Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/BA08.20250090
Published: December 20, 2025

A Comparative Study of Genetic Algorithm and Multi-Agent Dueling DQN for a Complex Deterministic VRP

Mahamudul Hassan Siddique, Minhaz Ahmed Likhon and Md. Ahabab Ul Haq
Department of Industrial and Production Engineering
Bangladesh University of Engineering and Technology
Dhaka, Bangladesh
hassansiddique632@gmail.com, minhazahmedlikhon@gmail.com,
ahababbuetipe20@gmail.com

Abstract

The Vehicle Routing Problem (VRP) has traditionally served as a significant problem in various fields and got more complicated as the time goes on. This paper considers a complex VRP by combining capacitated time-window time-deliveries on the problem of large-scale distribution of medicine in a megacity. Genetic Algorithms (GA) and Multi-Agent Dueling Deep Q-networks (DQN) are compared as typical representatives of metaheuristic method and reinforcement learning method. Demand coverage, distance, time and energy cost are considered as assessment of performance. The comparison brings forward the robustness of the evolutionary and the learning-based approaches, which provide both methodological and empirical frameworks of optimizing logistics in high-dimensional VRP logistics.

Keywords
Vehicle Routing Problem, Genetic Algorithm, Reinforcement Learning.

1. Introduction
Vehicle routing problem (VRP) is a well-known optimization problem in the sphere of supply chain issues. As the COVID-19 infection proved the fragility of the medical supply chains, the necessity to optimize them has become more significant. Numerous studies have been conducted on this very issue with numerous metaheuristic algorithms but there has been a little disregard in applying the concept of reinforced learning even though they can be considered as a promising candidate.

1.1 Objectives
In our study, it is desired to locate the venue in densely populated cities with prime objective with its medicine distribution and further optimize it using RL and compare it against another metaheuristic algorithm like genetic algorithm. There has been a scarcity of application of RL in the medical sector(Sar & Ghadimi, 2023) and the accuracy of RL compared with GA since RL has shown promises in the aspects of route optimization(Kommey et al., 2024).

2. Literature Review
From early 21st century, hybrid genetic algorithms and parallel metaheuristics were shown to be effective for Time Windows VRPs, improving both solution quality and speed on benchmark instances (Berger & Barkaoui, 2004). Recently, there have been instances with machine learning with metaheuristics. For example, Zhao et al (2024) (Qi et al., 2024) introduced a joint approach that couples a Genetic Algorithm with a Graph Convolutional Network (GCN) for VRP, using the GCN to guide search and enable near real time decisions. Another interesting work has been done in the form of Genetic Algorithm Neural Cost Predictor (GANCP) based on graphical neural networks multi head

attention which can streamline search in problems such as multi depot and capacitated VRP (Sobhanan et al., 2024). It shows that domain heuristics and learned guidance can be combined to increase solution quality while keeping budgets moderate.

Recent advancements in RL makes it a strong alternative for heuristics. Nazari et al. (2018) (Nazari et al., n.d.) framed VRP as a sequential decision problem and trained an attention-based policy that generalizes across instances without problem specific tweaks, demonstrating competitive performance against classical heuristics. Subsequent work advanced decentralized multi-agent reinforcement learning (MARL) for cooperative routing, enabling scalable coordination without a central controller (Ngu et al., n.d.) and also Graph-driven deep reinforcement learning (GDRL) approach for pickup and delivery related VRP (Yan et al., 2025). Now for the healthcare industry, RL has been used for to optimize routes under operational uncertainty (Maheshwari et al., 2023) and to control multi compartment cold chain distribution with temperature and capacity constraints (Hu & Wang, 2025). Also, RL becomes a catalyst for better solution for other heuristic methods. It has been in Evolutionary Algorithm with Reinforcement Learning Initialization (EARLI) to solve NP hard VRP problem (Greenberg et al., n.d.).

There has been some work on the benchmark of machine learning algorithms and metaheuristic algorithm for solving VRP problem but more work needs to done on this matter. It is noticed that the GA is more effective than the Epsilon-Greedy Q-Learning Algorithm (EQLA) in solving the Traveling Salesman Problem (TSP), providing faster and more optimal solutions. However, GA may struggle with larger city numbers, and hybridizing GA and EQLA could lead to cost-effective optimal solutions (Uthayasuriyan et al., 2023).

