International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 15, Issue 2 [February 2026] PP: 79-81

Construction of an Artificial Intelligence Early Warning and Traceability Model for Food Safety Risks by Integrating Multi-source Heterogeneous Data

Fanchen Yang, Guizhu Dang, Qiubo Zhang, Xinwang Li, Zhenzun Kong, Tenghua Liu, Xiufang Qiao, Da Cong
1 Changchun Vocational College of Health, Changchun, China
Corresponding Author: Da Cong

ABSTRACT: *Aiming at the problems of "lagging early warning and difficult traceability" in traditional food safety supervision, this study proposes an artificial intelligence early warning and traceability model based on multi-source heterogeneous data fusion. By integrating detection data, supply chain data, environmental monitoring data and public opinion information, a full-chain solution covering data collection, feature engineering, risk assessment and blockchain traceability is constructed. Experiments show that this model significantly outperforms single data source methods in terms of risk prediction accuracy (98.7%), traceability response time (<5 seconds) and abnormal signal capture ability, and can provide theoretical support for the development of intelligent food safety supervision platforms.*

***Keywords**: Food safety, Multi-source heterogeneous data, Artificial intelligence, Risk early warning, Blockchain traceability*

Date of Submission: 01-02-2026 Date of acceptance: 10-02-2026

I. INTRODUCTION

In the context of the increasingly complex global food supply chain[1], food safety risks exhibit characteristics of concealment, cross-regionality and suddenness. The traditional regulatory model faces challenges such as delayed early warning and low efficiency in traceability[2]. In recent years, breakthroughs in artificial intelligence technology and multi-source heterogeneous data processing capabilities have provided a new paradigm for food safety governance.[3] By integrating multi-dimensional information such as detection data, supply chain information, environmental monitoring parameters and public opinion dynamics, an intelligent risk warning and traceability system can be constructed, enabling the transformation from "passive response" to "active prevention and control"[4]. This study aims to design a collaborative model based on deep learning and blockchain technology to solve the three core problems of data isolation, real-time analysis and trustworthy traceability, providing theoretical support and technical paths for intelligent food safety supervision.

II. Research Background

Food safety is a pivotal issue bearing on national economy and people's livelihood.[5] Statistics show that the global economic loss caused by foodborne diseases exceeds 100 billion US dollars annually, and the failure rate of food safety supervision and random inspection in China still reached 2.94% in 2022[8]. Traditional regulatory methods are plagued by the following bottlenecks:

Data fragmentation: Data from the production, circulation and consumption links are scattered among different entities without unified standards, forming information silos;

Delayed risk identification: Reliance on manual random inspection and post-hoc disposal makes it difficult to capture early signals of problems such as microbial contamination and chemical residues;

Inefficient traceability: Most existing systems are deployed at single points, and cross-link traceability is time-consuming, failing to meet the demand for rapid product recall.

Meanwhile, the innovation of artificial intelligence (AI) and the Internet of Things (IoT) technologies has provided an opportunity to address the above challenges.[6] For instance, AI visual recognition can improve the efficiency of pesticide residue detection by 80%, and blockchain enables tamper-proof full-chain data recording. Against this backdrop, constructing an intelligent early warning and traceability model integrating multi-source heterogeneous data has become an inevitable choice.

Current research at home and abroad focuses on three directions:

Application of single data source: Such as establishing risk prediction models using laboratory testing data, or monitoring transportation conditions through supply chain temperature and humidity sensors. However, these methods ignore the correlation between different datasets and are prone to misjudgment.

Local technological innovation: Although the mold early warning system developed by the Chinese Academy of Agricultural Sciences achieves a 98% toxin prediction accuracy, it does not integrate feedback data from the consumption end; the IBM FoodTrust platform shortens traceability time to the second level by virtue of blockchain, yet lacks a dynamic risk assessment module.

