

# Development and evaluation of an adaptive routing algorithm for C2C logistics\*

Danylo Kovalenko<sup>1,†</sup> and Iryna Zamrii<sup>1,\*†</sup>

<sup>1</sup> State University of Information and Communication Technologies, Solomyanska Street 7, 03110, Kyiv, Ukraine

## Abstract

Modern C2C logistics systems face challenges related to demand variability, limited resources, and the need for adaptive routing. This study presents an approach to developing an adaptive routing methodology based on the combination of reinforcement learning (RL) and machine learning (ML) for demand prediction.

In the first stage of the research, an algorithm was developed that utilizes RL agents for dynamically determining optimal routes and LSTM networks for demand forecasting. Simulation testing demonstrated improvements in key metrics, including reduced delivery time and increased resource utilization efficiency.

Currently, an experimental validation of the algorithm in real-world conditions is being conducted, with its results to be used for formalizing the methodology. The obtained data is expected to contribute to the development of a universal adaptive routing methodology, enhancing the flexibility and efficiency of C2C logistics.

## Keywords

C2C logistics, adaptive routing, reinforcement learning, demand forecasting, multi-agent systems, last-mile logistics, route optimization, digital twins, cloud computing.

## 1. Introduction

C2C (Customer-to-Customer) logistics plays a crucial role in modern e-commerce by enabling fast and convenient deliveries between private individuals. Unlike traditional B2C logistics, where the delivery process is centralized, C2C models are characterized by high dynamism, uneven resource distribution, and fluctuating demand [1]. As a result, routing problems arise that cannot be effectively solved using static methods, since the system's state changes rapidly. Studies show that classical shortest-path algorithms, such as Dijkstra's algorithm or A\*, perform well in fixed road networks but fail to efficiently adapt routes to changes in traffic and demand distribution [2].

One approach to overcoming these limitations is the use of Reinforcement Learning (RL), which enables agents to make real-time decisions based on historical and current system data. RL is applied to routing problems where decisions need to be adapted to changing environmental conditions, such as traffic congestion, adverse weather, or demand fluctuations [3]. In [4], it was demonstrated that Q-learning-based algorithms provide more optimal routes for dynamic logistics systems compared to traditional methods. Additionally, it was found that deep neural networks (Deep Q-Network, DQN) significantly enhance the adaptability of RL models, but their effectiveness depends on data quality and the accuracy of future state predictions.

A critical component of successful RL application is the ability to forecast future system load. The use of deep neural networks, such as Long Short-Term Memory (LSTM), enables the

---

*The Second International Conference of Young Scientists on Artificial Intelligence for Sustainable Development (YAISD), May 8-9, 2025, Ternopil, Ukraine*

\* Corresponding author.

† These authors contributed equally.

✉ d.kovalenko@stud.duikt.edu.ua (D. Kovalenko); i.zamrii@duikt.edu.ua (I. Zamrii)

ORCID 0009-0002-2475-2883 (D. Kovalenko); 0000-0001-5681-1871 (I. Zamrii)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

incorporation of temporal dependencies in input data, allowing for the prediction of new order placements in different locations [5]. The study in [6] demonstrated that combining demand forecasting with routing algorithms can significantly improve the efficiency of logistics systems by reducing the number of "empty" trips and optimizing courier workload. This is particularly relevant for urban logistics platforms, where demand distribution can change rapidly.

This research focuses on developing an adaptive routing methodology for C2C logistics by integrating reinforcement learning and demand forecasting. At the first stage, an algorithm was designed to dynamically determine routes based on environmental variables. Its effectiveness was evaluated in a simulation environment, where it showed a significant reduction in deviation from the optimal route, as well as improvements in delivery time and transport network utilization [7]. However, it was also found that the algorithm incurs higher computational costs, which could be a limiting factor when scaling to real-world logistics platforms.

Currently, the algorithm is undergoing real-world testing to assess its performance when working with incomplete or noisy input data, as well as its ability to adapt to dynamic transport conditions. The results obtained are expected to refine the model and formulate a generalized methodology suitable for implementation in real logistics platforms [8].

## 2. Research Motivation and Problem Statement

With the growth of C2C logistics, traditional routing methods are losing their effectiveness due to the dynamic nature of demand, unpredictability of traffic conditions, and limited resources. Algorithms such as Dijkstra's algorithm or A\* perform well in environments with static road networks but fail to account for real-time changes in the transportation system [1]. In urban logistics, where congestion, road closures, and fluctuating delivery requests are common, classical methods cannot quickly adapt routes, leading to increased order fulfillment time and higher logistics costs [2].

One promising approach to addressing this challenge is the use of Reinforcement Learning (RL), which enables dynamic delivery routing by adapting to environmental changes. RL agents learn from historical and real-time data, leveraging a Markov Decision Process (MDP) model, where each system state depends on previous actions [3]. Studies show that Deep Q-Networks (DQN) improve routing efficiency by implementing an adaptive strategy for selecting optimal routes that considers real-time traffic conditions [4]. However, RL performance heavily depends on the accuracy of input data, particularly in forecasting future demand and courier distribution across the city.

