Journal of Data Analytics and Intelligent Decision-making

Journal homepage: jdaid.qom.ac.ir

University of Qom

Intelligent Counterfeit Detection in Supply Chain Through Hybrid Pattern Mining and Blockchain Traceability: A Drug Distribution Case Study

Meysam Jahani*[a], Fatemeh Raji[b]

A. Faculty of Computer Engineering, University of Isfahan, Isfahan, Iran.
B. School of Computer Science and Informatics, De Montfort University, Leicester, United Kingdom.

A R T I C L E I N F O

Keywords:
Detection of Counterfeit Products
Sequential Pattern Mining
Blockchain
Anomaly Detection

A B S T R A C T

The growing number of exchange points in distribution systems has increased the risk of counterfeit product infiltration, posing serious threats to public health and economic stability. Existing anti-counterfeiting strategies, such as blockchain-based traceability and machine learning–driven anomaly detection, remain constrained by vulnerabilities to data manipulation and limited automation. To address these challenges, this study proposes a hybrid approach that integrates sequential pattern mining with blockchain infrastructure for trajectory-based counterfeit detection. The system applies the PrefixSpan algorithm in combination with the longest common subsequence method to detect anomalous trajectories in product distribution networks. Blockchain technology ensures immutability, transparency, and decentralized validation of distribution records, while smart contracts enable automated anomaly detection. Experimental evaluation on a real-world dataset, supplemented with simulated counterfeit trajectories, achieves an overall accuracy of 87.4% and an F1-score of 0.843, outperforming existing models. Moreover, complexity analysis demonstrates the scalability of the proposed framework by offloading computationally intensive tasks to off-chain processes.

1. Introduction

The distribution system represents the flow of products within an enterprise, beginning with the manufacturer and progressing through wholesaler and retail participants until the product reaches the end consumer without disruption or tampering. In recent years, the number of exchange points within distribution systems has increased, particularly in industries such as the

* Corresponding author.
E-mail addresses: jahany.meysam@eng.ui.ac.ir (M. Jahani), f.raji@aston.ac.uk (F. Raji)
Received 23 August 2025; Received in revised form 8 September 2025; Accepted 13 September 2025
Available online 16 September 2025

Cite this article: Jahani, M., Raji, F. (2025). Intelligent Counterfeit Detection in Supply Chain Through Hybrid Pattern Mining and Blockchain Traceability: A Drug Distribution Case Study. *Journal of Data Analytics and Intelligent Decision making*, 1(3), 85 101.
https://doi.org/10.22091/jdaid.2025.14633.1023

pharmaceutical sector. While this expansion improves efficiency and accessibility, it also introduces significant security challenges (Liu et al., 2023).

Counterfeit products are one of the most critical challenges in modern distribution systems (Jahani et al., 2025; Rajalakshmi et al., 2024). These products are deliberately manufactured, labeled, and marketed to imitate authentic goods, thereby deceiving consumers into believing them to be genuine (Sharmila et al., 2023). Counterfeiting poses a serious threat to the global economy, with recent reports estimating that the market for counterfeit goods exceeds USD 4.5 trillion, accounting for more than 3.3% of international trade (Naoum-Sawaya et al., 2023). Furthermore, organizations have reported substantial financial losses, with some incurring costs exceeding USD 65,000 due to counterfeit products (Hohmann et al., 2014).

To address this challenge, various anti-counterfeiting strategies have been proposed, including decentralized frameworks and machine learning–based track-and-trace solutions. Decentralized approaches, particularly those utilizing blockchain technology, have been adopted in supply chains (Agrawal et al., 2022; Bapatla et al., 2024; Dash et al., 2024). Blockchain, as a recent technological innovation, demonstrates strong potential for enhancing product traceability (Ghadge et al., 2023). It functions as a distributed ledger and employs smart contracts to record and track assets across decentralized networks (Khan et al., 2024; Zwitter & Hazenberg, 2020). However, blockchain alone is insufficient and requires complementary methods to automate the detection of anomalous trajectories (Jahani et al., 2025).

Machine learning techniques address this limitation by applying anomaly detection methods and frequent pattern mining algorithms to extract legitimate distribution chain patterns from trace records (Benatia et al., 2022; Ding et al., 2022; Lee & Bang, 2018; Nguyen et al., 2021). These patterns are subsequently analyzed using classification algorithms to identify counterfeit products. The effectiveness of this approach, however, is highly dependent on the reliability of the data. Incomplete, manipulated, or inaccurately recorded data can undermine the performance of frequent pattern mining and classification algorithms, leading to errors in counterfeit detection and a reduction in overall system accuracy.

This paper proposes a counterfeit product detection approach, termed SPM-Chain, which integrates sequential pattern mining with blockchain technology. The approach classifies normal and anomalous trajectories within a distribution system by employing the PrefixSpan sequential pattern mining algorithm in combination with the Longest Common Subsequence (LCS) technique, implemented on a blockchain-secured infrastructure.

The PrefixSpan algorithm is a widely adopted technique for mining frequent subsequences from large-scale sequential datasets and represents a cornerstone in sequential pattern mining research (Sharma & Balakrishna, 2011). Unlike traditional candidate-generation approaches, PrefixSpan reduces computational complexity through a pattern-growth strategy, in which the database is recursively projected according to detected prefixes. This prefix-projection mechanism systematically narrows the search space, allowing frequent subsequences to be identified efficiently without exhaustive enumeration. By preventing candidate explosion and ensuring scalability, PrefixSpan has become a foundational method that underpins numerous applications in sequence analysis and knowledge discovery (Mazumdar & Sarma, 2025).

