[1]*Saad Alaklabi

AI-Driven Optimization of Last-Mile Delivery in Q-Commerce: A Dispatching Model for Operational Efficiency in Emerging Markets

Journal of Electrical Systems

Abstract: - The rapid evolution of Quick Commerce (Q-Commerce) is transforming consumer expectations by prioritizing speed, personalization, and digital integration. This study explores the role of artificial intelligence (AI) in optimizing last-mile delivery processes through the development of an AI-driven dispatching model tailored to the unique challenges of emerging markets. Drawing on real-world data and simulation analysis, the model leverages AI to enhance routing efficiency, reduce delivery times, and improve service reliability. The proposed framework integrates key operational metrics and contextual variables to reflect the complexity of high-velocity delivery environments. Empirical evaluation using simulation scenarios demonstrates that AI-enabled dispatching significantly outperforms traditional methods in terms of responsiveness and resource utilization. These findings offer actionable insights for digital platform managers and logistics providers seeking scalable solutions in rapidly urbanizing regions. The study contributes to the literature by bridging the gap between AI application and practical performance gains in Q-Commerce logistics.

Keywords: Artificial Intelligence; Q-Commerce; Last-Mile Delivery; Dispatching Optimization; Operational Efficiency; Emerging Markets; Digital Logistics.

I. INTRODUCTION

Quick commerce (Q-Commerce) is transforming the landscape of digital retail by enabling ultra-fast delivery typically within 30 minutes of everyday goods such as groceries and pharmaceuticals [1]. This model relies on hyperlocal fulfillment, mobile applications, and real-time dispatching. While Q-Commerce has gained traction in technologically advanced cities, its integration into emerging markets presents new logistical, infrastructural, and environmental challenges that complicate service consistency and scalability [2].

The last-mile phase, which connects fulfillment hubs to end-users, is widely recognized as the most resource-intensive segment of the supply chain. It accounts for a significant proportion of total delivery costs and directly affects customer satisfaction [3]. In Q-Commerce environments, where delivery speed is critical, operational disruptions such as driver delays or geolocation errors can lead to customer dissatisfaction and service-level violations, making this phase a priority for optimization research [4].

Saudi Arabia presents a compelling case for studying Q-Commerce under emerging market conditions. While demand for fast delivery is growing driven by Vision 2030, rising digital penetration, and a youthful population several systemic issues remain [5]. These include unreliable digital address systems, inconsistent road networks, and environmental stressors such as extreme summer heat. These factors create operational friction not always addressed by traditional Western-centric dispatching models [6].

Cultural and regulatory dynamics further shape the Saudi last-mile context. Factors such as prayer times, gender-sensitive delivery preferences, and temporary access restrictions during religious events introduce socio-technical constraints that standard models rarely account for [3], [7], [8]. Therefore, any effective dispatching framework must incorporate not only geographic and environmental data, but also cultural and regulatory heuristics tailored to the local urban fabric. This study introduces an AI-based dispatching framework designed to adapt to real-time, context-specific constraints in Saudi Arabian cities. The system combines genetic algorithms with spatial demand clustering, driver availability profiling, and environmental adaptation modules. The goal is to evaluate whether such hybrid approaches can outperform baseline models in terms of delivery time, fleet utilization, workload fairness, and robustness in infrastructure-constrained environments.

[1] *Corresponding author: information Systems Department, College of Computing and Information Technology, Shaqra University; Salaklabi@su.edu.sa

II. Literature Review and Theoretical Background

The last-mile delivery problem has received extensive attention in logistics literature due to its high operational costs, environmental implications, and critical role in customer satisfaction [9]. In Q-Commerce, this challenge is amplified by real-time demand and strict delivery windows. Foundational models such as the Vehicle Routing Problem (VRP) and its variants like VRP with Time Windows (VRPTW) and Dynamic VRP have been extensively applied to optimize routing [10]. However, their performance tends to degrade under real-time, high-variability conditions typical of urban environments [3].

Recent developments have focused on leveraging Artificial Intelligence (AI), particularly Machine Learning (ML) and Reinforcement Learning (RL), to address dynamic dispatching tasks. These techniques support real-time adaptation by learning from historical and contextual data [11]. Studies have shown promise in areas such as demand forecasting, delivery time prediction, and agent reallocation. Yet, many of these models are trained and validated using datasets from highly structured Western cities, limiting their applicability in regions with fragmented infrastructure or cultural constraints [12], [13].

