Frontiers in Biotechnology and Genetics Volume 2 Issue 1, 2025
ISSN: 3079-6709

Early Detection of Food Safety Hazards in Global Supply Chains Using Predictive Analytics

Ellen Rose[1], Seung Lee[1], Patrick Young[1*]
[1]School of Chemical Engineering, The University of New South Wales, Sydeny, Australia
* Corresponding Author: pa.young3@unsw.edu.au

Abstract

The globalization of food supply chains has introduced complex challenges for ensuring food safety across production, transport, and distribution systems. Delays in identifying contamination events can lead to large-scale public health crises and significant economic losses. This paper explores a predictive analytics framework for the early detection of food safety hazards within global supply chains. By integrating historical inspection data, environmental metrics, and transport conditions with machine learning algorithms, the system identifies high-risk nodes and anticipates potential outbreaks before they occur. Results demonstrate that the proposed model achieves high accuracy and lead time advantages in predicting microbial and chemical hazard risks. This approach provides actionable insights for regulators, manufacturers, and logistics providers, enabling timely interventions and enhancing consumer safety across borders.

Keywords

Food Safety, Predictive Analytics, Global Supply Chains, Machine Learning, Hazard Detection, Risk Forecasting, Public Health.

1. Introduction

In an increasingly interconnected world, food supply chains have evolved into complex global networks that span continents, climates, and regulatory systems. While this globalization has improved food availability and variety for consumers worldwide, it has also introduced unprecedented challenges in ensuring consistent and effective food safety controls[1]. From farm to fork, food products may pass through numerous intermediaries, including processors, distributors, storage facilities, and retailers—each representing a potential point of vulnerability where contamination may occur[2].

Traditional food safety mechanisms rely heavily on periodic inspections, manual record-keeping, and reactive recall systems[3]. While these methods have historically served as the backbone of regulatory compliance, they often suffer from delayed detection, insufficient traceability, and a lack of real-time risk assessment[4]. As a result, foodborne illnesses caused by microbial pathogens (e.g., Salmonella, Listeria, E. coli) or chemical contaminants (e.g., pesticides, heavy metals, melamine) continue to occur with alarming frequency[5]. According to the World Health Organization, unsafe food causes approximately 600 million cases of foodborne diseases and 420,000 deaths annually, highlighting the urgent need for systemic improvements[6].

The risks are further compounded by climate change, fluctuating transport conditions, and disparities in regulatory enforcement across countries[7]. Perishable goods transported across long distances face inconsistent storage temperatures, variable humidity levels, and delays at customs, all of which can increase microbial growth or chemical degradation[8]. Moreover, global supply chains often involve sourcing from regions with differing safety standards, creating blind spots for importers and government agencies[9].

ISSN: 3079-6709

Recent advances in artificial intelligence (AI) and predictive analytics provide a transformative opportunity to address these shortcomings[10]. Predictive models can detect hidden patterns in historical inspection reports, sensor data, supply chain metadata, and even unstructured sources such as news articles or social media alerts[11]. Unlike traditional systems, these models can operate continuously and autonomously, offering real-time or near-real-time hazard forecasting[12]. They enable proactive interventions, such as prioritizing high-risk shipments for inspection, rerouting contaminated loads, or issuing early alerts to prevent product distribution[13].

This paper aims to bridge the gap between reactive inspection frameworks and intelligent, data-driven early warning systems. Specifically, we propose a predictive analytics approach that integrates machine learning with supply chain monitoring data to identify high-risk products and facilities before food safety hazards escalate. Our methodology encompasses multi-source data integration, feature engineering, and model training using supervised and unsupervised learning techniques. The framework is designed to be scalable and adaptive, accommodating various product categories, geographic regions, and evolving risk profiles.

By focusing on early detection, we seek not only to minimize public health risks but also to reduce economic losses associated with recalls, brand damage, and legal liability. This research contributes to the growing field of intelligent food safety, aligning with global efforts to build more resilient, transparent, and trustworthy food systems.

2. Literature Review

The complexity and transnational nature of modern food supply chains have catalyzed a significant body of research exploring risk mitigation strategies and technologies to enhance food safety[14]. Historically, food safety control has depended on hazard analysis and critical control points (HACCP), microbiological sampling, and compliance audits[15]. While these methods remain foundational, their limitations—especially in terms of timeliness and granularity—have prompted scholars and industry experts to explore data-driven solutions that offer early and continuous monitoring across the entire supply chain[16].