The longest common subsequence (LCS) is a classical sequence comparison technique which is widely used for assessing similarity between data. Given two or more sequences, LCS identifies the longest subsequence that appears in all of them. Owing to its computational efficiency and robustness in handling noisy or partially matching data, LCS has been extensively applied in diverse domains, including bioinformatics, natural language processing, and trajectory analysis (Nikolic et al., 2021).

The novelty of this study is threefold:

1. SPM-Chain introduces an enhanced frequent-pattern mining approach that integrates smart contracts with LCS-based similarity filtering, enabling more accurate detection of unusual or suspicious product trajectories.
2. SPM-Chain provides a reliable, secure infrastructure that protects supply-chain distribution data from manipulation, tampering, or inaccurate recording.
3. SPM-Chain is rigorously evaluated using real-world distribution data, demonstrating its practical effectiveness in detecting anomalies linked to counterfeit product infiltration.
4. Overall, the proposed SPM-Chain framework is applied to a new application domain within supply chain security and is formulated to address a previously unexplored problem setting, establishing its contribution as both methodologically and practically novel.

This study provides a practical decision-support framework for managers in highly regulated industries, particularly pharmaceutical and healthcare supply chains, to detect counterfeit infiltration with greater accuracy and speed. By integrating secured traceability with sequential pattern mining, SPM-Chain delivers an automated and tamper-resistant mechanism for identifying abnormal product trajectories. This enhances the reliability of distribution records, reduces monitoring and verification costs, and strengthens organizational capabilities to protect consumers, ensure product authenticity, and maintain regulatory compliance in critical sectors such as drug distribution.

The remainder of this paper is organized as follows. Section 2 reviews related work on anti-counterfeiting strategies that employ product traceability and frequent pattern mining. Section 3 presents the proposed algorithm for detecting anomalous trajectories in the distribution system. Section 4 provides the performance evaluation of the proposed approach. Finally, Section 5 concludes the paper and outlines potential directions for future research.

2. Related work

This section reviews recent research addressing the challenges posed by counterfeit products. Prior studies have primarily concentrated on methods that leverage product tracking and tracing, as well as trajectory-based anomaly detection, for counterfeit detection.

2.1. Counterfeit Products Detection Using Track and Trace Methods

Bpatla et al. (2024) proposed an architecture that integrates blockchain, IPFS, and barcode technologies to enhance product traceability and mitigate counterfeiting in the pharmaceutical supply chain. The system records drug pedigree information in compliance with regulations such as the drug supply chain security act (DSCSA), while a novel serialization method reduces costs and facilitates system integration. Off-chain storage is employed to preserve patient access to drug data and minimize blockchain transaction costs. Agrawal et al. (2022) introduced a blockchain-based product recall management system that integrates blockchain technology with a mathematical model for both forward and reverse supply chains. The forward model ensures secure delivery of authentic products from manufacturers to end users, whereas the reverse model reduces the time and cost of recalling defective products. Dash et al. (2024) developed a blockchain-based model to manage transportation risks in the supply chain, addressing challenges such as fraud, traceability, data security, and product loss. The system simulates transportation processes to evaluate risk indicators, with assets and transactions managed on the blockchain through core, blockchain, and user interface modules. The results highlight that participant interaction and feedback are critical for effective risk mitigation.

Jahani et al. (2025) proposed a blockchain-integrated supply chain framework to address the infiltration of counterfeit products. First, they introduced a reliable manufacturer selection algorithm that enhanced product quality by incorporating participant evaluations. Next, they developed a smart traceability algorithm that monitored product movement through sensor data, thereby preventing the circulation of substandard items within the supply chain. Additionally, they designed two counterfeit tag detection algorithms that leveraged a weighted graph to simulate the shortest distribution paths, enabling the identification of cloning, application, and modification attacks on product tags. However, although these systems enable counterfeit tracking, none provide a comprehensive automated solution for counterfeit detection across the distribution system, leaving significant gaps in real-time identification. Moreover, they rely heavily on manual inspection for traceability and perform poorly in dynamic counterfeiting scenarios, such as product distribution from anomalous trajectories.

2.2. Counterfeit Products Detection Using Pattern-Based and Data-Driven Methods

Lee and Bang (2018) proposed a frequent pattern data mining algorithm to identify normal supply chain patterns from tracking records, combined with a classification algorithm for counterfeit detection. Their approach constructs a tree structure using regular expressions to extract frequent trajectories, thereby overcoming threshold limitations in existing methods. Benatia et al. (2022) introduced a framework for tracking and detecting counterfeit products in supply chains. The system applies the Apriori algorithm to extract frequent valid product trajectories, with counterfeit products detected by comparing actual movements against these reference paths. Training data are generated through supply chain simulations using a multi-agent system in AnyLogic. Nguyen et al. (2021) proposed two data-driven approaches to improve decision-making in supply chain management. Specifically, the study introduced an LSTM-based method for forecasting multivariate time series data and an LSTM autoencoder combined with a one-class support vector machine for anomaly detection in sales. Another study (Ding et al., 2022) addressed the challenge of efficiently mining frequent trajectory patterns from large and uncertain trajectory datasets for location prediction and location-based services. The proposed algorithm incorporates grouping and partitioning for preprocessing, prefix pruning to eliminate redundant patterns, and cluster memory computing in Spark to enhance computational efficiency. However, these approaches face notable limitations. The use of insecure infrastructures for data storage and management exposes systems to risks of trajectory alterations and data tampering, undermining the reliability. Furthermore, many experiments rely solely on virtual datasets, reducing real-world applicability.