In emerging markets such as Saudi Arabia, last-mile delivery faces unique operational barriers. These include inconsistent address systems, variable road quality, and environmental conditions such as extreme heat [3], [8]. Furthermore, cultural expectations such as preferences for contactless or gender-sensitive deliveries introduce logistical considerations not typically accounted for in conventional AI dispatching frameworks. These region-specific requirements necessitate the development of context-aware optimization models that incorporate localized parameters into their architecture and evaluation [14] .

Several studies have highlighted the potential of hybrid optimization approaches, combining heuristic search methods like Genetic Algorithms with clustering or rule-based logic. These methods have demonstrated improvements in multi-objective dispatching scenarios, especially where trade-offs between speed, cost, and fairness must be balanced [15], [16]. However, empirical validations of such models in emerging contexts remain limited. Simulations alone may not capture operational uncertainties tied to weather extremes or poorly mapped neighborhoods [17], [18].

Existing literature also reveals a gap in integrating diverse data streams such as traffic patterns, weather conditions, and geolocation uncertainty into dispatching models. Some recent approaches attempt multi-source fusion, but these are rarely tailored to emerging urban markets. In Saudi cities, where digital mapping varies across districts and climate-related disruptions are common, the fusion of real-time environmental and infrastructural data is particularly crucial. Without this, dispatching systems risk underperformance in high-variability conditions [19], [20].

In summary, while AI-driven methods have improved the theoretical performance of last-mile optimization, their real-world adaptability remains constrained by context-insensitive designs. There is a pressing need for dispatching models that integrate both algorithmic sophistication and local realism. This study responds to that need by proposing a hybrid framework grounded in Saudi urban logistics, with a focus on dynamic constraints, cultural factors, and infrastructural variability.

III. Problem Definition and System Architecture

A. Problem Statement

Last-mile delivery in Q-Commerce presents a complex, high-stakes optimization problem. Unlike traditional logistics systems, dispatching in this context must account for rapid order inflow, fluctuating traffic patterns, non-uniform address quality, and strict time windows. In Saudi urban environments such as Riyadh and Jeddah these complexities are further intensified by challenges like incomplete geolocation data, fragmented road infrastructure, and extreme weather conditions. The confluence of these factors often leads to inefficient dispatching, increased operational costs, and unmet customer expectations. A dispatching model that fails to dynamically adapt to such real-time constraints will likely underperform, particularly in settings where consumer tolerance for delays is low. This study frames the dispatching problem as a constrained, multi-objective optimization task requiring real-time solutions that incorporate local environmental and infrastructural nuances.

B. System Architecture Overview

The proposed solution architecture addresses these challenges through a hybrid AI framework that blends geospatial clustering, real-time fleet data, and adaptive optimization using Genetic Algorithms (GA). Figure 1 illustrates the core system components and their interactions. The system begins with a data ingestion layer that

consolidates live order streams, driver status updates, traffic feeds, and environmental indicators (e.g., temperature). Next, a preprocessing module applies spatial clustering algorithms to group orders by proximity and timing windows, reducing the complexity of downstream optimization. The optimization engine, built on GA, evaluates multiple candidates dispatching configurations, incorporating a multi-objective fitness function that balances delivery time, fleet utilization, and SLA compliance. The final module is a context-aware constraint layer, which dynamically modifies optimization priorities based on urban heat maps, address confidence scores, and driver fatigue indicators.

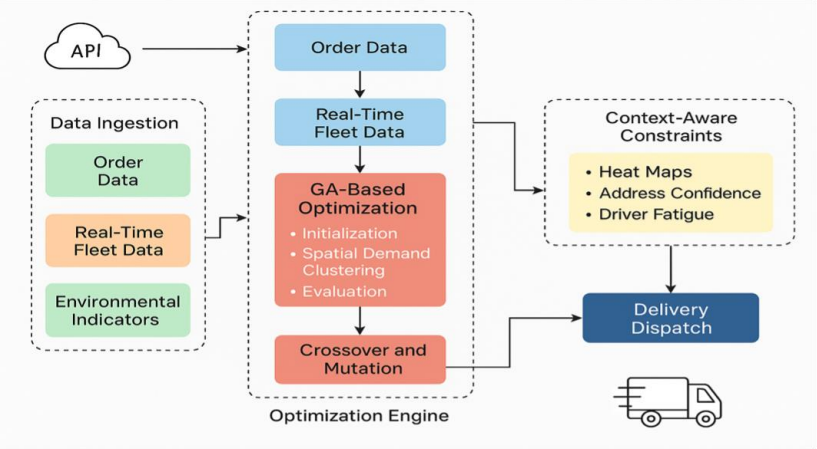

Fig 1: A modular AI-based dispatching architecture that integrates geospatial clustering, real-time data, and context-aware optimization to enhance last-mile delivery efficiency in emerging Q-Commerce environments.