To address this challenge, deep neural networks such as Long Short-Term Memory (LSTM) are employed, which can analyze temporal dependencies and predict logistics system fluctuations [5]. The combination of RL and LSTM enables not only real-time adaptation but also the anticipation of future delays and demand surges, which is crucial for stabilizing logistics operations. Research in this field indicates that integrating predictive models reduces overall system load, balances order distribution, and minimizes courier downtime [6].

As part of this study, a system combining RL agents for optimal decision-making and LSTM networks for demand forecasting was developed. Initial testing in a simulation environment demonstrated that the adaptive approach significantly reduces route deviation compared to traditional methods. Depending on the scenario, deviations ranged from 9.2% to 3.3%, whereas for classical algorithms, this metric reached 31.7% in heuristic-based methods. At the same time, the adaptive algorithm required higher computational resources, with an average route computation time of 50–60 ms, which was 2–3 times higher than that of traditional methods. Additionally, an increased frequency of route recalculations was observed, reaching 18 updates in the most complex scenarios, potentially leading to computational system overload.

The next phase of the research involves validating the algorithm in real-world conditions, allowing for an assessment of its performance when handling incomplete or noisy data and evaluating its robustness against sudden changes in the logistics system. The results of real-world experiments are expected to refine the model and establish a generalized adaptive routing

methodology for C2C logistics, making it suitable for scalability and integration into large logistics platforms.

### 3. Developed Algorithm and Its Simulation-Based Evaluation

The proposed routing algorithm is designed to enable adaptive route management for couriers in C2C logistics. The primary goal is to develop a system capable of dynamically adjusting routes in response to changes in demand and the state of the transportation network. To achieve this, the algorithm integrates Reinforcement Learning (RL), which allows agents to make optimal decisions based on both current and anticipated changes in the environment, and Long Short-Term Memory (LSTM) for predicting the future distribution of orders [1].

The overall structure of the algorithm consists of two key components:

- demand forecasting – utilizing LSTM to estimate the probability of new order appearances in different city zones. This enables proactive courier redistribution, preventing overload or idle time [2];
- reinforcement learning-based routing – an RL agent is trained to select optimal routes considering not only the current network state but also predicted changes [3].

This approach minimizes delivery delays, improves courier workload balance, and reduces overall route length. Previous studies have shown that RL-based traffic flow management can reduce average delivery time by 15–20% compared to classical routing methods [4]. Additionally, LSTM-based demand forecasting decreases the number of empty trips and contributes to a more balanced load distribution across the transportation system [5].

Expected Benefits of the Algorithm:

- flexibility – the ability to adapt to changes in the transportation environment in real time;
- resource optimization – balancing workloads among couriers and reducing idle time;
- reduction in delivery time – through dynamic route adjustments.

At the same time, the use of RL-based methods may lead to increased computational costs, as the algorithm requires significant resources for training and decision-making. This issue has been highlighted in real-time route optimization studies, emphasizing the need to balance prediction accuracy and computation speed [6, 7]. Another critical aspect is algorithm scalability as the number of delivery requests increases, which requires further investigation [8].

Thus, the proposed approach aims to strike a balance between routing efficiency, computational speed, and adaptability. The next subsection presents its mathematical model and formal algorithmic principles.

#### 3.1. Mathematical Model of the Algorithm

The proposed routing algorithm integrates Reinforcement Learning (RL) for adaptive route management and Long Short-Term Memory (LSTM) for demand forecasting. Its operation can be formalized as a Markov Decision Process (MDP), where each action affects the future state of the system and the agent's reward [1].

The algorithm operates in two stages:

1. Demand forecasting using LSTM – predicts the future distribution of orders, enabling proactive route planning [2].
2. Adaptive decision-making by the RL agent – optimizes routes in real-time based on updated data and demand forecasts [3].

##### 3.1.1. Formalization of the Learning Process

The system is modeled as an MDP (Markov Decision Process), defined as a five-tuple  $\langle \mathbf{s}, \mathbf{a}, \mathbf{P}, \mathbf{R}, \gamma \rangle$ , where:

- $\mathbf{s}$  – the set of states, including courier locations, active orders, and the current state of the transportation network;
- $\mathbf{a}$  – the set of actions available to the agent (e.g., assigning an order to a courier or modifying a route);
- $\mathbf{P}(\mathbf{s}' | \mathbf{s}, \mathbf{a})$  – the probability of transitioning between states based on the selected action;
- $\mathbf{R}(\mathbf{s}, \mathbf{a})$  – the reward function that determines the efficiency of the chosen route;
- $\gamma$  – the discount factor that regulates the long-term optimization of decision.