3. The Proposed Counterfeit Product Detection Approach

SPM-chain integrates blockchain, PrefixSpan, and longest common subsequence (LCS) techniques to detect anomalous trajectories in product distribution. The system architecture, illustrated in Figure 1, is organized into two main sections. Trajectory data acquisition and storage consists of the product trajectory layer (PTL), which captures the movement of products across the distribution system, and the blockchain data layer (BDL), which ensures decentralized, immutable storage of these trajectory records. Trajectory detection and analysis comprises the on-chain execution layer (OEL), which enforces predefined anomaly detection policies through contract-based execution, and the off-chain analysis layer (OAL), which applies sequential pattern mining and similarity measures to identify suspicious product flows. The operational mechanisms of both the OEL and OAL are formalized through algorithmic descriptions.

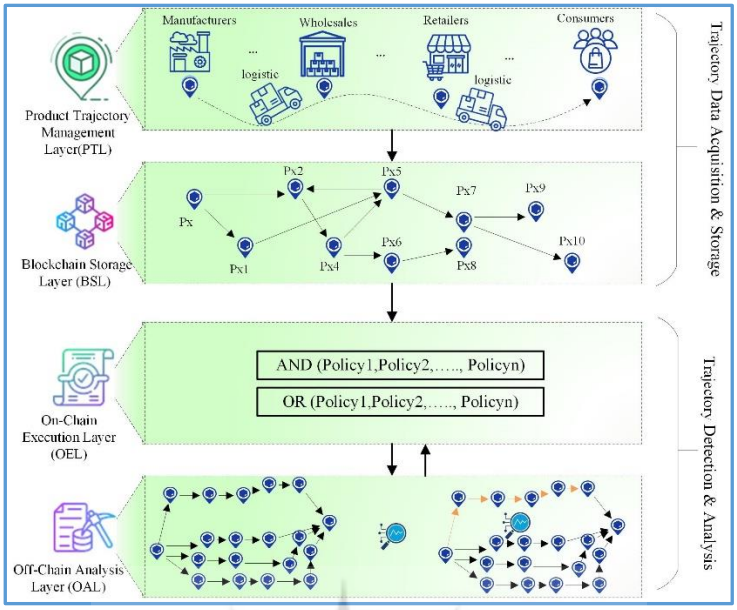


Figure 1

Overview of the Proposed System Architecture

These sections are described in detail in the following and with notation in Table 1.

Table 1

The Notations Used in SPM-Chain

Abbreviation	Full Name	Function and Role in the System
prefix	Prefix	The current trajectory prefix used to extend the pattern into its subsequences.
similarity	Similarity	Parameter that determines the degree of similarity between trajectories.
minSup	Minimum Support	The minimum support threshold for identifying frequent product distribution patterns; manually set.
maxPatLen	Maximum Pattern Length	Parameter that defines the maximum allowed length for a product distribution trajectory.
deReqDB	Demand Requests Database	Database of product orders, including order quantity and order date.
newTrj	New Trajectory	The new product trajectory to be analyzed for similarity against existing patterns.
RT	Reference Trajectory	The reference (standard) trajectory pattern sequence.

TL	Trajectory Label	Classification label indicating whether the trajectory is anomalous or normal.
findFrequentItems (deReqDB, minSup)	Find Frequent Items	Identifies frequent nodes (e.g., manufacturers or wholesale) whose occurrences in all trajectories exceed the minimum support threshold (*minSup*) in *deReqDB*.
projectDatabase (deReqDB, prefixItem)	Projected Database	Generates the search space for prefixes of the discovered frequent patterns (*prefixItem*).
LCS (pattern, newTrj)	Longest Common Subsequence	Computes the length of the longest common subsequence between a frequent pattern and a new trajectory; forms the basis for similarity evaluation.
Sim (newTrj, RT)	Similarity	Function that determines the degree of similarity between the new trajectory and the reference trajectory.

3.1. Trajectory Data Acquisition and Storage

SPM-chain introduces a blockchain-based approach for managing product distribution by securely storing and validating product-related data. As illustrated in Figure 1, the product trajectory layer (PTL) models the distribution flow, beginning with manufacturers, followed by wholesalers, retailers, and ultimately consumers. The roles of each participant are defined as follows:

- Manufacturers: Convert raw materials supplied by vendors into finished products, label them appropriately, and prepare them for distribution.
- Logistics providers: Collect packages from manufacturers and deliver them to wholesalers.
- Wholesalers: Purchase products in bulk from logistics providers and redistribute them to retailers in smaller quantities, often through carriers. They may include distribution companies or healthcare centers.
- Retailers: Acquire products from wholesalers and supply them directly to consumers.
- Consumers: Represent the final stage in the distribution system, receiving products from retailers and verifying their authenticity prior to use.

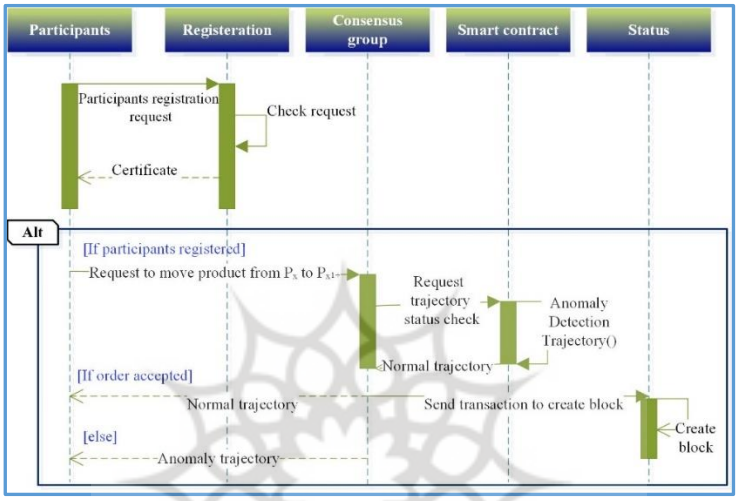
At each stage, participants record product-related location data, thereby creating a traceable history of the distribution process in the blockchain data layer (BDL). Product trajectories and transaction records are stored immutably on the blockchain, ensuring transparency, traceability, and tamper resistance across all participants. SPM-chain organizes these records into blocks, where each block represents transactions associated with a specific product location. Transaction validation is governed by smart contracts, which enforce policies based on consensus rules (e.g., logical conditions such as AND/OR: Policy1, Policy2, Policyn).