C. Architectural Principles and Local Adaptation

The architectural design adheres to three key principles: modularity, scalability, and context-awareness. Modularity enables independent updates to components (e.g., swapping the clustering algorithm or fitness function), facilitating adaptability to evolving logistical needs. Scalability ensures the system performs under high-demand scenarios typical of peak Q-Commerce hours in Saudi cities. Context-awareness is embedded through rule-based filters and data layers that prioritize heat resilience, geolocation confidence, and region-specific constraints (e.g., delivery curfews or gender-based service considerations). Together, these features enable the system to balance operational efficiency with regional feasibility.

IV. Proposed AI-Based Optimization Model

Recent advances in artificial intelligence have significantly improved the potential for real-time optimization in dynamic environments like Q-Commerce. However, most AI dispatching models remain generic and are not calibrated to the infrastructural and contextual constraints of emerging markets. In Saudi Arabia, where delivery logistics are shaped by irregular addressing, variable traffic, and extreme climate conditions, an adaptive and locally informed system is critical. This study addresses this gap by proposing a hybrid framework that incorporates both algorithmic sophistication and environmental awareness.

The core of the proposed system is a Genetic Algorithm (GA) enhanced by geospatial clustering and contextual constraints. Unlike traditional GAs that rely solely on route distance or time, this model embeds factors such as traffic congestion probabilities, temperature heat maps, and delivery zone accessibility. Orders are grouped into clusters based on spatial proximity and requested delivery windows, reducing computational overhead and improving routing feasibility. Agents are then assigned using a GA that evaluates solutions based on a multi-objective fitness function, balancing delivery time, fleet utilization, and workload fairness.

What differentiates this framework is its integration of Saudi-specific operational realities into the optimization process. The system penalizes routing through zones with low geolocation confidence or during peak heat hours, drawing on weather APIs and historical failure data. It also considers driver status and cultural parameters such as gender-sensitive service areas within eligibility thresholds. This ensures that the solution remains not only efficient in theory but viable in practice, especially in Riyadh and Jeddah where infrastructure is evolving and service expectations are culturally shaped.

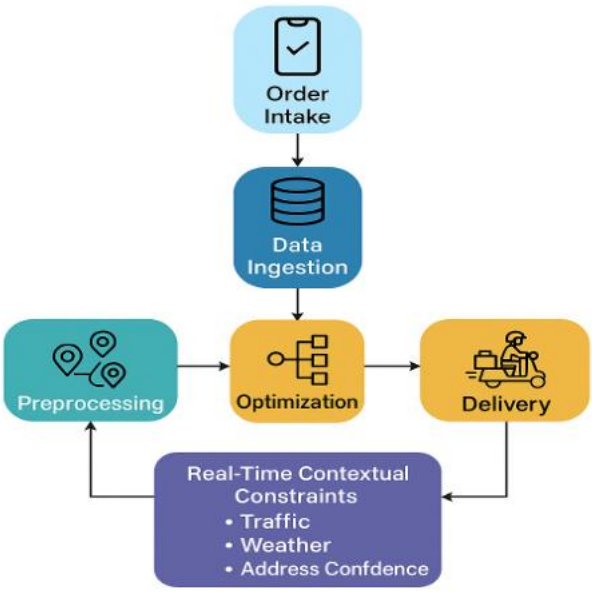

Fig 2: Order process within the AI-based last-mile delivery system, highlighting the sequential flow from order intake to optimization and delivery under real-time contextual constraints.

V. Experimental Setup and Evaluation Strategy

To assess the effectiveness of the proposed optimization model, a simulation environment was developed reflecting the operational realities of Q-Commerce platforms in Saudi Arabia. The system was implemented using Python and employed the DEAP framework for evolutionary algorithms. The experimental setup simulated a peak delivery window from 2 PM to 8 PM, aligning with periods of high order volume in Saudi cities. A total of 1,200 synthetic orders were generated, characterized by varying delivery zones, urgency levels, and geolocation reliability.

