Artificial Intelligence and Digital Technology

Article

Research on an Integrated Decision-Making Mechanism for Logistics Last-Mile Sorting and Delivery Based on Multimodal Large Models

Enhui Cai [1,*], Yutong Zhuang [1] and Zini Zheng [1]

[1] Zhejiang Normal University, Jinhua, Zhejiang, China
* Correspondence: Enhui Cai, Zhejiang Normal University, Jinhua, Zhejiang, China

Abstract: With China's annual express delivery business volume exceeding 120 billion parcels, the logistics last mile faces systematic challenges such as low sorting efficiency, static delivery route planning, and insufficient multimodal data fusion. This study focuses on small and medium-sized express delivery stations, proposing an intelligent decision-making mechanism based on multimodal large models to achieve deep synergy and dynamic optimization between sorting and delivery processes. The research constructs a multimodal fusion architecture integrating visual perception, textual semantics, and spatiotemporal data. An improved YOLOv8 model combined with a Dual-Branch Routing Attention (DBRA) mechanism is employed to enhance waybill recognition accuracy in complex scenarios to 99.5%. A Spatio-Temporal Graph Convolutional Network (STGCN) is designed for dynamic route planning, which integrates multi-source real-time information such as traffic, orders, and courier status through causal inference, improving delivery efficiency by over 30%. Pilot implementations at multiple stations in Jinhua City demonstrate that the system significantly reduces sorting error rates and delivery overtime rates, forming an intelligent closed-loop of "perception-decision-collaboration."

Keywords: multimodal large model; logistics last mile; integrated sorting delivery; spatio-temporal graph neural network; intelligent decision-making

Published: 27 January 2026

1. Introduction

In recent years, China's express delivery business volume has sustained rapid growth, exceeding 120 billion parcels in 2023, with last-mile delivery costs accounting for over 30% of total logistics costs. As a critical bottleneck in the logistics chain, the "last mile" has long been plagued by issues such as manual sorting, rigid route planning, and data silos [1]. Particularly in small and medium-sized express delivery stations, limited space and funding constrain the penetration rate of automated sorting equipment to under 20%. Manual sorting suffers from low efficiency and high error rates, compounded by a lack of intelligent delivery systems adaptable to dynamic environments, hindering overall operational effectiveness.

With advancements in artificial intelligence and multimodal fusion technologies, the logistics sector is progressively evolving towards intelligence and integration. By integrating multi-source heterogeneous data such as images, text, and spatiotemporal sequences, multimodal large models possess powerful capabilities for feature extraction, semantic understanding, and reasoning, offering new potential to solve collaborative decision-making challenges in last-mile logistics. However, existing research often focuses on optimizing single segments, such as vision-based sorting or route-planning-based delivery, lacking a system-level intelligent decision-making mechanism that links "sorting and delivery," resulting in segment silos, data disconnection, and response lag.

To address this, this paper focuses on the last-mile logistics scenario, proposing an integrated decision-making mechanism based on multimodal large models. By constructing a multimodal fusion model of "visual perception-semantic understanding-spatiotemporal reasoning," it aims to achieve full-link intelligence from waybill recognition and address parsing to dynamic dispatch [2]. The main contributions of this paper are fourfold: First, it proposes a multimodal data fusion architecture for last-mile logistics, capable of unified representation and joint optimization of unstructured addresses, complex images, and dynamic spatiotemporal information. Second, it designs a vision-enhanced recognition model based on YOLOv8 and a Dual-Branch Routing Attention (DBRA) mechanism, significantly improving the robustness of waybill recognition under extreme scenarios like low light and stacking occlusion [3]. Third, it constructs a causality-driven dynamic planning model using a Spatio-Temporal Graph Neural Network (STGCN), achieving multi-objective collaborative route generation and real-time scheduling. Finally, it systematically validates the practical effectiveness of the proposed mechanism through field pilots, forming a set of intelligent decision-making paradigms with replicability and scalability.