As illustrated in Figure 2, the sequence of operations proceeds as follows. First, participants are registered in the network and assigned cryptographic keys (public/private). When a product

is transferred between entities, the corresponding trajectory point and location are automatically recorded on the blockchain. The transaction is then forwarded to a consensus group, which verifies its validity before sending it to the smart contract. The smart contract invokes the AnomalyDetectionTrajectory algorithm to classify trajectories as normal or anomalous. If an anomalous trajectory is detected, SPM-chain issues a notification indicating the potential circulation of counterfeit products within the distribution system.

Figure 2

Transaction Flow in SPM-Chain in Blockchain (Jahani et al., 2025)



3.2. Trajectory Detection and Analysis

The primary objective of this section is to detect potentially counterfeit products that may be distributed through unusual trajectories. SPM-chain identifies normal distribution patterns by analysing frequent sequences in product distribution trajectories and detecting suspicious routes that may indicate counterfeit activity. Such anomalies may involve irregular product circulation across the distribution network, the use of multiple intermediaries with potential collusion, or deviations in distribution timing. Since product distribution trajectories generally follow stable patterns, the analysis of historical distribution data enables the detection of meaningful deviations and abnormal behaviors. This approach strengthens monitoring capabilities within the distribution system and improves the detection of counterfeit products.

Figure 3 presents the SPM-chain algorithm for anomalous trajectory detection. The algorithm employs a hybrid execution model that integrates both the on-chain execution layer (OEL) and the off-chain analysis layer (OAL). Within the OEL, the system verifies initial conditions, such as evaluating trajectory length. In contrast, computationally intensive operations, such as trajectory pattern extraction, sequential pattern mining, and anomaly detection, are executed in the OAL to reduce storage and computational overhead on the blockchain. The results are subsequently transmitted to and stored within the OEL.

In the SPM-chain, the inputs are represented as a set X of sequences corresponding to distribution system participants, including manufacturers, wholesalers, retailers, and consumers. The output consists of frequent path sequences derived from ordered subsets of X. More specifically, each input pair from X can be expressed as a sequence pair (x, y). The spatial order of transaction events is critical; for example, the consecutive event pairs {B, A} and {A, B} must be considered distinct itemsets.

In the first step (lines 1–4), the function patternAnomalyDetection examines the input trajectory *newTrj*. At this stage, if the trajectory contains fewer than two steps, does not begin with the manufacturer entity, or exceeds the predefined number of standard trajectory steps (*stc*), it is classified as anomalous. This rule-based inspection, implemented in the OEL, enables SPM-chain to detect a subset of trajectory anomalies without the need for extensive analysis.

Figure 3

Proposed SPM-Chain in Detection Anomaly Trajectory

Input: *deReqDB, prefix, minSup, maxPatLen, stc, newTrj*

Output: *" Anomaly trajectory" or " Normal trajectory"*

```
1   AnomalyDetectionTrajectory
2     if length(newTrj) < 2 OR NOT startsWith(newTrj, "Manufacturer")  OR length(newTrj) > stc
    then
3     return " Anomaly trajectory "
4     end if
5     patterns = ∅
6     freqItems = findFrequentItems(deReqDB, minSup)
7     if freqItems is empty then
8        return patterns
9     end if
10    for each item β in freqItems do
11       newPattern = prefix + β
12       patterns = patterns ∪ newPattern
13       sProjected = projectDatabase(deReqDB, β)
14       recursivePatterns = prefixSpan (sProjected, newPattern, minSup)
15       patterns = patterns ∪ recursivePatterns
16    end for
17    for each pattern in patterns do
18       similarity = lCS (pattern, newTrj)
19       if similarity < threshold then
20          return " Anomaly trajectory"
21       end if
22    end for
23    return " Normal trajectory"
```

```
24  end function
25  function findFrequentItems (deReqDB, minSup)
26      freqItems = ∅
27      Count frequencies of each item in db
28      return items with frequency ≥ minSup
29  end function
30  function projectDatabase (deReqDB, prefixItem)
31      sProjected = ∅
32      for each sequence s in db function
33          index = findFirstOccurrence (s, prefixItem)
34          if index ≠ -1 then
35              suffix = s[index + 1 : end]
36              sProjected = sProjected ∪ suffix
37          end if
38      end for
39      return sProjected
40  end function
```

Next, frequent distribution trajectories are extracted using a predefined threshold. To achieve this, the PrefixSpan algorithm is applied, as it efficiently mines frequent sequential patterns without the need to generate candidate sets. The product distribution system represents a large and complex search space with a hierarchical structure, where trajectories often involve multiple intermediary participants, such as wholesalers and retailers. Given these characteristics, the use of PrefixSpan is particularly well suited. Its sequential nature and ability to process ordered data enable more accurate modeling of product distribution behavior and facilitate the identification of both anomalous and recurring patterns. This capability is especially important for detecting abnormal distribution behaviors, such as deviations from authorized routes or the emergence of informal distribution channels.