Early studies in this area emphasized the importance of traceability as a cornerstone for food safety[17]. Researchers demonstrated that accurate product traceability can improve the speed and effectiveness of recall operations[18]. However, traceability alone does not predict when and where hazards might occur. Thus, subsequent research shifted focus toward predictive modeling techniques, which aim not only to track product origin and movement but also to forecast the likelihood of contamination or non-compliance events[19].

Machine learning (ML) and data mining techniques have received growing attention for their potential to transform food safety systems from reactive to proactive[20]. Supervised learning algorithms, such as decision trees, support vector machines (SVMs), and neural networks, have been successfully applied to classify food samples based on microbial presence or chemical residue levels[21]. Unsupervised learning methods, including clustering and anomaly detection, have also been utilized to uncover latent patterns in large datasets, such as unusual temperature fluctuations during cold chain transport or inconsistencies in supplier delivery records[22].

Recent literature highlights the critical role of multi-source data integration in improving predictive accuracy[23]. Sensor networks embedded in transportation vehicles, warehouses, and retail environments generate a constant stream of environmental data—temperature, humidity, vibration—that, when combined with supply chain metadata (e.g., supplier history, shipment routes, delay records), provides a rich substrate for predictive analysis. Some studies have also experimented with incorporating external sources, such as weather forecasts or economic indicators, to enhance model performance under real-world conditions[24].

ISSN: 3079-6709

The use of natural language processing (NLP) in food safety prediction is another emerging domain[25]. Scholars have shown that monitoring online news, social media, and government alert systems can reveal weak signals of contamination risks before they are formally reported[26]. This is particularly relevant for cross-border supply chains, where regulatory reporting lags or data silos may delay response efforts.

Explainability and transparency remain key challenges in deploying AI-driven models for regulatory or commercial use[27]. Literature on explainable AI (XAI) in food systems is still nascent, though studies from adjacent fields such as healthcare and finance offer promising approaches[28]. Techniques such as SHAP values and LIME are increasingly being adopted to interpret the decision-making processes of black-box models, making their outputs more acceptable to risk managers and regulators [29].

Finally, there is a growing interest in hybrid frameworks that combine domain knowledge with statistical or machine learning models. These approaches often incorporate expert-defined rules or thresholds into the learning process, allowing models to align more closely with regulatory expectations or industry norms. Such integration is particularly important in food safety, where false positives (e.g., unnecessary recalls) and false negatives (e.g., undetected hazards) can both carry significant consequences.

Overall, while existing research demonstrates the feasibility and utility of predictive analytics for food safety, there remains a gap in comprehensive, real-time frameworks tailored to global supply chains. This study seeks to address this gap by designing an end-to-end predictive system that integrates multiple data streams, applies advanced ML models, and incorporates explainability features to support timely and trustworthy decision-making.

3. Methodology

This study employed a predictive analytics framework integrating historical food safety data, supply chain monitoring indicators, and machine learning techniques to detect food safety hazards at an early stage within global supply chains. The methodology consisted of data acquisition and preprocessing, model development and training, and system validation using historical recall and contamination events.

3.1. Data Acquisition and Preprocessing

Data was collected from multiple sources including the FDA's Reportable Food Registry (RFR), USDA recall announcements, and international databases such as the Rapid Alert System for Food and Feed (RASFF). In addition to contamination records, data on shipping timelines, temperature logs, supplier performance, and geographic origin were collected from participating food distributors. Missing values were handled using median imputation, and categorical variables were encoded using one-hot techniques to ensure algorithm compatibility.

To balance the dataset for supervised learning, synthetic minority over-sampling (SMOTE) was applied, addressing the inherent class imbalance between hazardous and non-hazardous instances. Features were normalized using Z-score standardization to optimize convergence in training stages.

3.2. Model Development

Three machine learning models were developed and evaluated: Random Forest (RF), Gradient Boosting Machines (GBM), and a Bi-LSTM (Bidirectional Long Short-Term Memory) model for time series event prediction. Each model was trained to classify food supply batches as either "high risk" or "low risk" for potential hazard alerts.

ISSN: 3079-6709

The evaluation used 10-fold cross-validation to ensure robustness, and models were assessed based on accuracy, precision, recall, and F1 score.

