS: Dynamic Classifier Selection; DES: Dynamic Ensemble Selection; DWM: Dynamic Weighted Majority; AUC: Area under the Curve.

methods, suggesting the strong ability of the suggested method to identify accurately positive samples. When applied to different datasets, the suggested method consistently achieves a higher or comparable recall rate, compared to the other three methods. This indicates that the proposed method is useful in preventing the omission of crucial risk management tactics. The F1-score, a metric that combines accuracy and recall using the harmonic mean, considers both precision and recall in evaluating the model's performance. The suggested method achieves the highest F1-score across all datasets, particularly on the B1 dataset, where it attains a value of 0.8689. This value indicates a favorable equilibrium between precision and recall. The AUC value is a measure of the model's capacity to classify accurately, and the suggested technique outperforms other methods in terms of AUC values across all datasets, especially on datasets B1 and B2 with the respective AUC values are 0.9321 and 0.9546.

The suggested method exhibits superior performance compared to existing comparable methods (DCS, DES, and DWM) across several datasets, showcasing enhanced efficacy in crucial metrics, including accuracy, recall, F1-score, and AUC. The suggested method exhibits extraordinary performance on the B1 dataset, which may be attributed to its specific benefits in managing the risk management strategies employed in this dataset. The suggested method is a data-driven, multi-criteria decision-making tool that effectively and practically balances cost, benefits, and feasibility to design optimal risk management strategies.

By analyzing the data presented in Figure 7, one can compare the performance of various methodologies in terms of the accuracy of gold standard evaluation recommendations for risk management strategies. On the training set, the accuracy of each approach increases with increase in the number of folds. This is because a higher number of folds allows for more data utilization for training, which enables the models to learn more effectively the features of the data. The precision of the DCS approach improves from 0.868 at two folds to 0.994 at 10 folds, indicating an upgrade in performance as the amount of data grows. Both DES and DWM algorithms exhibit comparable performance, achieving an accuracy of 0.996 at the highest number of folds, which demonstrates their exceptional capacity to adapt to the data.

The proposed method in this paper demonstrates a marginally superior accuracy compared to DCS across all folds, and is on par with DES and DWM in the initial folds. However, its performance slightly declines in the subsequent folds, possibly because of variations in the stability and generalization abilities of the proposed method across different folds. The test set is primarily concerned with evaluating the model's capacity to generalize unfamiliar data. The suggested technique consistently maintains accuracy above 0.928 across all folds, with a peak of 0.934, demonstrating strong and consistent performance on the test set. The DCS technique exhibits variable accuracy ranging from 0.912 to 0.920, consistently lower than the suggested method. This implies that the proposed method may possess superior predictive capability if applied to unfamiliar data. The precision of DES is marginally superior to DCS, although it is still outperformed by the suggested approach, particularly in certain instances where the disparity is more evident. The performance of DWM on the test set is comparable to that of DCS, but it does not exceed the accuracy of the suggested technique.

The suggested method exhibits comparable or slightly improved performance on the training set, compared to