Subsequently, a set of frequent participants (e.g., manufacturers) is extracted from the deReqDB database using the findFrequentItems function (lines 6–11). These participants represent key elements filtered by the minimum support threshold (minSup). If no frequent anomaly trajectory is identified at this stage, the function terminates without further analysis. In the next step (lines 13–19), for each frequent item, a new pattern is created by appending it to the current prefix. Using prefix projection, a new projected database (sProjected) is constructed, containing only the suffixes that appear after the first occurrence of that item (i.e., a wholesale or logistic appended as the next step to each of the previously frequent items). The algorithm is then recursively executed to extract sequential frequent patterns, all of which are added to the patterns set. This stage constitutes the core process of sequential pattern discovery. During the anomaly analysis phase (lines 22–27), the input path *newTrj* is compared against

each discovered frequent pattern. The comparison is performed using the longest common subsequence (LCS) similarity function (Sim (*newTrj,RT*)). If the similarity ratio (the length of the LCS (*newTrj,RT*) divided by the length of the new trajectory (*newTrj*), as defined in Eq. (1) falls below a predefined threshold, the trajectory is classified as an "Anomaly trajectory." Otherwise, the trajectory is considered valid and is labeled as a "Normal trajectory" (Eq. (2)). Notably, in Eq(1), a new trajectory sequence is indicated with *newTrj* and a reference pattern sequence with *Q*.

$$\mathrm{Sim}(\mathrm{newTrj},\mathrm{RT})=\frac{|\mathrm{LCS}(\mathrm{newTrj},\mathrm{RT})|}{|\mathrm{newTrj}|} \quad (1)$$

$$\mathrm{Label}(\mathrm{CT})=\begin{cases}\text{Anomaly trajectory} & \text{if } \mathrm{Sim}(\mathrm{newTrj},\mathrm{RT}) < \text{threshold}\\ \text{Normal trajectory} & \text{if } \mathrm{Sim}(\mathrm{newTrj},\mathrm{RT}) \geq \text{threshold}\end{cases} \quad (2)$$

The auxiliary functions findFrequentItems() (lines 32–36) and projectDatabase() (lines 38–48) are responsible for extracting frequent items and constructing the projected database, respectively. These functions implement the core components of the PrefixSpan algorithm, which efficiently mines sequential patterns from the data. Specifically, in findFrequentItems(), the frequency of each item in the sequences of deReqDB is calculated, and only those items that meet the minimum support criterion are returned as frequent items. In projectDatabase(), when provided with *deReqDB* and a frequent item (prefixItem), a new database is generated. For each sequence in *deReqDB*, the first occurrence of prefixItem is located using findFirstOccurrence(), and the corresponding suffix is extracted. These suffixes are collected in a set named sProjected, which is then returned as the function output.

4. Evaluation

This section describes the processes of data collection, preprocessing, and evaluation, with particular emphasis on security, computational complexity, and accuracy in detecting anomalous trajectories. SPM-chain is evaluated within a drug distribution network as a case study. The findings demonstrate its potential to improve distribution management by enabling reliable detection of anomalous trajectories within a secure infrastructure. The approach can substantially reduce the circulation of counterfeit drugs by identifying irregular distribution paths.

4.1. Data Collection and Preprocessing

In this study, both real-world data from official repositories and simulated data based on drug distribution system scenarios were utilized to enhance the credibility of the results and ensure a comprehensive experimental evaluation. In total, 58,804 records were collected, providing a robust foundation for statistical analysis, performance assessment, and outcome validation. For performance evaluation, data were sourced from the automation of reports and consolidated orders system (ARCOS) database. ARCOS, managed by the U.S. drug enforcement administration (DEA), records all legal transactions of controlled substances between 2006 and 2014 (Amico et al., 2024). The database contains more than 500 million transaction records, capturing the distribution of medications from manufacturers to distributors, pharmacies, hospitals, and physicians. In this study, 52,804 ARCOS records were selected to design and evaluate multiple experimental scenarios. To prepare the data for advanced analysis, the following preprocessing steps were applied:

- Reconstruction: Complete distribution trajectory were reconstructed from dyadic transaction records (Amico et al., 2024).
- Selection of sales transactions: Because ARCOS records each transaction in duplicate (purchase and sale), only sales transactions to end consumers were retained to eliminate duplicates and prevent double-counting.
- Removal of incomplete or ambiguous data: Transactions from early 2006 were flagged as low reliability due to incomplete initial inventory records and were used only for secondary analyses.
- Standardization of formats: All dates were standardized to the daily level to enable temporal analysis. Negative or zero values, which lacked analytical relevance, were filtered out.
- Exclusion of exceptions: International transactions (imports/exports), representing only 0.01% of the dataset, were excluded from the main analysis.

Table 2 presents an example of a drug distribution trajectory involving three participants along with transaction dates and transferred quantities.

Table 2

Example of a Drug Distribution Trajectory

Participant 1 (Manufacturer)	Participant 2 (Wholesaler Distributor)	Participant 3 (Pharmacy Consumer)	Transfer Date 1→2	Transfer Date 2→3	Quantity
RR0284632	RP0257938	PT0165983	2006/02/08	2006/02/10	122,767