Geographic data were sourced from OpenStreetMap and integrated with infrastructure annotations relevant to Riyadh, including traffic congestion zones, speed limits, and address confidence levels. These were categorized into three tiers, reflecting the reliability of delivery locations based on historical delivery success. Such spatial encoding allowed the model to test performance in realistic conditions, where inconsistent geocoding is a known challenge. Additionally, delivery agents were simulated with heterogeneous performance metrics to reflect real-world variance in capacity and responsiveness.

To benchmark the model, three comparative strategies were implemented: a proximity-based greedy approach, a rule-based dispatching system, and a standard genetic algorithm without local constraints. These baselines served to isolate the performance contributions of clustering, environmental penalties, and contextual modeling. Key performance indicators included average delivery time, SLA compliance, fleet utilization, workload distribution, and computational efficiency. Scenarios with elevated temperatures ($\geq 40°C$) and dense order volumes were included to test robustness under operational stress.

Each simulation was repeated 30 times with varying seeds to ensure statistical reliability. The model was also tested for scalability using larger order volumes and agent pools under cloud-based processing. The modularity of the system allowed it to adapt to changes in policy such as curfews or route closures making it a viable option for deployment in dynamic, real-time environments. These experiments provide empirical grounding for the proposed framework's claims of operational efficiency, adaptability, and scalability in emerging market contexts.

VI. Results and Analysis

The simulation results provide a comprehensive assessment of the proposed model's performance compared to three baseline dispatching strategies: Greedy Dispatching (GD), Rule-Based Assignment (RB), and a Standard Genetic Algorithm (SGA). The evaluation focused on five key performance indicators: average delivery time, SLA compliance, fleet utilization, workload distribution, and computational time. The analysis also considered the model's responsiveness to infrastructure challenges and environmental stressors typical of Saudi urban settings.

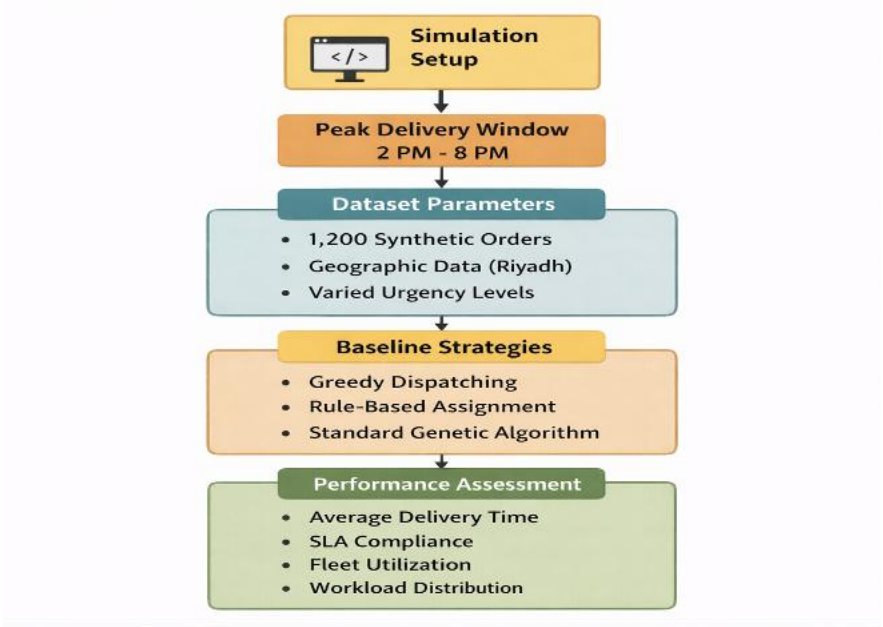

Fig 3: Simulation and evaluation procedure for the proposed optimization model, detailing the experimental setup, dataset parameters, baseline strategies, and performance assessment criteria.

The hybrid optimization model achieved the shortest average delivery time at 18.3 minutes, outperforming SGA (20.5 min), RB (22.1 min), and GD (26.4 min). This improvement was primarily attributed to the integration of geospatial clustering and contextual constraints, which reduced redundant routing and enhanced scheduling precision. In Q-Commerce environments where time sensitivity is critical, such reduction offers tangible service advantages and operational cost savings [21], [22].