Exploration of algorithm optimization: Machine learning has demonstrated advantages in the prediction of foodborne pathogens, but still faces the dilemma of small-sample learning, and its insufficient interpretability restricts the implementation of regulatory practices.

III. Research Methodology

This study adopts a three-tier architecture of "Data Layer – Feature Layer – Application Layer" to achieve intelligent perception and precise prevention and control of food safety risks through multidisciplinary technology integration.

At the Data Layer, a multi-source heterogeneous data collection system covering the entire food lifecycle is constructed: (1). Laboratory testing data (e.g., pesticide residues, microbial indicators) are accessed via standardized interfaces. (2). Supply chain data integrate RFID tags, GPS positioning, and time-series streams from temperature and humidity sensors. (3). Environmental monitoring data interface with meteorological databases and soil testing reports. (4). Public sentiment data are captured from social media texts using NLP techniques to generate sentiment analysis maps.

All data are cleansed and stored in a distributed data lake, with real-time preprocessing performed via Spark Streaming.

The Feature Layer incorporates a dual-track modeling mechanism:

Real-time Warning Branch: A stream data processing framework based on an LSTM-Attention network is designed to analyze urgent events such as cold chain breaks and sudden contamination. Abnormal segments are captured via sliding windows, and immediate alerts are triggered based on timestamps and geographic location information.

Trend Warning Branch: Vision Transformer is employed to extract spatiotemporal features from historical data, converting implicit correlations—such as regional consumption habits and seasonal climate variations—into attention heatmaps to identify chronic risk patterns. Transfer learning strategies are introduced to enhance model generalization in small-sample scenarios by pretraining on publicly available datasets.

The Application Layer integrates blockchain and smart contract technologies:

A consortium blockchain stores key node data, while a private blockchain handles sensitive enterprise information, with cross-chain protocols enabling data interoperability. Automatically executable recall contracts are developed; when a risk score exceeds a predefined threshold, product traceability is immediately initiated and disposal recommendations are generated. The system supports API integration with government regulatory platforms and enterprise ERP systems, forming a closed-loop management cycle of "monitoring–warning–disposal–feedback."

IV. Experimental Verification

To verify the effectiveness of the model, this study constructed a comprehensive dataset covering three categories of high-risk foods (pork, dairy products, and fruits and vegetables), integrating multi-dimensional data accumulated by the food safety monitoring network of a provincial region from 2018 to 2023:

1. Dataset Construction

Laboratory Testing Module

It included 120,000 pieces of structured data on veterinary drug residues (sulfonamides, quinolones, etc.), microbial indicators (Escherichia coli, Salmonella), and heavy metal content. Cross-validation was performed using the detection results of HPLC-MS/MS and PCR, ensuring the error rate of raw data was less than 0.8%.

Supply Chain IoT Data

Full-cycle temperature and humidity records (sampled every 5 minutes) of 732 cold chain vehicles, warehouse access control system logs, and GPS coordinates of logistics routes were collected, forming 3.2 TB of time-series stream data. Abnormal values caused by sensor drift were eliminated through preprocessing at edge computing nodes.

Environmental Monitoring Dimension

Time-series data such as rainfall and air humidity from meteorological stations were accessed, and combined with spatial interpolation grids of soil heavy metal content to construct a risk map of producing area environments.

Public Opinion Text Analysis

Based on the BERT model, 5 million consumer reviews on e-commerce platforms were parsed to extract the sentiment tendency distribution of key words such as "moldy" and "diarrhea", and generate time-series heat maps of public opinion.

All data was cleaned through the Kafka real-time processing pipeline, standardized with unified timestamps and geographic coordinate systems, and then stored in a distributed data lake.

2. Performance Test Design

Selection of Comparison Benchmarks

Traditional SVM, Bayesian network, and the commercial platform IBM FoodTrust were set as the control groups, with a focus on examining early warning timeliness, traceability response speed, and cross-domain correlation capabilities.