4.1.1. Simulated Data

The simulated datasets in this study were designed using a combination of field-collected data, official organizational reports, and information extracted from scientific publications and reputable international repositories (Kellyrx, 2024; Mackey & Nayyar, 2017; Mackey et al., 2024). These sources provided critical insights into counterfeit drug distribution practices, illegal distribution networks, and common fraudulent patterns within the drug distribution system. Based on these inputs, approximately 6,000 anomalous drug distribution trajectories associated with counterfeit drugs were simulated, with their characteristics summarized in Table 3. To capture the dynamic behavior of the drug distribution system and to generate probabilistic scenarios under uncertainty, the Monte Carlo simulation method was employed. This method was chosen for its strong capacity to model stochastic processes and evaluate rare but meaningful events. In the simulation, each trajectory comprised a random number of participants (e.g., manufacturers, wholesalers, retailers, and consumers), structured to reflect real-world distribution patterns. The transit time of drugs between participants was modeled using a normal distribution with a mean of three days and a proportional standard deviation. Approximately 12% of the simulated trajectories (around 6,000) were classified as anomalous. Specifically, in 5% of the cases, drugs entered the distribution system through wholesalers or retailers without the involvement of manufacturers. In 1% of the cases, manufacturers were simulated as sources of counterfeit drugs containing inactive or substandard ingredients. Finally, in 6% of the cases, drugs exited the official supply chain and entered unauthorized distribution networks.

Table 3

Probability of Counterfeit Drug Distribution

Condition of counterfeit drug distribution	Probability	Reference
Percentage of counterfeit drugs in total	10–12%	Mackey et al. (2015)
Drugs entering the distribution system without the involvement of manufacturers	5%	Kellyrx (2024)
Manufacturers identified as sources of counterfeit drugs	<1%	Kellyrx (2024)
Drugs leaving the official supply chain and entering unauthorized distribution networks	6%	Mackey & Nayyar (2017)

4.2. Comparison of SPM-Chain in Terms of Time Complexity

The algorithmic complexity of SPM-chain was evaluated in comparison with BDM (Xiao et al., 2022), DSCAM (Datta & Namasudra, 2024), and DTSB (Kochovski et al., 2024), with particular attention to their execution within smart contracts. For this analysis, it was assumed that each function call completes in a single time unit. Accordingly, the complexity of the algorithms is determined by the number of iterations required for a given input size.

In SPM-Chain, the smart contract verifies initial conditions, such as assessing the trajectory length and transmitting the result. Since each smart contract operation executes in constant time and is implemented using nested conditionals without loops, the overall time complexity for n inputs is O(n). As summarized in Table 4, the complexity of SPM-Chain is comparable to other blockchain-based models. This efficiency is achieved by delegating computationally intensive tasks to off-chain (OAL) processes, while restricting on-chain (OEL) execution to lightweight comparison and validation functions.

Table 4

Complexity Comparison of SPM-Chain and Related Models Implemented On-Chain

Model	SPM-Chain	DTSB	DSCAM	BDM
Complexity	O(n)	O(n)	O(n)	O(n)

Table 5 compares the performance of the proposed system's off-chain module for frequent-pattern discovery against two benchmark approaches, DCPSCE (Lee & Bang, 2018) and DCPFPM (Benatia et al., 2022). The following parameters are considered across all models:

- n: number of sequences in the dataset
- l: average sequence length
- m: number of distinct sequences

- p: number of frequent sequential patterns detected

The time complexity of these algorithms is primarily influenced by the number of frequent sequences and their distribution within the dataset. For DCPFPM and DCPSCE, the complexity can be expressed as O(n ·m · p · l). In contrast, sequence analysis in SPM-Chain is constrained by a maximum depth (*stc*), beyond which sequences are discarded. Additionally, only sequences containing at least two steps *(k ≥ 2)* and those starting with the producer (*d*) are considered. These restrictions reduce the number of candidate sequences, thereby lowering computational costs. As a result, the effective number of sequences can be approximated as $n - (d + k + stc) \approx n$. Taking these factors into account, SPM-Chain achieves both theoretically and practically lower complexity compared with DCPFPM and DCPSCE.

Table 5

Complexity Comparison of SPM-Chain and Related Models Implemented Off-Chain

Model	*DCPFPM*	*DCPSCE*	*SPM-Chain*
Complexity	O(n · m · p · l)	O(n · m · p · l)	O((stc − (d + k)) · m · p · l)

4.3. Security Analysis of proposed system

In this section, the security of SPM-Chain is analyzed based on key security concerns, data integrity, non-repudiation, resistance to distributed denial-of-service, access-level security evaluation.

4.3.1. Data Integrity

In SPM-Chain, data integrity is ensured through mechanisms that prevent stored information from being tampered with or altered. Each block contains a cryptographic hash that functions as a digital fingerprint, linking it to the preceding block. This hash is derived from the block's contents, including the data, timestamp, and the previous block's hash. Even minimal modifications to a block's content alter the hash value, thereby exposing tampering attempts. To further safeguard integrity, the system employs consensus mechanisms to validate data, ensuring that information is stored only when a subset of nodes verifies its correctness. While collusion within the consensus group remains a theoretical risk, the integration of smart contracts and frequency pattern mining-based evaluation substantially reduces the likelihood of fraudulent activity.

4.3.2. Non-repudiation

In the proposed system, users are registered and authenticated in blockchain, which also manage their access permissions. Each transaction is digitally signed using the sender's private key. Since transactions are uniquely linked to the sender and include the recipient's identity, no party can repudiate the transaction once it has been executed.

4.4. Evaluation of SPM-Chain for Detecting Counterfeit Drugs in Anomalous Trajectory

Table 6 reports the evaluation results of the anomaly detection algorithm using the dataset introduced in Section 5.1. Within this dataset, 25% of the trajectories correspond to standard distributions, while 12% involve counterfeit drugs. In Table 6, a value of 1 indicates an anomalous distribution trajectory, whereas a value of 0 denotes a normal trajectory.