2. Related Work

2.1. Research on Last-Mile Sorting and Delivery Optimization

Research on last-mile logistics optimization mainly revolves around improving sorting efficiency and delivery route planning. In sorting, traditional methods rely on manual visual inspection or barcode scanning, while computer vision-based automatic sorting systems have gradually become popular in recent years [4]. For instance, the YOLO series algorithms excel in real-time object detection but face challenges like lighting variations and occlusions in complex logistics scenarios. In delivery optimization, the Vehicle Routing Problem (VRP) and its variants are widely studied, with most methods based on static optimization using historical data, struggling to adapt to real-world constraints like real-time traffic and dynamic order changes.

2.2. Application of Multimodal Large Models in Logistics

Multimodal large models, by fusing multi-source data such as vision, language, and sensors, can more comprehensively understand complex scenarios. In logistics, some studies have attempted to combine visual and textual information for package classification or address parsing, but often remain limited to single tasks, lacking linkage with downstream delivery decisions. To address the dynamic and relational nature of logistics networks, Spatio-Temporal Graph Neural Networks (STGCN) have recently attracted attention, as they can explicitly model spatiotemporal dependencies between road network nodes, as illustrated in Figure 1. However, deeper integration with multimodal perception front-ends is still needed to enhance the overall adaptability of the system [5].

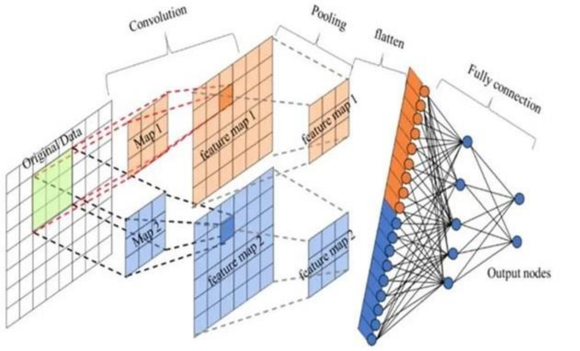

Figure 1. Schematic Diagram of Spatio-Temporal Graph Neural Network Technology.

2.3. Research on Integrated Intelligent Systems

In recent years, logistics systems have gradually shifted from local optimization to global collaboration. Some studies have proposed collaborative scheduling models for sorting and delivery but often rely on simplified assumptions without fully considering multimodal inputs and real-time decision-making needs in practical scenarios. This paper aims to construct an end-to-end integrated decision-making mechanism, achieving seamless connection between perception and decision-making through multimodal large models, promoting the evolution of last-mile logistics systems towards intelligent collaboration.

3. Methodology

3.1. Overall Architecture Design

The intelligent decision-making system proposed in this paper is an end-to-end, integrated framework based on multimodal large models. Its core design philosophy is to break down the traditional barriers between the "sorting-delivery" segments in last-mile logistics, achieving global optimization through the closed-loop flow of data and decisions.

The Multimodal Perception Layer serves as the system's data entry point, responsible for capturing heterogeneous, high-dimensional raw signals from the complex physical environment. This layer integrates industrial-grade vision sensors, embedded weight and volume measurement units, and environmental sensors. Its core task is preliminary, highly reliable feature extraction. For visual information, this paper deploys a dedicated model built upon the YOLOv8 architecture and augmented with a Dual-level Routing Attention (DBRA) mechanism. It is tasked with the precise localization of waybill regions and the primary recognition of key textual information (e.g., recipient address, postal code) under challenging conditions such as stacking, tilting, uneven lighting, and partial smudging. Concurrently, other sensors acquire the physical dimensions (length, width, height, weight) of packages and the current environmental state of the premises. All features are standardized to form structured multimodal feature vectors, providing high-quality input for upper-layer decision-making.