To target the healthcare industry's VRP, specific constraints arise such as cold chain integrity, patient/client priority, carbon aware operations etc. There has been some work regarding the pharmaceutical supply chains for example solving with GA coupled with neighbourhood search, systemic comparison of algorithmic families and offering guidance on method selection under practical constraints (Li et al., 2024; Mamoun et al., 2024). Related research focuses on improving the vehicle routing problem for emergency logistics in the context of significant epidemics, particularly tackling the difficulties brought about by the COVID-19 pandemic by ensuring the prompt delivery of medical supplies to hospitals (Tan et al., 2023).

2.1 Contextual Challenges: Mega Cities

In the context of the medicine distribution in mega cities, significant work needs to done both in choosing proper constraints and mathematical modelling. It is imperative for the medicine to arrive on time, in proper time windows, under properly maintained cold chains and also reducing the environmental impact. The dynamic conditions of a mega city require the combination of time window delivery, capacitated and split delivery VRP. The reviewed studies suggest that (i) hierarchical modelling for multi depot and location routing structures (Sobhanan et al., 2024), (ii) carbon- and priority aware objective design (Li et al., 2024), and (iii) cold chain control under multi compartment constraints (Hu & Wang, 2025) are directly transferable to a mega city's operating environment. Also, this provides a wonderful opportunity to determine the performance of RL and GA of tackling such problems. Since there is a probability that GA might struggle against complex scenarios of a city and also to how RL fairs against it.

3. Methods

3.1 Problem Formulation

The VRP was modelled as an undirected graph $G = (V, E)$, with depots $D1$, $D2$ and customer nodes C. Each edge (i, j) is associated with distance d_{ij}, time t_{ij}, and energy cost e_{ij}. Vehicle k has capacities $Q_k \in \{10, 15, 20\}$, and customer demands qi must be satisfied within time windows $[ai, bi]$. Split deliveries and transit-node usage were permitted. Objectives were to maximize total demand D' minimize distance, time, and energy lexicographically (Figure 1):

$$\max D', \min \left\{ \sum_{(i,j)\in E} d_{ij}x_{ij}, \sum t_{ij}x_{ij}, \sum e_{ij}x_{ij} \right\} \quad (1)$$

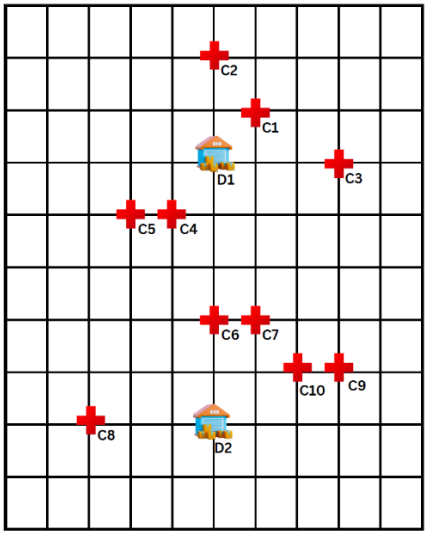


Figure 1. Study Area with Designated Zones (C1–C10 and D1–D2)

3.2 Genetic Algorithm Implementation

An initial population was generated randomly and heuristically. Selection was performed via tournament sampling. Partially Matched Crossover (PMX) was applied to exchange sub lists while maintaining feasibility. Mutation operations included swaps, insertions, and segment cuts to allow re-visitation and transit-node usage. The fitness function was defined as

$$F(X) = \alpha D'(X) - \beta \sum_{(i,j)\in r_k} \left(d_{ij} + t_{ij} + e_{ij}\right) \tag{2}$$

prioritizing demand fulfilment while penalizing distance, time, and energy. Iterative reproduction, crossover, and mutation were used to converge towards routes optimizing demand coverage and efficiency.