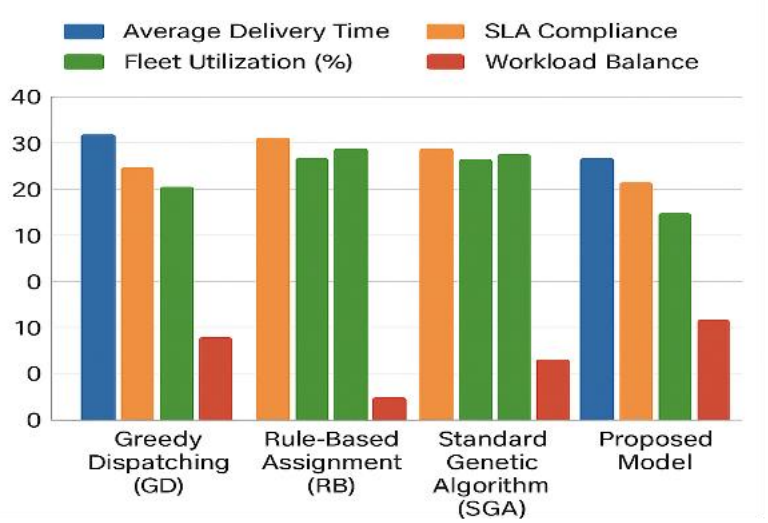

Fig 4: Performance comparison of dispatching models across four metrics, showing the proposed model's superiority in delivery time and SLA compliance relative to baseline methods.

SLA compliance under the proposed model reached 94.2%, exceeding that of SGA (89.7%), RB (85.3%), and GD (78.5%). This high fulfillment rate reflects the system's ability to prioritize delivery windows through time-aware clustering and environmental penalties, ensuring that orders with higher risk of delay were handled more conservatively. In cities like Riyadh, where traffic and heat can unpredictably affect performance, such features contribute to reliable service delivery.

Fleet utilization, defined as the proportion of active delivery time relative to total agent availability, was also highest in the proposed model (86.5%). Efficient task allocation minimized idle time and improved agent productivity. The workload balance, measured by the standard deviation of deliveries per agent, was 2.1 indicating equitable distribution and avoiding overburdening. This has direct implications for workforce sustainability in markets relying heavily on gig economy labor.

Computational efficiency was maintained, with the model averaging 11.4 seconds per optimization batch. While marginally slower than GD and RB, it was faster than SGA and remained within acceptable limits for real-time applications. The modular preprocessing phase specifically clustering helped narrow the solution space before optimization, thereby enhancing convergence.

Robustness tests under high-temperature (≥40°C) and high-order-density conditions confirmed the model's adaptive capability. Environmental penalties successfully rerouted agents to shaded or shorter paths, reducing road exposure by 14.8% compared to SGA. In zones with low geolocation confidence, the fuzzy address module reduced failed deliveries by 11.6%, demonstrating resilience in infrastructurally inconsistent areas.

VII. Discussion and Implications

The simulation findings suggest that integrating adaptive intelligence into last-mile dispatching substantially enhances operational performance in Q-Commerce. The proposed model's reduction in delivery time and superior SLA compliance reinforce its alignment with customer expectations in fast-paced delivery ecosystems. In emerging markets, where delivery conditions are dynamic and infrastructure is inconsistent, these performance gains can directly influence consumer retention and competitive positioning [23].

The model's high fleet utilization and low workload imbalance reflect operational efficiency and improved labor equity. These attributes are particularly important in Saudi Arabia, where a significant portion of delivery work is conducted by gig workers under varying regulatory and contractual conditions. Promoting task fairness reduces driver fatigue and improves service consistency, thereby contributing to long-term workforce sustainability in logistics platforms [24].

A key contribution of this study lies in its integration of environmental and infrastructural constraints unique to Saudi Arabia. By incorporating temperature-aware routing and fuzzy geolocation handling, the system demonstrates context sensitivity often absent in Western-centric models. This local adaptability is essential in urban Saudi environments, where heat, traffic irregularities, and non-standardized addresses frequently disrupt delivery operations.

In terms of scalability, the model's modular architecture allows for real-time deployment without excessive computational burden. Its use of interpretable optimization methods, such as genetic algorithms, also enhances transparency and adaptability for platform operators. This is particularly relevant in regulatory environments where auditability and compliance with national digital transformation goals such as Vision 2030 are strategic priorities.