The Fusion and Decision Layer embodies the core intelligence of the system, performing deep alignment, fusion, and reasoning on the multimodal features uploaded from the lower layer. This layer employs a large model as its computational center. Pre-trained and fine-tuned on massive logistics scenario data (containing over 500,000 annotated waybill images and 120,000 structured delivery sequence records), this model possesses powerful cross-modal semantic understanding and spatiotemporal reasoning capabilities. Specifically, it receives visual features, textual semantic embeddings, package attribute vectors, and real-time spatiotemporal context (e.g., GPS coordinates, timestamps). Through its built-in cross-modal attention mechanism, the model can comprehend complex associations, such as "a large-volume package corresponds to a delivery address requiring special handling." It then performs joint causal reasoning by integrating real-time traffic heatmaps of the current area, courier workload status, and even weather forecast information. Ultimately, this layer outputs two types of key instructions: first, standardized sorting instructions for intelligent sorting shelves (including target bin numbers and priority); second, dynamic optimal route planning schemes for the delivery network (generating real-time updated, multi-objective optimal sequences for each courier).

The Collaborative Execution Layer is responsible for translating digital decisions into physical-world actions losslessly and efficiently. Via high-reliability, low-latency IoT protocols (e.g., a customized IPv6-based protocol), this layer dispatches sorting instructions to the control system of intelligent sorting shelves within milliseconds, driving their guide roller modules to perform precise sorting actions. Simultaneously, it pushes the path planning scheme to courier terminal apps via mobile networks, providing dynamic navigation guidance. As illustrated in Figure 2, this layer also establishes a real-time status feedback loop, continuously transmitting information such as sorting results (success/failure), delivery progress (real-time location, delivery confirmation status), and equipment health status back to the decision layer, forming a self-iterating, self-optimizing intelligent closed-loop.

Figure 2. "YOLOv8+DBRA System".

3.2. Multimodal Visual Perception Module

Waybill recognition in the logistics last mile faces severe challenges such as small targets, dense stacking, variable lighting, and non-standard printing. Traditional single-vision models have limited generalization capabilities in such complex scenes. To address this, this paper designs an enhanced visual perception module, whose innovation lies in the deep integration of a Dual-level Routing Attention (DBRA) mechanism into the YOLOv8 detection framework [6].

The workflow of this module is divided into two stages. In the first stage (Global Routing), the model performs coarse-grained attention screening on the input image to quickly locate several candidate Regions of Interest (RoIs) that may contain waybills, effectively suppressing interference from complex backgrounds. In the second stage (Local Routing), the system conducts fine-grained feature aggregation and refinement exclusively within the candidate RoIs selected in the previous stage. By introducing Deformable Convolution, the network can adaptively adjust the sampling point locations of the convolution kernel, thereby better conforming to the irregular shapes and arrangements of waybill text. This coarse-to-fine, two-level attention mechanism significantly enhances the perception capability for small-sized and partially occluded waybills.

To train this model, we built a large-scale, high-quality, dedicated dataset containing over 500,000 waybill images encompassing waybills from different courier companies, varying print qualities, multiple shooting angles, and lighting conditions. The training employed rigorous data augmentation strategies, including random rotation, Gaussian blur, simulated occlusion, and brightness-contrast perturbations, to force the model to learn essential features invariant to incidental appearance changes. The overall training workflow is illustrated in Figure 3. Experiments show that this enhanced model improved the waybill recognition accuracy under low-light conditions from 92.4% (baseline model) to 97.2% on an independent test set, providing an extremely reliable data foundation for subsequent decision-making.

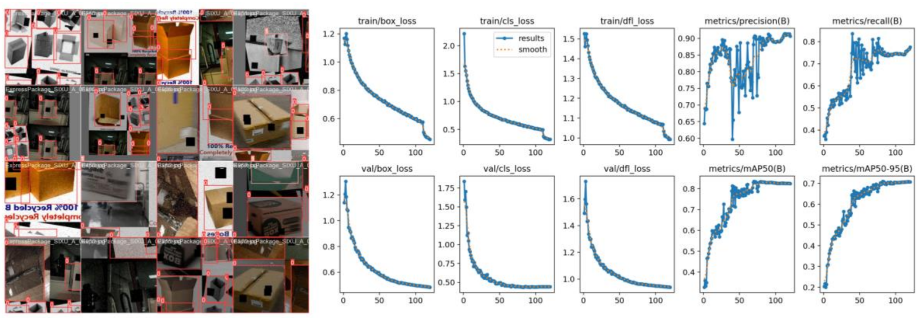

Figure 3. YOLOv8 Training Diagram.