3.3 Multi-Agent Dueling DQN Implementation

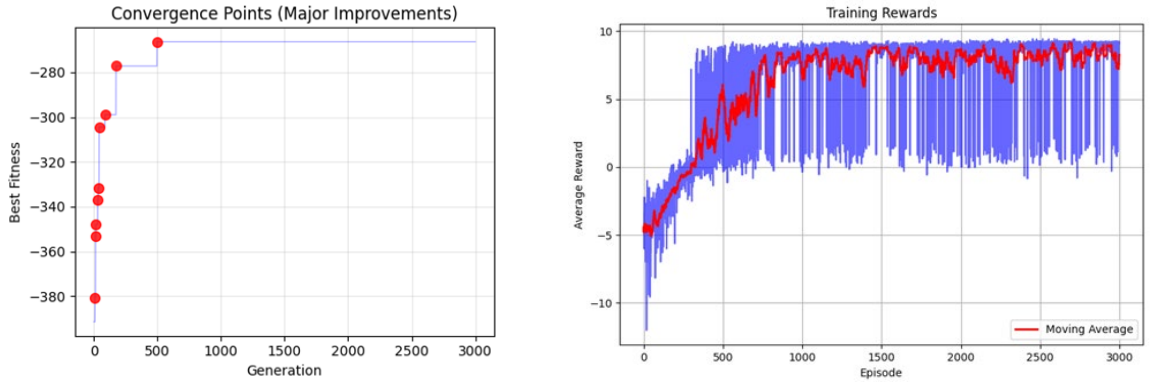
Each vehicle was modelled as an agent interacting in a common environment. The state included current location, remaining capacity ratio, normalized elapsed time, and demand fulfilment ratio. The action space comprised visiting depots or customers, serving, revisiting, or skipping infeasible actions. Invalid actions were masked to ensure feasibility. A duelling DQN architecture separated state value $V(s)$ and action advantages $A(s, a)$, recombined as

$$Q(s,a) = V(s) + (A(s,a) - \frac{1}{|A|}\sum_{a'} A(s,a')) \tag{3}$$

reducing variance and improving decision quality. Rewards were provided for demand fulfilment, efficiency, and depot visits for refill, while penalties were applied for delays, detours, or infeasible actions and visiting same location etc.

4. Results and Discussion

This section presents a case-by-case comparative analysis of the performance of a Genetic Algorithm (GA) and a Multi-Agent Dueling Deep Q-Network (RL) in addressing a complex Vehicle Routing Problem (VRP). The problem incorporates capacitated vehicles, strict time windows, and split deliveries across two depots. Performance is evaluated under six distinct scenarios, where the number of vehicles per depot (one or two) and their respective capacities (10, 15, or 20) are systematically varied. The quality of the generated routes is measured using normalized metrics: the Distance-to-Demand Ratio (D Ratio), the Time to-Demand Ratio (T Ratio), and the Cost-to-Demand Ratio (C Ratio). Lower values of these ratios indicate greater efficiency, signifying reduced travel distance, shorter time, or lower costs per unit of demand fulfilled. The following analysis discusses the results of each configuration, highlighting percentage differences in performance and relating the findings to efficiency benchmarks drawn from radar plot analysis.



(a) Best Fitness in Each Generation (b) Average Reward in Each Episode

Figure 2. Genetic Algorithm Best Fitness and MADDQN Reward Achievements

Here, Figure 2(a) illustrates how the best fitness of the population is improved across generations in the Genetic Algorithm. Fig.2(b) depicts how the agents in Reinforcement Learning increase their average reward over successive episodes (Table 1).

Table 1. Comparative Results of GA vs RL Algorithms

Case(Depot Config)	Demand Completion	Algorithm	D Ratio	T Ratio	C Ratio
(10,10)	106 95	**GA** RL	**0.849** 0.758	**2.123** 1.895	**1.724** 1.538
(15,15)	112 112	**GA** RL	**0.493** 0.654	**1.232** 1.634	**1.001** 1.327
(20,20)	112 112	**GA** RL	**0.536** 0.632	**1.339** 1.580	**1.087** 1.283
(10,10,10,10)	112 112	**GA** RL	**0.771** 1.018	**1.929** 2.545	**1.566** 2.066
(15,15,15,15)	112 112	**GA** RL	**0.514** 0.707	**1.286** 1.768	**1.044** 1.435
(20,20,20,20)	112 112	**GA** RL	**0.525** 0.943	**1.313** 2.357	**1.066** 1.914