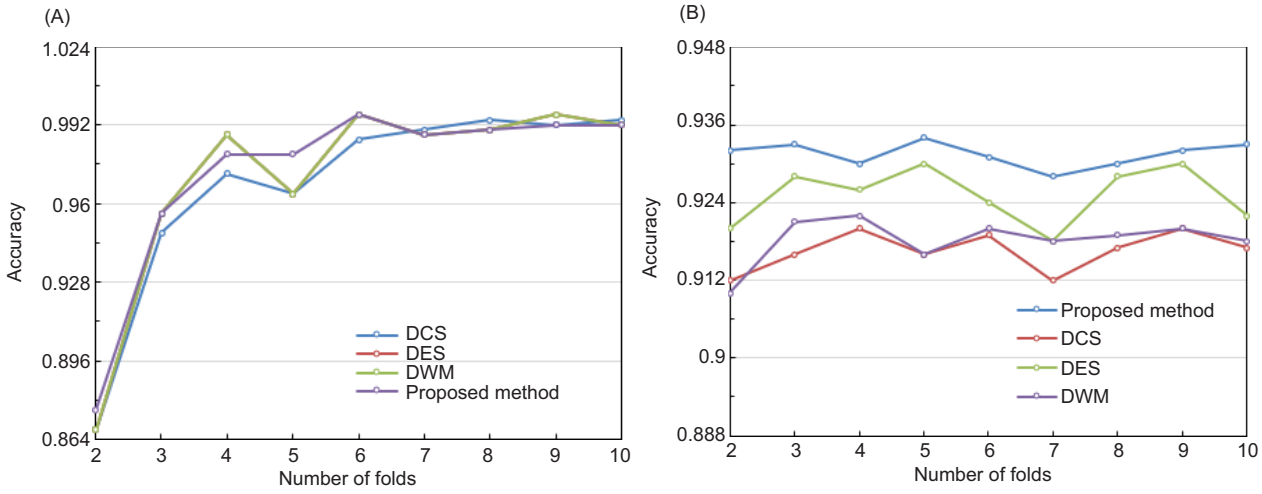

Figure 7. Comparison of gold standard evaluation recommendation accuracy between different methods. (A) Training set and (B) Test set.

the existing methods while demonstrating higher accuracy and stability on the test set. The suggested method demonstrates a stable and consistent high accuracy on the test set, which suggests its strong generalization potential and dependable forecast performance on unfamiliar data. While DES and DWM achieved slightly higher accuracy when using the maximum number of folds on the training set, the suggested technique demonstrated superior performance on the test set, thereby showcasing its effectiveness in practical applications.

Conclusion

This research has elucidated a data-driven, multi-criteria decision-making methodology, meticulously designed to aid managers in the nuanced balancing of diverse risk management strategies within the intricate realm of agricultural supply chains. Fundamental to this study is the implementation of decision tree algorithms, notably those incorporating transition structures, which serve to analyze methodically risk factors and their consequential impacts. Furthermore, the adoption of dynamic selection techniques within a multi-classifier system has been instrumental in augmenting the precision and dependability of evaluations pertaining to risk management strategies. The algorithm proposed herein meticulously constructs a model that assimilates an array of risk factors and their potential interrelations, derived from a comprehensive analysis of historical data, expert evaluations, and practical instances. This model not only facilitates decision-makers in the intuitive discernment of risks and the mapping of their interconnections but also weighs judiciously various elements, such as costs, benefits, and feasibility to devise optimal management strategies.

Empirical evidence demonstrates that this method outperforms traditional risk assessment approaches across various key performance indicators, showcasing superior learning and generalization capabilities. Although the research findings are encouraging, limitations still exist. For example, responding to specific risk factors may require further refined data support, and the applicability of this approach in different cultural and geographical contexts remains to be examined. The future development directions could include expanding the dataset size to accommodate a more diverse range of agricultural environments and further enhancing the adaptability and robustness of the algorithm. Additionally, integrating this method with modern agricultural technologies, such as the Internet of things, devices, and real-time data monitoring, may pave way for more efficient and real-time pathways for agricultural risk management. Moreover, exploring improvements in algorithm interpretability and decision-maker interactivity could be an important part of the future work, ensuring that technological solutions are as comprehensible and applicable to users as possible.

Funding

Heilongjiang Provincial Natural Science Foundation Project "Research on the impact mechanism of relationship embeddedness on enterprise innovation performance—based on inter-organizational science learning ability perspective" (grant No.: LH2019G019).

Data Availability Statement

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declared no conflict of interest.

References

Asfaw, Z., Alemneh, D., and Jimma, W., 2023. Data-driven decision-making and its impacts on education quality in developing countries: a systematic review. In: 2023 International conference on information and communication technology for development for Africa, ICT4DA 2023, pp. 198–203.

Awad, H., Aboalganam, K., Hijazeen, O., Alhanatleh, H., Bakir, S.M.A., Altahrawi, M., et al. 2023. Investigation study of the cloud supply chain management system. Ingénierie des Systèmes d'Information 28(1): 183–188. https://doi.org/10.18280/isi.280120

Cao, Y., Yi, C., Wan, G., Hu, H., Li, Q. and Wang, S., 2022. An analysis on the role of blockchain-based platforms in agricultural supply chains. Transportation Research Part E: Logistics and Transportation Review 163: 102731. https://doi.org/10.1016/j.tre.2022.102731

Chandrasekaran, K.S., Mahalakshmi, V. and Anathapadmanabhan, M.R., 2021. Forecasting parameter strategy using data analytics in supply chain management. Ingénierie des Systèmes d'Information 26(5): 477–482. https://doi.org/10.18280/isi.260507

Chen, L. and Su, S., 2022. Optimization of the trust propagation on supply chain network based on blockchain plus. Journal of Intelligent Management Decision 1(1): 17–27. https://doi.org/10.56578/jimd010103

Chen, M., Zhichuan, W., Zhu, H., Xia, J. and Yue, X., 2022. Research on system dynamics of agricultural products supply chain based on data simulation. In: Proceedings of the 2022 5th international conference on e-business, information management and computer science, pp. 54–66. https://doi.org/10.1145/3584748.3584758

Chennouk, H., Ziyati, E.H. and El Bhiri, B., 2022. Business value creation through project management based on big data approach.