SPM-Chain achieves an overall classification accuracy of 0.874, demonstrating strong generalization capability and resistance to overfitting. For anomalous patterns, the model attains recall and precision values of 0.847 and 0.839, respectively. The corresponding F1-score of 0.843, together with macro-average recall and precision values of 0.868 and 0.869, further substantiates the robustness and reliability of SPM-Chain compared with benchmark models.

Table 6

SPM-Chain Performance in Detecting Counterfeit Drug

Class	Precision	Recall	F1-score
0	0.898	0.892	0.895
1	0.839	0.847	0.843
Macro Avg	0.868	0.869	0.869
Accuracy			0.874

Table 7 compares SPM-Chain with DCPSCE (Lee & Bang, 2018) and DCPFPM (Benatia et al., 2022) in terms of anomaly detection performance. A notable strength of SPM-Chain is its ability to evaluate natural and anomalous patterns independently of the training process, whereas DCPFPM relies on the same training data for evaluation, potentially introducing bias. The results demonstrate that SPM-Chain achieves a more balanced trade-off between accuracy and F1-score. Its higher F1-score highlights a superior capability to correctly identify both normal and anomalous classes. Although DCPFPM attains marginally higher overall accuracy, its lower F1-score indicates weaker detection of positive samples (i.e., anomalous trajectories), likely due to class imbalance or evaluation on training data. These findings suggest that SPM-Chain provides more reliable and robust performance in error-sensitive contexts, such as anomaly detection in a drug distribution system. In terms of overall priority, SPM-Chain outperforms both DCPFPM and DCPSCE.

Table 7

Performance Comparison of SPM-Chain with Existing Models in Detecting Counterfeit Drug

Models	Accuracy	F1-score	*Dataset*
SPM-chain	0.87	0.84	ACROS+ Simulated
DCPFPM	0.89	0.76	Simulated
DCPSCE	0.80	–	Simulated

4.5. Management Contribution

This study offers a practical framework for managers in highly regulated industries, particularly in pharmaceutical and healthcare supply chains, to detect counterfeit infiltration with greater accuracy and speed. By integrating blockchain-secured traceability with sequential pattern mining, SPM-Chain provides an automated and tamper-resistant mechanism for identifying abnormal product trajectories. This enhances the reliability of distribution records, reduces monitoring costs, and strengthens organizational capability to safeguard patients, ensure product authenticity, and maintain compliance in critical sectors such as drug distribution.

5. Conclusion and Future Work

This study introduces SPM-Chain, a hybrid approach that integrates sequential pattern mining and blockchain to detect anomalous trajectories in product distribution systems. By combining the PrefixSpan algorithm with the longest common subsequence method, the approach enables reliable classification of normal and anomalous product flows. The blockchain infrastructure, augmented with smart contracts, ensures tamper-resistant traceability, decentralized validation, and automated anomaly detection. Experimental evaluation, using real-world ARCOS data and simulated counterfeit trajectories, demonstrated strong performance, achieving an overall accuracy of 87.4% and an F1-score of 0.843, outperforming existing benchmark models. Beyond classification effectiveness, the proposed approach reduces computational complexity by leveraging a hybrid on-chain/off-chain execution model, thereby ensuring both scalability and operational efficiency.

These findings provide several important managerial implications for organizations aiming to enhance supply-chain security and mitigate the risks of counterfeit product infiltration. First, SPM-Chain offers managers a practical, data-driven tool that is capable of detecting suspicious products, enabling faster and more informed decision-making within distribution networks. By integrating smart contracts with similarity-based filtering, the framework facilitates more accurate anomaly detection, allowing managers to proactively intervene before counterfeit products spread throughout the system. Second, the blockchain-secured infrastructure of SPM-Chain ensures that distribution data remains tamper-resistant and reliably recorded across all supply-chain partners. This significantly enhances trust, transparency, and coordination in multi-tier supply chains, where data manipulation or inaccurate reporting are common challenges. Managers can leverage these immutable records to improve auditability, compliance monitoring, and cross-organizational accountability. Despite these promising results, several directions remain for future research. First, extending the framework to support real-time streaming data would improve responsiveness in dynamic distribution environments. Second, incorporating advanced deep learning models, such as graph neural networks or attention-based architectures, could enhance the detection of subtle and evolving counterfeit patterns. Finally, real-world deployment and longitudinal evaluation will be essential to validate the system's usability, trustworthiness, and long-term impact on counterfeit mitigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

Supplementary data are available in supplementary files or on https://www.slcg.com/opioid-data

Acknowledgements

The author thanks the anonymous referee and the editor-in-chief for their close attention and helpful comments.

Reference

Agrawal, D., Minocha, S., Namasudra, S., & Gandomi, A. H. (2022). A robust drug recall supply chain management system using hyperledger blockchain ecosystem. *Computational Biology and Medicine, 140*, 105100. https://doi.org/10.1016/j.compbiomed.2021.105100

Amico, A., Verginer, L., & Schweitzer, F. (2024). Tracing opioids across the US: A high-resolution pharmaceutical distribution dataset. *Scientific Data*, *11*(1), 1-8. https://doi.org/10.1038/s41597-024-03534-3

Bapatla, A. K., Mohanty, S. P., & Kougianos, E. (2024). PharmaChain 3.0: Efficient tracking and tracing of drugs in pharmaceutical supply chain using blockchain integrated product serialization mechanism. *SN Computer Science*, *5*(1), 22-32. https://doi.org/10.1007/s42979-023-02510-9

Benatia, M. A., Baudry, D., & Louis, A. (2022). Detecting counterfeit products by means of frequent pattern mining. *Journal of Ambient Intelligence and Humanized Computing*, *13*(7), 3683–3692. https://doi.org/10.1007/s12652-020-02237-y