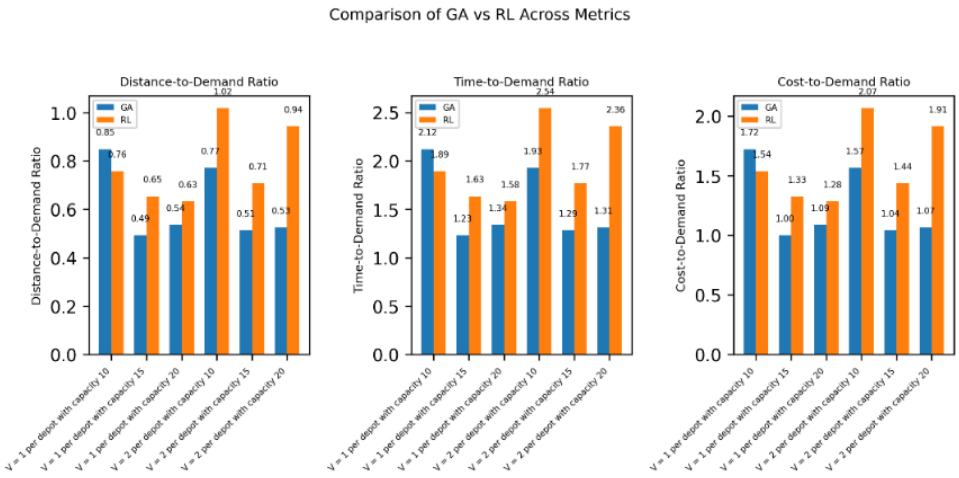


Figure 3. Ratio metrics for Each Case Genetic and MADDQN(RL)

In Figure 3 illustrates the demand-to-distance ratio, demand-to-time ratio, and demand-to-cost ratio for each case across both algorithms. A lower ratio is indicative of greater minimization of time, cost, and distance.

In the first scenario, where each depot had a single vehicle with capacity 10, the GA fulfilled 106 units of demand, achieving a D Ratio of 0.849, a T Ratio of 2.123, and a C Ratio of 1.724. In contrast, RL fulfilled only 95 units of demand but was more efficient in terms of ratios, recording a D Ratio of 0.758, a T Ratio of 1.895, and a C Ratio of 1.539. This represents a 10.7% improvement in distance efficiency for RL compared to GA, placing RL's D Ratio close to the efficient benchmark of 0.75 observed for two-vehicle configurations. However, this efficiency came at the expense of demand fulfilment, as RL failed to deliver 11 units—equivalent to a 10.4% shortfall. This suggests that RL prioritized shorter routes over complete coverage, struggling with strict time-window constraints, whereas GA identified a longer but comprehensive solution.

In the second scenario, with one vehicle of capacity 15 per depot, both GA and RL fulfilled the total demand of 112 units. However, GA was substantially more efficient. Its D Ratio was 0.493 compared to RL's 0.654, representing a 24.6% lower value. Similarly, GA's T Ratio of 1.232 and C Ratio of 1.001 were 24.6% lower than RL's respective values of 1.634 and 1.327. Notably, GA's T Ratio was close to the highly efficient benchmark of 1.03, underscoring the near-optimal quality of its solution. RL, while delivering all demand, demonstrated lower efficiency, indicating declining route optimization performance as complexity increased slightly.

In the third scenario, with single vehicles of capacity 20 per depot, both algorithms again fulfilled the demand of 112 units. GA maintained its efficiency advantage, with a D Ratio of 0.536 versus RL's 0.632 (15.2% lower), a T Ratio of 1.339 compared to RL's 1.580 (15.2% lower), and a C Ratio of 1.087 compared to RL's 1.283 (15.3% lower). These results show that although RL managed the higher capacity without compromising demand, it generated slower and more resource-intensive routes, particularly in terms of time. GA continued to balance efficiency across all metrics through global optimization.