Ingénierie des systèmes d information 27(5): 823–828. https://doi.org/10.18280/isi.270516

Cui, X. and Gao, Z., 2022. Application of internet of things and traceability technology in the whole industrial chain in mountainous areas by blockchain. In: 2022 IEEE 2nd international conference on power, electronics and computer applications (ICPECA), Shenyang, China, pp. 720–723. https://doi.org/10.1109/ICPECA53709.2022.9718882

Dai, M. and Liu, L., 2020. Risk assessment of agricultural supermarket supply chain in big data environment. Sustainable Computing: Informatics and Systems 28: 100420. https://doi.org/10.1016/j.suscom.2020.100420

Ge, H.Y., Wu, X.Y., Jiang, Y.Y., Zhang, Y., Sun, Z.Y., Cui, G.Y., et al. 2023. Research progress and prospect of grain and oil food traceability based on blockchain technology. Nongye Gongcheng Xuebao (Transactions of the Chinese Society of Agricultural Engineering) 39(5): 214–223. https://doi.org/10.11975/j.issn.1002-6819.202209205

Gueham, T. and Merazka, F., 2024. Packet loss concealment method based on hidden Markov model and decision tree for AMR-WB codec. Multimedia Tools and Applications 83(4), 11261–11297.

Hasan, I., Rizvi, S.A.M., Zaman, M., Bakshi, W.J. and Fayaz, S.A., 2022. A scalable framework to analyze data from heterogeneous sources at different levels of granularity. Information Dynamics and Applications 1(1): 26–34. https://doi.org/10.56578/ida010104

He, J., Lin, K.Y. and Dai, Y., 2022. A data-driven innovation model of big data digital learning and its empirical study. Information Dynamics and Applications 1(1): 35–43. https://doi.org/10.56578/ida010105

Hui, D., 2021. Study on risk control of fresh supply chain demand interruption under computer E-commerce environment. Journal of Physics: Conference Series 1992(3): 032053. https://doi.org/10.1088/1742-6596/1992/3/032053

Jiang, S., Wang, T., and Zhang, K-H., 2023. Data-driven decision-making for precision diagnosis of digestive diseases. BioMedical Engineering Online 22(1): 87. https://doi.org/10.1186/s12938-023-01148-1

Krska, R., Leslie, J.F., Haughey, S., Dean, M., Bless, Y., McNerney, O., et al. 2022. Effective approaches for early identification and proactive mitigation of aflatoxins in peanuts: an EU–China perspective. Comprehensive Reviews in Food Science and Food Safety 21(4): 3227–3243. https://doi.org/10.1111/1541-4337.12973

Kusrini, E., Prakoso, I. and Hidayatuloh, S., 2022. Improving efficiency for retail warehouse using data envelopment analysis. Mathematical Modelling of Engineering Problems 9(1): 261–267. https://doi.org/10.18280/mmep.090132

Land, K. and Siraj, A., 2021. Blockchain-based farm-to-fork supply chain tracking. In: 2021 IEEE international conference on big data (big data), Orlando, FL, pp. 3416–3425. https://doi.org/10.1109/BigData52589.2021.9671969

Lazarevska, Z.B., Tocev, T. and Dionisijev, I., 2022. How to improve performance in public sector auditing through the power of big data and data analytics? The Case of the Republic of North Macedonia. Journal of Accounting, Finance and Auditing Studies 8(3): 187–209. https://doi.org/10.32602/jafas.2022.023

Le, T.H.P. and Chu, T-C., 2023. A weighted centroids approach based trapezoidal interval type-2 fuzzy TOPSIS method for evaluating agricultural risk management tools. Soft Computing, 27(22): 17153–17173.

Li, B.Z., Chen, J.X. and Lv, X.T., 2023. Analysis of influencing factors of agricultural products supply chain quality risk based on ISM. In: Second international conference on green communication, network, and internet of things (CNIoT 2022), vol. 12586, pp. 279–286. https://doi.org/10.1117/12.2670199

Li, R. and Gao, Y.Y., 2022. Research on agricultural enterprise accounting information resource-sharing model based on big data technology. Journal of Commercial Biotechnology 27(1): 53–63.