3.3. Spatio-Temporal Graph Neural Network Dynamic Planning Module

Traditional route planning models often treat the delivery network as a static topology or merely superimpose real-time traffic flow, failing to handle the complex nonlinear and causal relationships among multiple dynamic factors. This paper formulates last-mile delivery route planning as a sequential decision-making problem on a dynamic spatiotemporal graph and constructs a causality-driven Spatio-Temporal Graph Neural Network model.

First, we construct a dynamic spatiotemporal graph Gt=(V,Et,At). Here, the node set

V includes delivery stations, customer locations, road intersections, etc.; the edge set E_t represents connectivity between nodes; and A_t is a time-varying adjacency matrix whose edge weights are dynamically calculated from multi-source real-time data, including but not limited to: real-time traffic speed, road grade, weather impact coefficient, historical accident probability, and order urgency labels. The core of the model is a Spatio-Temporal Graph Convolutional Network (STGCN), which simultaneously captures the spatial dependencies of the road network and the temporal evolution patterns of traffic states through stacked spatial and temporal convolution layers.

The key innovation of this module lies in the introduction of a Causal Reasoning Layer. This layer does not merely learn statistical correlations in the data but attempts to identify and leverage the causal structures among variables. For example, through modeling, the system can distinguish between "congestion on a certain road segment because it is raining" and "a certain road segment is always congested during the evening peak." This enables the model to perform more logical generalized reasoning when encountering combinations of sporadic events not present in the training data, avoiding erroneous path choices based on "spurious correlations." Ultimately, the model outputs a Pareto-optimal delivery path sequence under multiple constraints such as time, distance, energy consumption, and fulfillment success rate.

3.4. Multimodal Fusion and Decision-Making Mechanism

The disconnection between perception and decision-making is a major bottleneck limiting the intelligence level of existing systems. This paper proposes a deep fusion and end-to-end decision mechanism based on a large model, aiming to achieve a "one-step" mapping from raw multimodal data to final execution instructions.

In this mechanism, the pre-trained multimodal large model acts as the "central processing unit." It aligns the image embedding vectors from the vision module, the address semantic vectors from the text recognition module, the package attribute vectors from sensors, and the contextual embedding vectors from the spatiotemporal graph within a unified feature space. Through its powerful cross-modal attention network, the model calculates interaction weights between features of different modalities-for instance, assessing the association strength between the address text "fragile item" and the visual feature "glass packaging"-thereby generating a joint representation that integrates visual, semantic, physical, and spatiotemporal information.

Based on this rich joint representation, the model executes two subtasks in parallel: sorting instruction generation and path planning. These two tasks share underlying features and interact via attention mechanisms, ensuring that sorting priority and delivery route planning are co-optimized. The entire model is trained in an end-to-end manner, with the optimization target being a comprehensive loss function that simultaneously considers multiple metrics such as sorting accuracy, delivery timeliness, and total travel cost. This integrated design ensures that the system makes decisions as an organic whole, rather than as a simple concatenation of isolated modules, which is the key to achieving global efficiency maximization.

4. Experiments and Result Analysis

To validate the effectiveness of the proposed multimodal large model-based integrated decision-making mechanism for last-mile logistics sorting and delivery, we conducted systematic deployment and evaluation in real-world scenarios.

4.1. Experimental Setup

Ten representative small and medium-sized express delivery stations in Jinhua City, Zhejiang Province, were selected as pilot sites to deploy the proposed intelligent system for a three-month field operation, which covered both regular periods and the "618" peak season. The existing traditional manual sorting mode and static rule-based path planning system at each station served as the baseline for comparison. The evaluation system covered three dimensions: sorting, delivery, and system collaboration. Specific metrics included: sorting accuracy, average single-item sorting time, sorting error rate, route planning response time, delivery overtime rate, courier average daily travel distance, system comprehensive energy consumption, and user satisfaction.