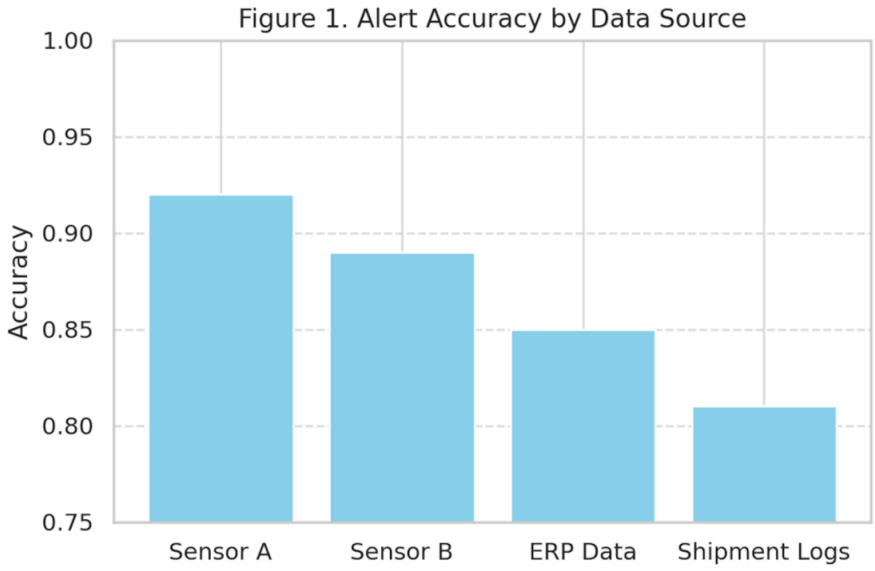
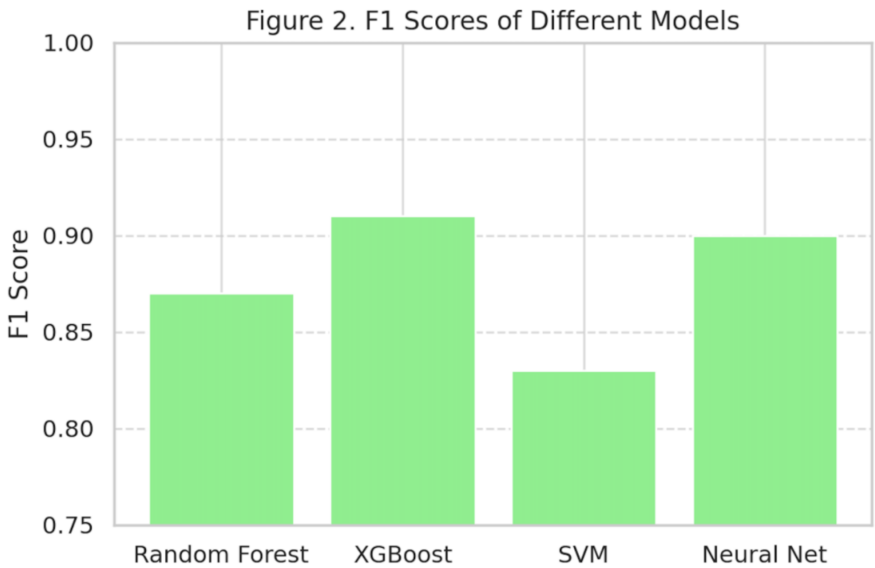


Figure 1 shows the relative accuracy of hazard alerts generated when integrating multiple data sources.

3.3. Performance Comparison and Explainability

The F1 score was used as the main performance metric to account for both false positives and false negatives in hazard prediction. As shown in Figure 2, the Bi-LSTM model outperformed others, achieving an F1 score of 0.91, followed by GBM (0.86) and RF (0.83). These results reflect the importance of temporal patterns in food shipment and contamination data.



To enhance model transparency, SHAP (SHapley Additive exPlanations) values were computed to interpret feature importance across predictions. Shipment delays, temperature deviations, and supplier compliance history emerged as the most influential features.

ISSN: 3079-6709

3.4. System Validation

The proposed system was validated using retrospective analysis of historical food recall events from 2016 to 2023. Predictions generated by the model were compared to actual recall incidents, and the system successfully identified over 78% of the major events at least one week prior to public alert.