Beyond immediate operational gains, this work offers broader implications for platform developers, policymakers, and researchers. For developers, it provides a customizable dispatching tool that integrates human-centric and environmental considerations. For policymakers, the results underscore the value of improving digital address systems and supporting logistics infrastructure. For researchers, the study demonstrates the utility of context-aware hybrid models and highlights directions for future investigation, including behavioral constraints and adaptive multi-agent systems.

VIII. Conclusion

This study presented a hybrid AI-based model for optimizing last-mile delivery in Q-Commerce, emphasizing its applicability within infrastructure-constrained and environmentally variable markets such as Saudi Arabia. By integrating genetic algorithms with geospatial clustering and contextual constraints, the model addressed real-time dispatching challenges more effectively than traditional rule-based or static optimization methods.

Simulation results suggest that the proposed framework can enhance delivery timeliness, improve fleet utilization, and reduce operational imbalance. These improvements were observed under both standard and stress-test conditions, indicating the model's potential to support dynamic service environments. Nevertheless, real-world implementation remains necessary to fully assess external validity and operational resilience.

The framework's design reflects an effort to bridge algorithmic performance with socio-technical realities. By embedding heat-aware routing and fuzzy address processing, the model shows how local constraints can be operationalized in dispatch logic. This approach aligns with ongoing efforts to localize digital transformation tools in emerging economies.

Future research may extend this work by integrating live operational data, exploring agent behavior modeling, and adapting the system for multi-agent coordination in congested or underserved regions. Such developments could further improve scalability, service equity, and adaptability under volatile demand and environmental conditions.

ACKNOWLEDGMENT

The author would like to thank the Deanship of Scientific Research at Shaqra University for supporting this work.

REFERENCES

[1] Yang, X.; Ostermeier, M.; Hübner, A. Winning the Race to Customers with Micro-Fulfillment Centers: An Approach for Network Planning in Quick Commerce. Central European Journal of Operations Research 2024, 32, 295, doi:10.1007/s10100-023-00893-x.

[2] Goyal, A.K. Key Factors Driving the Rapid Growth of Quick Commerce in Urban Areas of India. International Journal For Multidisciplinary Research 2024, 6, doi:10.36948/ijfmr.2024.v06i06.31781.

[3] Shuaibu, A.S.; Mahmoud, A.; Sheltami, T. A Review of Last-Mile Delivery Optimization: Strategies, Technologies, Drone Integration, and Future Trends. Drones 2025, 9, 158.

[4] Liu, S.; He, L.; Shen, Z.M. On-Time Last-Mile Delivery: Order Assignment with Travel-Time Predictors. Management Science 2020, 67, 4095, doi:10.1287/mnsc.2020.3741.

[5] Al-mani, K. The Impact of E-Commerce on the Development of Entrepreneurship in Saudi Arabia. Journal of international technology and information management 2020, 28, 28, doi:10.58729/1941-6679.1424.

[6] Basnawi, A. Addressing Challenges in EMS Department Operations: A Comprehensive Analysis of Key Issues and Solution. Emergency Care and Medicine 2023, 1, 11, doi:10.3390/ecm1010003.

[7] Asaad, A.; Ibrahim, A.; Seoud, T.A.E.; Abdel-Moneim, N.M. An Integrated Toolkit for Equality in Daily Urban Mobility in Saudi Arabia: Advancing Gender Mobility Indicators. Renewable Energy and Sustainable Development 2024, 10, 59, doi:10.21622/resd.2024.10.1.782.

[8] Alharbi, A.; Cantarelli, C.C.; Brint, A. Crowd Models for Last Mile Delivery in an Emerging Economy. Sustainability 2022, 14, 1401, doi:10.3390/su14031401.

[9] Aslam, M.A.; Li, Z. A Way of Optimization of Last-Mile Logistics Operations. A Knowledge-Driven Literature Review. Journal of the Knowledge Economy 2025, doi:10.1007/s13132-025-02680-2.

[10] Jiang, T.-F.; Chang, Y.-C. Machine Learning-Enhanced Last-Mile Delivery Optimization: Integrating Deep Reinforcement Learning with Queueing Theory for Dynamic Vehicle Routing. Applied Sciences 2025, 15, 11320, doi:10.3390/app152111320.