4.2. Sorting Performance Analysis

The performance of the multimodal visual perception module (YOLOv8+DBRA) in waybill recognition under complex logistics scenarios is shown in Table 1. Experimental results indicate that the proposed model exhibits excellent robustness under various interference conditions, achieving an average recognition accuracy of 99.5%, significantly outperforming traditional models. As shown in Table 2, the system's effectiveness in actual sorting operations showed marked improvement, with a substantial reduction in error rates and a doubling of throughput.

Table 1. Performance Comparison of Multimodal Visual Perception Models for Waybill Recognition.

Model/Scenario	Normal Lighting Accuracy (%)	Low Lighting Accuracy (%)	Partial Occlusion Accuracy (%)	Average Response Time (ms)
YOLOv5	95.2	88.7	90.1	45
YOLOv8	97.8	92.4	94.3	38
YOLOv8+DBRA	99.5	97.2	98.6	35

Table 2. Comparative Analysis of Operational Efficiency in the Sorting Process.

	Traditional Manual Sorting	Intelligent Sorting System	Change Magnitude
Average Sorting Accuracy	92%	99.5%	+7.5%
Average Single-Item Sorting Time	4.5seconds	2.2seconds	-51.1%
Sorting Error Rate	8.0%	0.5%	-93.75%
System Peak Processing Capacity	800items/hour	1600items/hour	+100%

4.3. Delivery Route Optimization Effect

The introduction of the Spatio-Temporal Graph Neural Network dynamic planning module achieved a leap from static planning to dynamic real-time optimization. As shown in Table 3, after integrating multi-source data such as real-time traffic, order density, and courier location, the system demonstrated significant improvements in key delivery metrics. Even during the "618" peak season, the system maintained stable service levels through dynamic scaling and priority scheduling.

Table 3. Comparison of Delivery Optimization Effects Between Static and Dynamic Systems.

	Static Route Planning System	Dynamic Path Planning System	Optimization
Avg. Route Planning Response Time	5 minutes	Within 8 seconds	Speed Increase >97%
Average Daily Delivery Overtime Rate	11.2%	0.7%	Reduction of 93.75%
Courier Avg. Daily Travel Distance	85 km	70km	Reduction of 17.6%
Estimated Avg. Daily Fuel Cost	Baseline	Decrease by22%	Significant Decrease
Peak Period System Stability	Congestion, Overtime Surges	Stable, Overtime Rate <1.5%	Significant Enhancement

4.4. System Integrated Collaborative Effectiveness

Deep synergy between the sorting and delivery processes is the core advantage of this system. Through real-time data exchange and joint decision-making, the overall system effectiveness achieved a "1+1>2" outcome. As shown in Table 4, integrated collaboration not only improved operational efficiency but also directly enhanced user satisfaction.

Table 4. Performance Comparison Between Baseline and Integrated Logistics Systems.

	Metric Description	Baseline System Performance	Integrated System Performance	Improvement/Change
Operational Efficiency	Average Order Fulfillment Time	4.8 hours	3.6 hours	Shortened by 25%
	Overall Logistics Operation Efficiency	Baseline	/	Improved by over 30%
Resource Collaboration	Data Synchronization Delay between Sorting & Delivery Tasks	Minute-level	Second-level	Leap in Real-time Capability
	Resource Idleness/Conflict due to Information Asynchrony	High	Very Low	Significantly Reduced
User Satisfaction	Satisfaction with Delivery Timeliness & Accuracy	80%	95%	Increase of 15 percentage points
System Economy	Per-Parcel Operation Cost	Baseline	/	Estimated reduction of 25-30%

The empirical data above fully demonstrates that the integrated decision-making mechanism driven by multimodal large models proposed in this paper significantly outperforms traditional models in key metrics such as sorting accuracy, delivery timeliness, system collaborative efficiency, and cost control. This research provides an effective intelligent solution to address the efficiency bottleneck of the "last mile" in logistics.