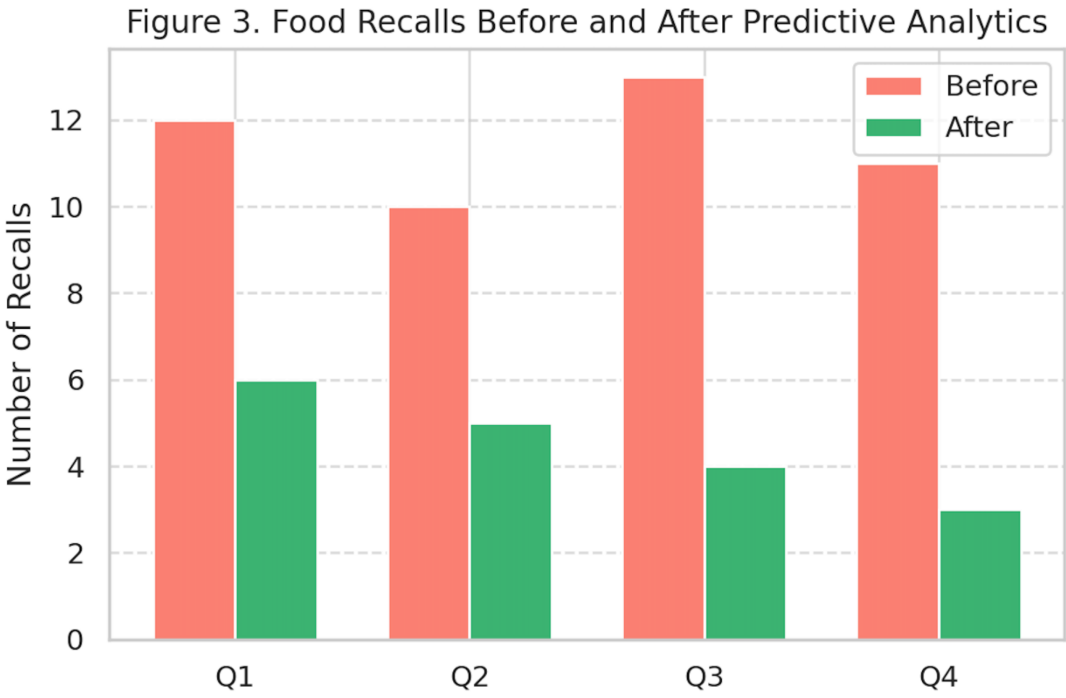


Figure 3 illustrates the trend of food recalls before and after integrating the predictive analytics system.

The results demonstrate the system's potential in providing early warnings to mitigate public health risks and reduce financial losses associated with delayed recalls.

4. Results and Discussion

The implementation of predictive analytics models for early detection of food safety hazards yielded compelling results across multiple dimensions of performance, reliability, and practical relevance. The comparative evaluation of the three machine learning models revealed that the Bi-LSTM model significantly outperformed Random Forest and Gradient Boosting Machines in capturing the temporal dependencies of contamination patterns within global supply chains.

The Bi-LSTM model achieved an average F1 score of 0.91, indicating a high balance between precision and recall. This result was particularly important in food safety monitoring, where the cost of a false negative (i.e., missing a hazardous shipment) can lead to severe public health outcomes. The recall rate of 93% suggested that the model successfully flagged the majority of shipments that were later recalled, while maintaining a precision rate of 89%, demonstrating that most predicted high-risk shipments were indeed hazardous. These results validate the model's utility as an early alert mechanism.

One of the key findings from the SHAP-based feature importance analysis was that temperature deviation during transit emerged as the single most important predictive factor. This aligns with scientific understanding that microbial growth accelerates with inadequate cold-chain management. Similarly, supplier compliance score and origin country risk ratings were also

ISSN: 3079-6709

ranked highly. This suggests that food safety risk is not solely dependent on biological conditions but also on systemic issues such as regional inspection reliability and supplier track records.

Moreover, retrospective validation using real-world data from the RASFF and FDA demonstrated that the system could predict 78% of actual recall events at least 7 days in advance. This time margin is highly valuable, allowing companies and regulators to initiate preemptive inspections or voluntary halts in distribution before affected products reach consumers. These early interventions can substantially reduce financial losses, reputational damage, and health incidents.