Lin, K.Y. and Hu, L., 2022. Supply and demand optimization of agricultural products in game theory: A state-of-the-art review. Journal of Engineering Management and Systems Engineering 1(2): 76–86. https://doi.org/10.56578/jemse010205

Liu, L., 2022. Research on the operation of agricultural products e-commerce platform based on cloud computing. Mathematical Problems in Engineering 2022: 8489903. https://doi.org/10.1155/2022/8489903

Liu, H., Ye, B., Qin, Z. and Zhang, J., 2022. The high-performance solution with federated learning in supply chain system. In: Proceedings of the 8th international conference on computing and artificial intelligence, pp. 241–245. https://doi.org/10.1145/3532213.3532249

Lu, X., Wang, W., Li, W., Jing, Y. and Li, X., 2022. A generic digital twin framework for collaborative supply chain development. In: 2022 5th International conference on computing and big data (ICCBD), Shanghai, China, pp. 177–181. https://doi.org/10.1109/ICCBD56965.2022.10080555

Modupalli, N., Naik, M., Sunil, C.K. and Natarajan, V., 2021. Emerging non-destructive methods for quality and safety monitoring of spices. Trends in Food Science & Technology 108: 133–147. https://doi.org/10.1016/j.tifs.2020.12.021

Nagahama, D., Nose, T., Koriyama, T. and Kobayashi, T., 2014. Transform mapping using shared decision tree context clustering for HMM-based cross-lingual speech synthesis. In: Proceedings of the annual conference of the international speech communication association, INTERSPEECH, pp. 770–774.

Nagendra, N.P., Narayanamurthy, G., Moser, R., Hartmann, E. and Sengupta, T., 2022. Technology assessment using satellite big data analytics for India's Agri-insurance sector. IEEE Transactions on Engineering Management 70(3): 1099–1113. https://doi.org/10.1109/TEM.2022.3159451

Nguyen, P.H., 2022. Agricultural supply chain risks evaluation with spherical fuzzy analytic hierarchy process. Computers, Materials & Continua 73(2): 4211–4229. https://doi.org/10.32604/cmc.2022.030115

Wang, C. and Wu, Y.G., 2022. Analysis of risk factors of fresh agricultural product supply chain based on grey correlation degree – take Chengdu as an example. In: Modern management based on big data III, Proceedings of MMBD 2022, Vol. 352, pp. 307–312. https://doi.org/10.3233/FAIA220111

Xing, Z.X. and Zhao, D.W., 2013. Research on risks evaluation of the agricultural products supply chain. Journal of Applied Sciences 13(14): 2735–2739.

Xu, J.P., Zhang, B.Y., Zhang, X., Wang, X.Y., Li, F. and Zhao, Y.D., 2022. Optimization of grain and oil quality and safety block chain based on DEMATEL-ISM. Nongye Jixie Xuebao (Transactions of the Chinese Society for Agricultural Machinery) 53(11): 412–423.

Ye, M., 2021. Risk assessment for agricultural supply chain financial sustainable development. In: 2021 3rd International conference on management science and industrial engineering, pp. 95–103. https://doi.org/10.1145/3460824.3460840

Yu, X. and Liang, Y., 2022. Construction strategy of China's agricultural risk management system under the background of COVID-19. In: Proceedings of SPIE – The international society for optical engineering, international conference on agri-photonics and smart agricultural sensing technologies, ICASAST 2022, p. 12349.

Zhai, H., 2023. A dynamic model for risk assessment of cross-border fresh agricultural supply chain. International Journal of Advanced Computer Science and Applications 14(7): 511–518. https://doi.org/10.14569/IJACSA.2023.0140756

Zhang, S., Liu, Z., Zhang, Y. and Luan, C., 2022. Research on patent selection of high-tech enterprises under supply chain risk. In: 2022 International conference on cloud computing, big data and internet of things (3CBIT), Wuhan, China, pp. 93–99. https://doi.org/10.1109/3CBIT57391.2022.00027

Zhang, X., Zhang, X. and Jin, Z., 2023. Big data technology driven 5G network optimization analysis. In: The international conference on cyber security intelligence and analytics. Springer Nature, Cham, Switzerland, vol. 172, pp. 151–159. https://doi.org/10.1007/978-3-031-31860-3_16

Zhao, Y., Zhang, W. and Huang, R., 2022. Research on the impact of big data technology on management accounting. In: 2022 the 5th International conference on data storage and data engineering, pp. 26–32. https://doi.org/10.1145/3528114.3528119