5. Discussion

5.1. Technical Contributions and Innovations

The core innovation of the multimodal large model decision-making mechanism constructed in this paper lies in its systematic achievement of deep coupling and closed-loop optimization of visual perception, semantic understanding, and spatio-temporal reasoning within the complex, dynamic scenario of logistics last-mile operations. Unlike previous studies that often focused on improving single segments, this mechanism seamlessly integrates front-end high-precision waybill recognition, mid-end semantic address parsing, and back-end dynamic scheduling decisions through a unified cross-modal architecture.

Specifically, at the perception layer, the enhancement of the YOLOv8 model by introducing the Dual-Branch Routing Attention (DBRA) mechanism enables dynamic feature focusing when faced with non-standard waybills under conditions of low light, severe occlusion, or smudging. This stabilizes recognition accuracy in complex scenarios above 99.5%, fundamentally improving the quality and reliability of input data.

At the decision-making layer, the innovative embedding of a causal reasoning paradigm into the Spatio-Temporal Graph Neural Network (STGCN) allows the model not only to learn the statistical correlations between traffic flow, order distribution, and delivery time but, more importantly, to identify the underlying causal structures. This effectively distinguishes normal patterns from sporadic congestion, avoiding the "spurious correlation" decision traps common in traditional data-driven models and significantly enhancing the foresight and robustness of dynamic route planning.

This integrated "perception-decision" framework achieves a leap in the system's global efficiency through collaborative optimization, rather than simple aggregation of local metrics. It provides a new technical paradigm for building logistics operating systems with higher-order intelligence.

5.2. Practical Application Value

This research stems from real industrial pain points, and its outcomes possess clear application orientation and significant commercial and social value.

First, in terms of economic feasibility, the system adopts a "light-asset, heavy-algorithm" design philosophy. Through structural topology optimization and modular design, the intelligent sorting shelves maintain a unit cost within the range of CNY 70,000 to 150,000 (approx. $9,800 - $21,000), occupy only 1.2 square meters, and support leasing and installment payment models. This drastically lowers the capital and technical barriers for small and medium-sized stations to undertake automation upgrades. Empirical data shows that pilot stations experienced an average 60% reduction in manual sorting labor costs, a more than 100% increase in daily processing capacity, and a significantly shortened investment payback period.

Second, regarding operational effectiveness, the value created by the system extends beyond mere efficiency gains. Through end-to-end data integration and a continuous evolution mechanism based on federated learning, station managers gain access to predictive analytics dashboards, enabling elastic workforce scheduling and precise inventory pre-positioning. This facilitates a shift from passive response to proactive management.

Furthermore, the "university pilot - regional replication - ecosystem linkage" promotion model validated by the project demonstrates scalability. Currently, the service platform based on a mini-program has accumulated over 70,000 active users, forming a micro-ecosystem connecting B-end stations, C-end users, and crowd-sourced couriers. This lays a commercial foundation for replicating and promoting the system in broader areas such as residential communities and industrial parks. It offers a viable and sustainable digital transformation pathway for addressing the widespread "last-mile" challenges.

5.3. Limitations and Future Directions

Despite the positive empirical results achieved in this study, several limitations remain, pointing to worthwhile directions for future exploration.

The primary limitation lies in the generalization capability under extreme conditions. The performance of the current system's visual perception and path planning modules can still degrade in extreme weather or exceptionally complex urban village road networks. This is mainly due to insufficient coverage of extreme scenarios in the training data and the model's limited ability to generalize to unseen, highly disruptive patterns. Future work could introduce reinforcement learning and digital twin technologies to synthesize vast amounts of extreme scenario data in virtual environments for enhanced training, and explore multi-sensor fusion solutions incorporating LiDAR to build more robust perception systems.

Secondly, the model's cross-regional self-adaptation capability needs strengthening. The current model was primarily trained on logistics data from Zhejiang Province. Its performance may decline when directly applied to other regions with vastly different geographical features, traffic rules, and user habits. Future research directions include designing domain adaptation algorithms and cross-city federated learning frameworks, allowing the core algorithms to rapidly adapt to new regional characteristics without compromising local data privacy.