When expanded to two vehicles per depot with capacity 10, both GA and RL fulfilled the 112 units of demand, but the efficiency gap widened considerably. GA achieved a D Ratio of 0.771, a T Ratio of 1.929, and a C Ratio of 1.566, while RL recorded much poorer results: a D Ratio of 1.018, a T Ratio of 2.545, and a C Ratio of 2.066. This indicates that GA outperformed RL by 24.3% in distance, 24.2% in time, and 24.2% in cost. GA's D Ratio was nearly identical to the efficient benchmark for a four-vehicle configuration, confirming the global quality of its solution. RL, however,

produced outcomes comparable to poor-performing cases in the radar plot (e.g., V-6, V-7, V-9), indicating a breakdown in multi-agent coordination that resulted in redundant routing and inefficiencies.

The fifth scenario, Involving two vehicles per depot with a capacity of 15, further confirmed GA's superiority. Both algorithms fulfilled the demand of 112 units, but GA achieved a D Ratio of 0.514, a T Ratio of 1.286, and a C Ratio of 1.044, while RL recorded 0.707, 1.768, and 1.435, respectively. This reflects a consistent 27.3% improvement across all three metrics for GA. The uniformity of this advantage highlights GA's holistic efficiency in coordinating multiple vehicles, while RL continued to exhibit coordination problems that inflated resource consumption.

The sixth and final scenario, with two vehicles per depot each of capacity 20, produced the largest disparity. Both GA and RL delivered 112 units, but GA's D Ratio was 0.525 compared to RL's 0.943 (44.3% lower). Similarly, GA's T Ratio of 1.313 and C Ratio of 1.066 were 44.3% lower than RL's respective values of 2.357 and 1.914. RL's ratios were worse than benchmarks for larger fleets, with its T Ratio suggesting poorly scheduled, time-window-violating routes. By contrast, GA demonstrated remarkable efficiency, effectively coordinating the four high-capacity vehicles and minimizing overall effort.

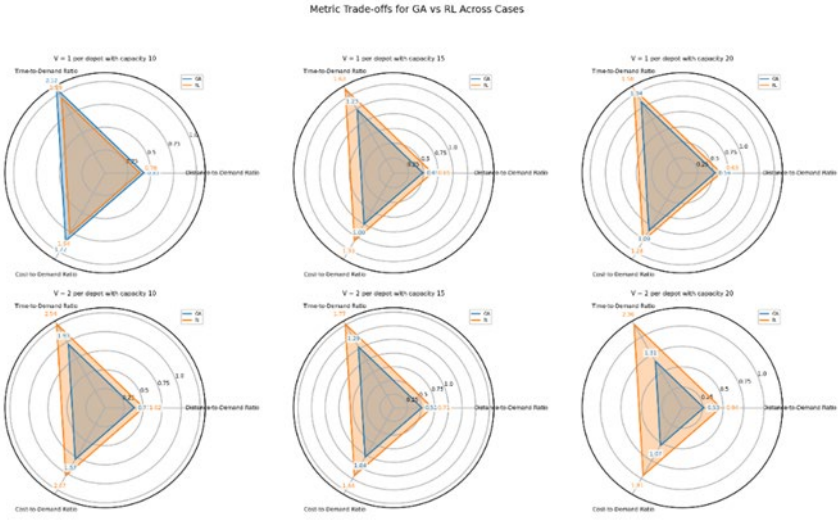


Figure 4. Radar Plot for Genetic and MADDQN(RL)

Figure 4 presents radar plots that demonstrate the performance differences between the two algorithms based on the demand-to-distance ratio, demand-to-time ratio, and demand-to-cost ratio. The blue shaded region represents the Genetic Algorithm, while the orange shaded region represents the Multi-Agent Dueling Deep Q-Network. A smaller shaded region indicates a greater minimization of cost, time, and distance in the optimal solution.

By contrast, the reinforcement learning approach, even with a Multi-Agent Dueling DQN framework, exhibited fundamental weaknesses. The vast state action space of the VRP posed severe challenges of combinatorial complexity and non-stationarity, especially in multi-agent settings where each agent perceives an evolving environment due to the actions of others. Reward function design was difficult, as balancing trade-offs between distance, time, cost, and demand often led to suboptimal strategies such as unmet demand improving efficiency ratios, as in Case 1. Furthermore, RL was highly sensitive to hyperparameters (learning rate, discount factor, architecture, exploration strategy), requiring extensive fine-tuning. Finally, the exploration–exploitation trade-off in such a large action space often caused premature convergence to mediocre policies, preventing the discovery of coordinated strategies. GA's diversified population-based approach avoided these pitfalls.