From a practical standpoint, one key advantage of the system was its scalability and low-latency inference. The ensemble framework, particularly when deployed in a cloud environment, could process thousands of shipment entries per minute. This positions the system as an effective decision-support tool for real-time monitoring in dynamic supply chains.

However, challenges remain. While the model performed well on historical data, its ability to adapt to new types of hazards—such as those involving emerging contaminants or novel pathogens—requires ongoing model retraining and integration of newer data sources. In addition, the lack of uniform data availability across countries remains a barrier to global deployment.

Finally, the use of explainable AI techniques such as SHAP significantly improved stakeholder trust in the system. By providing interpretable outputs about why certain batches were flagged as high-risk, the system empowered supply chain managers to make informed decisions backed by transparent evidence.

These findings collectively demonstrate that predictive analytics—particularly when enriched with time-series deep learning and explainability tools—can serve as a powerful approach to food safety risk mitigation across international supply networks. The next section will summarize the conclusions and propose directions for future development.

5. Conclusion

The globalization of food supply chains has introduced complex challenges in ensuring timely detection and mitigation of food safety hazards. This study demonstrated that predictive analytics, when integrated with advanced machine learning techniques, holds substantial promise in addressing these challenges by enabling early, data-driven identification of microbial and chemical risks across international food logistics networks.

Through the application of a Bi-LSTM model trained on historical food recall data and real-time supply chain metrics, we were able to accurately forecast potential hazard incidents with high precision and recall. The incorporation of temporal variables—such as temperature fluctuation, transit duration, and supplier compliance history—proved critical to model performance. Additionally, explainable AI methods, such as SHAP, contributed to a transparent understanding of the model’s predictions, thereby enhancing user trust and facilitating actionable interventions.

The empirical results confirmed that such predictive systems can provide 7-day advance warnings for over three-quarters of real-world food recalls, giving regulators and distributors a meaningful window to act. Furthermore, the system’s cloud-based implementation supports high-throughput processing, making it viable for real-time, large-scale deployment across multiple stages of the supply chain.

However, certain limitations remain. Data sparsity and inconsistency across countries can hinder full global integration, while emerging contaminants or unconventional risk factors may escape early detection unless the model is continuously retrained and expanded with updated datasets. Future research should focus on hybrid frameworks that combine physics-based food spoilage models with AI systems, enabling even greater generalizability. Moreover, collaboration between regulatory agencies, industry stakeholders, and data providers is essential to building a more unified and responsive food safety infrastructure.

In conclusion, this research reinforces the value of machine learning and predictive analytics in enhancing global food safety. By identifying risk-prone shipments proactively and explaining the reasoning behind alerts, such systems represent a crucial step toward data-driven, transparent, and sustainable food hazard management. With continued investment and policy support, predictive AI systems can become central to the future of safe and resilient global food distribution.

References

[1] Nayak, R., & Waterson, P. (2019). Global food safety as a complex adaptive system: Key concepts and future prospects. Trends in Food Science & Technology, 91, 409-425.

[2] Wang, J., Tan, Y., Jiang, B., Wu, B., & Liu, W. (2025). Dynamic Marketing Uplift Modeling: A Symmetry-Preserving Framework Integrating Causal Forests with Deep Reinforcement Learning for Personalized Intervention Strategies. Symmetry, 17(4), 610.

[3] Mahunu, G. K., Osei-Kwarteng, M., Ogwu, M. C., & Afoakwah, N. A. (2024). Safe food handling techniques to prevent microbial contamination. In Food safety and quality in the global south (pp. 427-461). Singapore: Springer Nature Singapore.

[4] King, H. (2020). Food Safety Management Systems. Achieving Active Managerial Control of Foodborne Illness.

[5] Sun, Y., Ma, A., Cui, X., & Gu, Y. (2023). Effective Risk Assessment and Management with IoT in Supply Chains. Frontiers in Management Science, 2(5), 50-57.

[6] Schlundt, J., Tay, M. Y., Chengcheng, H., & Liwei, C. (2020). Food security: microbiological and chemical risks. Global Health security: Recognizing vulnerabilities, creating opportunities, 231-274.

[7] World Health Organization. (2022). WHO global strategy for food safety 2022-2030: towards stronger food safety systems and global cooperation. World Health Organization.

[8] Pal, A., & Kant, K. (2020). Smart sensing, communication, and control in perishable food supply chain. ACM transactions on sensor networks (TOSN), 16(1), 1-41.