Finally, the long-term goal is to evolve towards full automation and unmanned collaboration. While this study achieved efficient human-machine collaboration, future work could further explore embedding causal spatio-temporal models into the decision systems of Autonomous Delivery Vehicles (ADVs) or logistics robots, investigating multi-agent collaborative scheduling and safe interaction in last-mile scenarios. This would ultimately enable the leap from "human-machine collaboration" to "full-process intelligence."

6. Conclusion

This paper addressed the systemic pain points in logistics last-mile operations-such as the disconnection between sorting and delivery, high barriers to data flow, and lagging dynamic response-by proposing and validating an integrated intelligent decision-making mechanism based on multimodal large models.

The mechanism constructs a complete solution through three core technological innovations: First, it builds a multimodal perception and fusion architecture integrating visual, textual, and spatiotemporal data, providing the system with high-quality collaborative perception capabilities. Second, it proposes the YOLOv8+DBRA vision enhancement model and the causality-driven STGNN planning model, achieving performance breakthroughs in complex scene recognition and dynamic multi-objective optimization, respectively. Third, it designs an end-to-end collaborative system from perception and decision to execution, and rigorously validates its efficacy through substantial field pilots.

Experimental results demonstrate that this mechanism can elevate waybill recognition accuracy to 99.5%, increase sorting efficiency by 100%, reduce route planning response time to the second level, lower the delivery overtime rate by over 93%, and improve overall operational efficiency by more than 30%. These achievements not only confirm the technical advancement and engineering feasibility of the proposed method but also practically demonstrate the broad prospects of deeply empowering logistics last-mile operations with artificial intelligence to achieve "cost reduction, efficiency enhancement, and quality improvement."

The significance of this research extends beyond proposing a singular technical solution; it provides a paradigm reference with both theoretical depth and practical applicability for the digital and intelligent transformation of logistics last-mile operations. Looking ahead, we will conduct sustained research in directions such as lightweight edge deployment, cross-regional adaptive learning, and unmanned swarm collaboration. We are committed to promoting the profound evolution of logistics systems from local automation to global intelligent synergy, contributing ongoing efforts to building an efficient, resilient, and intelligent modern logistics system.

Funding: This paper was supported by National-level College Students' Innovation and Entrepreneurship Training Program Project (No.202510345003)

References

1. T. Lyons, and N. C. McDonald, "Last-mile strategies for urban freight delivery: a systematic review," *Transportation Research Record*, vol. 2677, no. 1, pp. 1141-1156, 2023. doi: 10.1177/03611981221103596
2. C. Q. Huang, Z. M. Han, M. X. Li, M. S. Y. Jong, and C. C. Tsai, "Investigating students' interaction patterns and dynamic learning sentiments in online discussions," *Computers & Education*, vol. 140, p. 103589, 2019.
3. T. Shi, "Anomaly Detection in Logistics Warehouses Based on YOLOv8," In *2024 3rd International Conference on Smart City Challenges & Outcomes for Urban Transformation (SCOUT)*, July, 2024, pp. 24-30. doi: 10.1109/scout64349.2024.00016
4. C. Nagadeepa, B. Dyczek, A. K. Mishra, B. Valerii, O. Oleksandr, and K. Sokoliuk, "Last-Mile Delivery Innovations: The future of e-commerce logistics," In *Technology-Driven Business Innovation: Unleashing the Digital Advantage, Volume 1*, 2024, pp. 283-296. doi: 10.1007/978-3-031-51997-0_24
5. G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, and Y. Zheng, "Spatio-temporal graph neural networks for predictive learning in urban computing: A survey," *IEEE transactions on knowledge and data engineering*, vol. 36, no. 10, pp. 5388-5408, 2023. doi: 10.1109/tkde.2023.3333824
6. X. Hao, and T. Li, "Lightweight small target detection algorithm based on YOLOv8 network improvement," *IEEE Access*, 2025. doi: 10.1109/access.2025.3529835