In conclusion, the Genetic Algorithm emerged as the more effective, efficient, and reliable solution to the VRP. It consistently matched or outperformed RL, with performance gaps reaching up to 44.3% in key efficiency metrics. Importantly, GA always fulfilled demand completely, whereas RL's apparent efficiency in the simplest case (10, 10) resulted from leaving demand unmet. RL demonstrated fragility, performing reasonably only in low-complexity cases but collapsing in multi-vehicle scenarios due to coordination challenges in large combinatorial spaces. For problems of this scale and complexity, population-based optimization methods such as GA represent a more robust and superior approach. Future research could explore enhancements to RL frameworks to address nonstationary and improve exploration, but in their current form, GA remains the more practical and effective solution.

5. Limitations and Future work

The limited scope of the research is that it is applied to a deterministic vehicle routing problem (VRP) setting and thus does not account for real-world uncertainties, including but not limited to traffic variations, demand variability, and disruptions. Reinforcement learning with manually designed reward functions was highly sensitive to hyperparameters and lacked adaptability and coordination in multi-agent applications. Furthermore, the genetic algorithm (GA) and one reinforcement learning model (Multi-Agent Dueling DQN) were considered, with no attempt made to investigate hybrid or more advanced algorithms. Future studies may include stochastic and dynamic conditions, explore hybrid GA–RL or graph-based multi-agent approaches, and take advantage of adaptive reward-shaping and coordination to make them more robust, scalable, and realistic.

6. Conclusion

GA was observed to be the more effective and reliable solution, outperforming RL in five out of six scenarios, with the advantage increasing as complexity grew. Its metrics consistently matched radar plot benchmarks. RL performed adequately only in a single-vehicle, low-capacity scenario but failed to handle multiagent coordination and combinatorial complexity. For practical large-scale VRPs, population-based methods like GA provide a more robust and efficient solution. Future work may investigate advanced multi-agent RL frameworks, such as centralized critics or curriculum learning, to better manage complex state representations.

Nomenclature. V = Locations; E = Edges; D = Depot; C = Customer Node; Q = Capacity set of vehicle k; α = weighting factor of demand; β = weighting factor of distance, time & energy cost

References

Berger, J., and Barkaoui, M., "A Parallel Hybrid Genetic Algorithm for the Vehicle Routing Problem with Time Windows," *Computers and Operations Research*, Vol. 31, No. 12, pp. 2037–2053, **2004**, doi: 10.1016/S0305-0548(03)00163-1.

Greenberg, I., Sielski, P., Linsenmaier, H., Gandham, R., Mannor, S., Fender, A., Chechik, G., and Meirom, E., "Accelerating Vehicle Routing via AI-Initialized Genetic Algorithms," *Working Paper / Preprint*, **n.d.**

Hu, J., and Wang, C., "A Deep Reinforcement-Learning-Based Route Optimization Model for Multi-Compartment Cold Chain Distribution," *Mathematics*, Vol. 13, No. 13, **2025**, doi: 10.3390/math13132039.

Kommey, B., Isaac, O. J., Tamakloe, E., and Opoku, D., "A Reinforcement Learning Review: Past Acts, Present Facts and Future Prospects," *Journal of Research and Development (ITJRD)*, Vol. 8, No. 2, **2024**, doi: 10.25299/itjrd.2024.13474.

Li, J., Peng, K., Deng, X., Wang, J., and Liu, A., "Model and Algorithm for Pharmaceutical Distribution Routing Problem Considering Customer Priority and Carbon Emissions," *Data-Centric Engineering*, Vol. 5, **2024**, doi: 10.1017/dce.2024.13.

Maheshwari, S., Jain, P. K., and Kotecha, K., "Route Optimization of Mobile Medical Unit with Reinforcement Learning," *Sustainability*, Vol. 15, No. 5, **2023**, doi: 10.3390/su15053937.