Datta, S., & Namasudra, S. (2024). Blockchain-based secure and scalable supply chain management system to prevent drug counterfeiting. *Cluster Computing*, *3*(7), 9243–9260. https://doi.org/10.1007/s10586-024-04417-3

Dash, S., Ghugar, U., Godavarthi, D., & Mohanty, S. N. (2024). HCSRL: Hyperledger composer system for reducing logistics losses in the pharmaceutical product supply chain using a blockchain-based approach. *Scientific Reports*, *14*(1), 1-20. https://doi.org/10.1038/s41598-024-61654-7

Ding, J., Li, Y., Li, L., & Jia, L. (2022). Prefix-pruning-based distributed frequent trajectory pattern mining algorithm. *Mathematical Problems in Engineering*, *2022*. https://doi.org/10.1155/2022/3838147

Ghadge, A., Bourlakis, M., Kamble, S., & Seuring, S. (2023). Blockchain implementation in pharmaceutical supply chains: A review and conceptual framework. *International Journal of Production Research*, *61*(19), 6633–6651. https://doi.org/10.1080/00207543.2022.2125595

Hohmann, N., Mikus, G., & Czock, D. (2014). Effects and risks associated with novel psychoactive substances. *Deutsches Ärzteblatt International*, *111*(9), 139–147. https://doi.org/10.3238/arztebl.2014.0139

Jahani, M., Raji, F., & Zojaji, Z. (2025). Securing supply chain through blockchain-integrated algorithmic system: Ensuring product quality and counterfeiting tags detection. *Cluster Computing*, *28*(1), 51. https://doi.org/10.1007/s10586-024-04764-1

Jahani, M., Zojaji, Z., & Raji, F. (2025). Blockchain-driven peer-to-peer system: Elevating trust in pharmaceutical manufacturer selection through BERT-based sentiment analysis. *Peer-to-Peer Networking and Applications*, *18*(3), 159. https://doi.org/10.1007/s12083-025-01975-0

Kellyrx. (2024). How do counterfeit drugs enter the supply chain? https://www.idlogiq.com/how-do-counterfeit-drugs-enter-the-supply-chain/#:~:text=The

Khan, A. A., Laghari, A. A., Baqasah, A. M., et al. (2024). Blockchain-enabled infrastructural security solution for serverless consortium fog and edge computing. *PeerJ Computer Science*, *10*, 1–34. https://doi.org/10.7717/peerj-cs.1933

Kochovski, P., Masmoudi, M., Bouhamoum, R., Stankovski, V., Baazaoui, H., Ghedira, C., Vodislav, D., & Mecharnia, T. (2024). Drug traceability system based on semantic blockchain and on a reputation method. *World Wide Web*, *27*(5), 62. https://doi.org/10.1007/s11280-024-01301-3

Liu, S., Zhang, R., Liu, C., & Shi, D. (2023). P-PBFT: An improved blockchain algorithm to support large-scale pharmaceutical traceability. *Computational Biology and Medicine*, *154*, 106590. https://doi.org/10.1016/j.compbiomed.2023.106590

Mackey, T. K., Liang, B. A., York, P., & Kubic, T. (2015). Counterfeit drug penetration into global legitimate medicine supply chains: A global assessment. *American Journal of Tropical Medicine and Hygiene*, *92*(6), 59–67. https://doi.org/10.4269/ajtmh.14-0389

Mackey, T. K., & Nayyar, G. (2017). A review of existing and emerging digital technologies to combat the global trade in fake medicines. *Expert Opinion on Drug Safety*, *16*(5), 587–602. https://doi.org/10.1080/14740338.2017.1313227

Mazumdar, N., & Sarma, P. K. D. (2025). Sequential pattern mining algorithms and their applications: A technical review. *International Journal of Data Science and Analytics*, *20*(3), 1683–1726. https://doi.org/10.1007/s41060-024-00659-x

Naoum-Sawaya, J., Elhedhli, S., & De Carvalho, P. (2023). Strategic blockchain adoption to deter deceptive counterfeiters. *European Journal of Operational Research*, *311*(1), 373–386. https://doi.org/10.1016/j.ejor.2023.04.031

Nguyen, H. D., Tran, K. P., Thomassey, S., & Hamad, M. (2021). Forecasting and anomaly detection approaches using LSTM and LSTM autoencoder techniques with applications in supply chain management. *International Journal of Information Management*, *57*, 102282. https://doi.org/10.1016/j.ijinfomgt.2020.102282

Nikolic, B., Kartelj, A., Djukanovic, M., Grbic, M., Blum, C., & Raidl, G. (2021). Solving the longest common subsequence problem concerning non-uniform distributions of letters in input strings. *Mathematics*, *9*(13). 1515. https://doi.org/10.3390/math9131515

Rajalakshmi, B., Jayashree, & Dharan, A. (2024). Fake product identification system. In *2024 5th International Conference for Emerging Technology (INCET 2024)* (pp. 1–6). https://doi.org/10.1109/INCET61516.2024.10593456

Sharma, P., & Balakrishna, G. (2011). PrefixSpan: Mining sequential patterns by prefix-projected pattern. *International Journal of Data Mining and Knowledge Management Process, 2*(4), 111–122.

Xiao, L., Huang, G., Pedrycz, W., Pamucar, D., Martínez, L., & Zhang, G. (2022). A q-rung orthopair fuzzy decision-making model with new score function and best-worst method for manufacturer selection. *Information Sciences, 608*, 153–177. https://doi.org/10.1016/j.ins.2022.06.061

Zwitter, A., & Hazenberg, J. (2020). Decentralized network governance: Blockchain technology and the future of regulation. *Frontiers in Blockchain, 3*, 12. https://doi.org/10.3389/fbloc.2020.00012