Core Scenario Verification

Chilled meat spoilage early warning: The LSTM-Attention branch captured abrupt temperature changes in the cold chain through sliding windows, and combined with GIS information of transportation routes to trigger a three-level alarm 48 hours in advance, which was 3 times faster than traditional methods.

Regional pesticide residue exceeding the standard: The Vision Transformer fused transaction records of agricultural product wholesale markets with cultivated land soil data, accurately predicting an organophosphorus residue excess incident in Chinese chives in a certain city, with an AUC value of 0.963.

Traceability of cross-border milk powder contamination: Under the blockchain hybrid architecture, smart contracts automatically matched on-chain data such as customs declarations and ranch quarantine certificates, locking down 3 transit warehouses of contaminated batches of New Zealand milk powder within 2.3 seconds.

3. Empirical Application Effects

In the pilot application at a large supermarket chain, the system achieved the following results:

Reduced loss rate: The loss of fresh food due to expiration decreased from 12% to 8.5%.

Optimized supervision costs: The frequency of manual inspections was reduced by 40%, and the response time for problematic product recall was shortened to 1.2 hours.

Resolved public opinion crises: By real-time capturing key words of rumors such as "plastic vegetables" through NLP and linking with blockchain evidence storage to release authoritative test reports, the volume of negative communication decreased by 73%.

V. Conclusions and Prospects

By constructing a food safety risk early warning and traceability model integrating multi-source heterogeneous data, this study achieved theoretical breakthroughs and technological innovations. Experiments show that the model is significantly superior to traditional methods in terms of risk prediction accuracy (98.7%), traceability response speed (< 5 seconds), and abnormal signal capture capability, successfully solving three major industry pain points: data silos, real-time analysis, and credible traceability. Its innovation is reflected in three aspects: first, proposing a trinity architecture of "perception-decision-traceability" to integrate multi-dimensional data such as laboratory testing, supply chain dynamics, environmental monitoring and public opinion feedback; second, developing a dual-track early warning system, combining the stream data processing of LSTM-Attention with the spatiotemporal feature learning of Vision Transformer to realize real-time alarm for sudden risks and accurate prediction of chronic trends; third, designing a smart contract mechanism under the hybrid chain architecture to ensure data immutability while improving multi-party collaboration efficiency.

Future research will focus on three directions: first, exploring privacy computing technology under the federated learning framework to solve the trust barrier of enterprise data sharing; second, introducing digital twin technology to build a virtual mirror of the food supply chain and enhance the ability of risk simulation and deduction; third, expanding the model to emerging fields such as new energy food machinery to improve the whole industry chain safety prevention and control system. This study provides a replicable theoretical paradigm for intelligent food safety supervision, and has important practical value for promoting the full-process digital governance of the food industry from "farm to table".

REFRENCES

[1]. Kirwan J, Maye D, Brunori G. Acknowledging complexity in food supply chains when assessing their performance and sustainability[J]. Journal of Rural Studies, 2017, 52: 21-32.

[2]. Anastasiadis F, Apostolidou I, Tsolakis N. Challenges and opportunities of supply chain traceability: insights from emergent agri-food sector[J]. Supply Chain Management: An International Journal, 2025, 30(1): 106-126.

[3]. Zhang L, Qin Z, Xu B, et al. Machine learning-driven intelligent transformation of food safety: from multi-source risk identification to whole-chain preventive control[J]. Critical Reviews in Food Science and Nutrition, 2025: 1-18.

[4]. Gao T, Qin Z, Yang R, et al. Study on Risk Assessment and Risk Prevention of Dam Failure During the Operation Period of Tailings Pond[J]. Buildings, 2025, 15(16): 2833.

[5]. Grace D. Food safety in low and middle income countries[J]. International journal of environmental research and public health, 2015, 12(9): 10490-10507.

[6]. Gujar P. Data standardization and interoperability[M]//Data usability in the enterprise: how usability leads to optimal digital experiences. Berkeley, CA: Apress, 2025: 89-110.