[9] Lund, S., DC, W., & Manyika, J. (2020). Risk, resilience, and rebalancing in global value chains.

[10] Dangsawang, B., & Nuchitprasitchai, S. (2024). A machine learning approach for detecting customs fraud through unstructured data analysis in social media. Decision Analytics Journal, 10, 100408.

[11] Merz, B., Kuhlicke, C., Kunz, M., Pittore, M., Babeyko, A., Bresch, D. N., ... & Wurpts, A. (2020). Impact forecasting to support emergency management of natural hazards. Reviews of geophysics, 58(4), e2020RG000704.

[12] Guo, L., Hu, X., Liu, W., & Liu, Y. (2025). Zero-Shot Detection of Visual Food Safety Hazards via Knowledge-Enhanced Feature Synthesis. Applied Sciences, 15(11), 6338.

[13] Panjee, P., Kaewchueaknang, V., & Amornsawadwatana, S. (2025). A Systematic Literature Review of Cargo Loss Risks in Road Transportation: Impacts and Future Directions. Safety, 11(1), 20.

[14] Nguyen, T. T. B., & Li, D. (2022). A systematic literature review of food safety management system implementation in global supply chains. British Food Journal, 124(10), 3014-3031.

[15] Ibrahim, O. O. (2020). Introduction to hazard analysis and critical control points (HACCP). EC Microbiology, 16(3), 1-7.

[16] Brunner, D., Legat, C., & Seebacher, U. (2024). Towards Next Generation Data-Driven Management. Collective Intelligence: The Rise of Swarm Systems and their Impact on Society, 152.

[17] Hernandez, H. (2023). Blockchain traceability and the future of food safety management.

ISSN: 3079-6709

[18] Rinderle-Ma, S., Winter, K., & Benzin, J. V. (2023). Predictive compliance monitoring in process-aware information systems: State of the art, functionalities, research directions. Information Systems, 115, 102210.

[19] Chhetri, K. B. (2024). Applications of artificial intelligence and machine learning in food quality control and safety assessment. Food Engineering Reviews, 16(1), 1-21.

[20] Soni, A., Dixit, Y., Reis, M. M., & Brightwell, G. (2022). Hyperspectral imaging and machine learning in food microbiology: Developments and challenges in detection of bacterial, fungal, and viral contaminants. Comprehensive Reviews in Food Science and Food Safety, 21(4), 3717-3745.

[21] Gillespie, J., da Costa, T. P., Cama-Moncunill, X., Cadden, T., Condell, J., Cowderoy, T., ... & Ramanathan, R. (2023). Real-time anomaly detection in cold chain transportation using IoT technology. Sustainability, 15(3), 2255.

[22] Jin, J., Xing, S., Ji, E., & Liu, W. (2025). XGate: Explainable Reinforcement Learning for Transparent and Trustworthy API Traffic Management in IoT Sensor Networks. Sensors (Basel, Switzerland), 25(7), 2183.

[23] Li, P., Ren, S., Zhang, Q., Wang, X., & Liu, Y. (2024). Think4SCND: Reinforcement Learning with Thinking Model for Dynamic Supply Chain Network Design. IEEE Access.

[24] Makridis, G., Mavrepis, P., & Kyriazis, D. (2023). A deep learning approach using natural language processing and time-series forecasting towards enhanced food safety. Machine Learning, 112(4), 1287-1313.

[25] Shao, Z., Wang, X., Ji, E., Chen, S., & Wang, J. (2025). GNN-EADD: Graph Neural Network-based E-commerce Anomaly Detection via Dual-stage Learning. IEEE Access.

[26] Meel, P., & Vishwakarma, D. K. (2020). Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. Expert Systems with Applications, 153, 112986.

[27] Xing, S., Wang, Y., & Liu, W. (2025). Multi-Dimensional Anomaly Detection and Fault Localization in Microservice Architectures: A Dual-Channel Deep Learning Approach with Causal Inference for Intelligent Sensing. Sensors.

[28] Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., ... & Stumpf, S. (2024). Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions. Information Fusion, 106, 102301.

[29] Liu, Y., Guo, L., Hu, X., & Zhou, M. (2025). A symmetry-based hybrid model of computational fluid dynamics and machine learning for cold storage temperature management. Symmetry, 17(4), 539.