Mamoun, K. A., Hammadi, L., El Ballouti, A., Novaes, A. G. N., and De Cursi, E. S., "Vehicle Routing Optimization Algorithms for Pharmaceutical Supply Chain: A Systematic Comparison," *Transport and Telecommunication*, Vol. 25, No. 2, pp. 161–173, **2024**, doi: 10.2478/ttj-2024-0012.

Nazari, M., Oroojlooy, A., Takáč, M., and Snyder, L. V., "Reinforcement Learning for Solving the Vehicle Routing Problem," *Preprint*, **n.d.**

Ngu, E., Parada, L., Javier, J., Macias, E., and Angeloudis, P., "A Decentralised Multi-Agent Reinforcement Learning Approach for the Same-Day Delivery Problem," *Preprint*, **n.d.**

Qi, D., Zhao, Y., Wang, Z., Wang, W., Pi, L., and Li, L., "Joint Approach for Vehicle Routing Problems Based on Genetic Algorithm and Graph Convolutional Network," *Mathematics*, Vol. 12, No. 19, **2024**, doi: 10.3390/math12193144.

Sar, K., and Ghadimi, P., "A Systematic Literature Review of the Vehicle Routing Problem in Reverse Logistics Operations," *Computers and Industrial Engineering*, Vol. 177, Article 109011, **2023**, doi: 10.1016/j.cie.2023.109011.

Sobhanan, A., Park, J., Park, J., and Kwon, C., "Genetic Algorithms with Neural Cost Predictor for Solving Hierarchical Vehicle Routing Problems," *arXiv Preprint*, **2024**, http://arxiv.org/abs/2310.14157.

Tan, K., Liu, W., Xu, F., and Li, C., "Optimization Model and Algorithm of Logistics Vehicle Routing Problem under Major Emergency," *Mathematics*, Vol. 11, No. 5, **2023**, doi: 10.3390/math11051274.

Uthayasuriyan, A., G. H. C., K., U. V., Mahitha, S. H., and G., J., "A Comparative Study on Genetic Algorithm and Reinforcement Learning to Solve the Traveling Salesman Problem," *Research Reports on Computer Science*, pp. 1–12, **2023**, doi: 10.37256/rrcs.2320232642.

Yan, D., Guan, Q., Ou, B., Yan, B., and Cao, H., "Graph-Driven Deep Reinforcement Learning for Vehicle Routing Problems with Pickup and Delivery," *Applied Sciences*, Vol. 15, No. 9, **2025**, doi: 10.3390/app15094776.

Biographies

Mahamudul Hassan Siddique is an undergraduate senior year student of Department of Industrial and Production Engineering at BUET. He is studying the multi-tier supply chain optimization under the reinforcement learning in his thesis. His areas of research concern machine learning, deep learning, natural language processing, large language models, operations research, and optimization. He integrates the use of statistics and the AI techniques to come up with data-driven answers to complicated industrial challenges. His research is based on the interplay of conventional optimization with state-of-the-art computational models in aid of intelligent decision making.

Minhaz Ahmed Likhon is an undergraduate senior year student of the Department of Industrial and Production Engineering at BUET. He is working on his thesis in the area of multi-tier supply chain optimization using reinforcement learning and graph neural networks. Alongside his thesis, he has participated in a global supply chain case competition, where he explored innovative solutions for digital transformation and efficiency improvement. He has also carried out a project on a smart industrial system, published on ResearchGate. His research interests include supply chain management, operations research, optimization, and artificial intelligence, with a focus on applying data-driven techniques to solve complex industrial challenges.

Md. Ahabab Ul Haq is studying in his senior years as an undergrad student of Industrial and Production Engineering at Bangladesh University of Engineering and Technology (BUET). He used to be an active member of BUET Entrepreneurship Development Club as the Logistics Director and had an active participation in arranging programs like Projonmowave and EDC Talk and currently an active member of the executive panel of the Association of Industrial & Production Engineers (AIPE) BUET as the Design Secretary. He secured 70% scholarship in ISCEA Prize Global Case Competition 2024. He also recently debuted in ResearchGate, along with his team members, through his project of IoT-Based Smart Conveyor System. Currently his thesis involves the environmental impact of manufacturing practices and the optimization of manufacturing equipment and processes.