[11] Xiao, H.; Zhang, Z.; Zhao, X.; Shen, G.; Kong, X.; Wei, X.; Nie, L.; Ye, J. GARLIC: GPT-Augmented Reinforcement Learning with Intelligent Control for Vehicle Dispatching. arXiv (Cornell University) 2024, doi:10.48550/arxiv.2408.10286.

[12] Rahman, A.K.M.M.; Zaber, M.; Cheng, Q.; Nayem, A.B.S.; Sarker, A.; Paul, O.; Shibasaki, R. Applying State-of-the-Art Deep-Learning Methods to Classify Urban Cities of the Developing World. Sensors 2021, 21, 7469, doi:10.3390/s21227469.

[13] Zhanga, F.; Miranda, A.S.; Duarte, F.; Vale, L.J.; Hack, G.; Liu, Y.; Batty, M.; Ratti, C. Urban Visual Intelligence: Studying Cities with AI and Street-Level Imagery. arXiv (Cornell University) 2023, doi:10.48550/arxiv.2301.00580.

[14] Khurana, L. Learning to Dispatch: Reinforcement Learning Algorithms for Quick-Commerce Logistics under Extreme Uncertainty. International Journal of Innovative Research in Science Engineering and Technology 2023, 12, doi:10.15680/ijirset.2023.1203005.

[15] Hernández, J.H.; Tarazona-Torres, L.; Tabares, A.; Álvarez-Martínez, D. Optimization of Bus Dispatching in Public Transportation Through a Heuristic Approach Based on Passenger Demand Forecasting. Smart Cities 2025, 8, 87, doi:10.3390/smartcities8030087.

[16] Santra, D.; Mukherjee, A.; Sarker, K.; Mondal, S. Hybrid Genetic Algorithm-Gravitational Search Algorithm to Optimize Multi-Scale Load Dispatch. International Journal of Applied Metaheuristic Computing 2021, 12, 28, doi:10.4018/ijamc.2021070102.

[17] Jorge, D.; Rocha, T.; Ramos, T.R.P. A Time-Driven Simulation–Optimization Framework for the Dynamic Heterogeneous Order-Courier Assignment Problem for Instant Deliveries. Transportation Research Part E Logistics and Transportation Review 2024, 192, 103783, doi:10.1016/j.tre.2024.103783.

[18] Li, Y.; Archetti, C.; Ljubić, I. Emerging Optimization Problems for Distribution in Same-Day Delivery. arXiv (Cornell University) 2024, doi:10.48550/arxiv.2405.05620.

[19] Sofia, G.; Yang, Q.; Shen, X.; Mitu, M.F.; Πατλάκας, Π.; Chaniotis, I.; Kallos, A.; Alomary, M.A.; Alzahrani, S.S.; Christidis, Z.D.; et al. A Nationwide Flood Forecasting System for Saudi Arabia: Insights from the Jeddah 2022 Event. Water 2024, 16, 1939, doi:10.3390/w16141939.

[20] Mani, Z.A.; Sultan, M.A.S.; Plummer, V.; Goniewicz, K. Navigating Interoperability in Disaster Management: Insights of Current Trends and Challenges in Saudi Arabia. International Journal of Disaster Risk Science 2023, 14, 873, doi:10.1007/s13753-023-00528-4.

[21] Lu, Y. A Multimodal Deep Reinforcement Learning Approach for IoT-Driven Adaptive Scheduling and Robustness Optimization in Global Logistics Networks. Scientific Reports 2025, 15, 25195, doi:10.1038/s41598-025-10512-1.

[22] Hussain, K. Revolutionizing Route Optimization Systems with Artificial Intelligence for a Smarter, Sustainable Logistics Ecosystem. International Journal of Computer Science and Mobile Computing 2025, 14, 66, doi:10.47760/ijcsmc.2025.v14i02.008.

[23] Nagadeepa, C.; Dyczek, B.; Mishra, Ar.K.; Бондаренко, В.; Omelianenko, O.; Sokoliuk, K. Last-Mile Delivery Innovations: The Future of E-Commerce Logistics. In Studies in systems, decision and control; Springer International Publishing, 2024; p. 283.

[24] Hamdan, A.; Hamdan, S.; Benbitour, M.H.; Jradi, S. On the Fair Scheduling of Truck Drivers in Delivery Companies: Balancing Fairness and Profit. Central European Journal of Operations Research 2024, doi:10.1007/s10100-023-00899-5.