At each step, the agent receives a reward  $G_t$  which defines the effectiveness of the routing process:

$$G_t = \mathbb{E} \left[ \sum_{k=0}^{\infty} \gamma^k R(\mathbf{s}_{t+k}, \mathbf{a}_{t+k}) \right]. \quad (1)$$

The Q-function is used to estimate the quality of an action  $\mathbf{a}$  in state  $\mathbf{s}$ . The Q-value update follows the rule:

$$Q(\mathbf{s}, \mathbf{a}) \leftarrow Q(\mathbf{s}, \mathbf{a}) + \eta [R(\mathbf{s}, \mathbf{a}) + \gamma \max_{\mathbf{a}'} Q(\mathbf{s}', \mathbf{a}') - Q(\mathbf{s}, \mathbf{a})], \quad (2)$$

where  $\eta$  – is the learning rate, and  $\max_{\mathbf{a}'} Q(\mathbf{s}', \mathbf{a}')$  represents the best expected value for the new state  $\mathbf{s}'$ .

### 3.1.2. Demand Forecasting

An LSTM network is used to predict future demand based on historical data. This allows the identification of potentially overloaded regions, enabling route adjustments before imbalances occur [5].

The model uses the parameters  $\theta_{\text{LSTM}} = \{W_i, U_i, b_i, W_f, U_f, b_f, W_g, U_g, b_g, W_o, U_o, b_o\}$  which determine the behavior of the LSTM. The state update in the LSTM model is formalized as follows:

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + b_i), \quad (3)$$

$$f_t = \sigma(W_f X_t + U_f h_{t-1} + b_f), \quad (4)$$

$$o_t = \sigma(W_o X_t + U_o h_{t-1} + b_o), \quad (5)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh(W_g X_t + U_g h_{t-1} + b_g), \quad (6)$$

$$h_t = o_t \circ \tanh(c_t), \quad (7)$$

where  $i_t$ ,  $f_t$ ,  $o_t$  are sigmoid activation functions that regulate the flow of information between memory states [6]. Where  $i_t$  is the input gate and  $f_t$  is the descending gate, which depend on the activation function  $\sigma$ . The memory state is updated according to formula (10), where  $\circ$  is means element-wise multiplication. The output gate is calculated according to formula (5), and the hidden state is updated according to (7).

### 3.1.3. Optimization of the Learning Process

To evaluate forecasting accuracy, the loss function is computed as the mean squared error between the predicted values  $\mathbf{P}$  and the actual data  $\mathbf{y}$  [7]:

$$L_{\theta} = \frac{1}{N} \sum_{i=1}^N \|y_i - P_i\|^2. \quad (8)$$

The model is trained using gradient descent, where parameter updates follow the rule [8]:

$$\theta = \theta - \eta \cdot \nabla_{\theta} L_{\theta}, \quad (9)$$

where  $\eta$  is the learning rate, and  $\nabla_{\theta} L_{\theta}$  is the gradient of the loss function with respect to the parameters.

The training process is repeated for each epoch until the specified number of **epochs** is reached or the loss function stabilizes. After completing all epochs, the optimized parameters are determined as follows [9]:

$$\theta^{\text{final}} = \theta^0 - \eta \sum_{e=1}^{\text{epochs}} \sum_{i=1}^N \nabla_{\theta} L_{\theta}(P_i, y_i). \quad (10)$$

Ultimately, the algorithm generates a demand forecast  $P$ , which is used to update the system state and optimize routing decisions. The proposed approach is expected to reduce route deviations from optimal values, decrease average delivery time, and enhance system adaptability.

### 3.2. Results of Experiments in a Simulation Environment

Evaluating the efficiency of adaptive routing algorithms is a critical step in their validation before deployment in real logistics systems. Previous studies have shown that combining reinforcement learning (RL) with demand forecasting significantly improves the efficiency of urban logistics platforms. In [9], it was noted that RL-based algorithms enable flexible decision-making, which enhances the utilization of transportation resources in urban environments. The study in [10] demonstrated that integrating RL with optimization methods reduces overall delivery delays by 13%.

At the same time, as highlighted in [11], traditional routing methods, such as Vehicle Routing Problem (VRP) algorithms, are significantly less effective in scenarios with highly variable demand. In [12], it was demonstrated that LSTM-based demand forecasting reduces the number of empty trips, leading to increased overall logistics network efficiency.

#### 3.2.1. Testing Methodology

In this study, the developed algorithm was tested in a simulation environment that mimics an urban transportation network. The simulator models dynamic delivery demand, changing road conditions, and courier behavior variability. The primary objective of the testing was to evaluate the algorithm's ability to adapt routes in real time and compare its effectiveness with traditional routing methods. The Python platform with AnyLogic libraries for visualization and calculations was used. Real-world city maps, historical demand data, transportation routes, and traffic served as the basis for the simulations. The test scenarios covered static and dynamic delivery conditions that took into account factors such as traffic, weather conditions, and variable demand. The simulations involved three types of agents: couriers, orders, and vehicles, which could adapt their behavior depending on the state of the environment.

Three different approaches were considered:

1. Traditional heuristic algorithm – constructs routes based on the shortest path without accounting for real-time changes in the transportation system.
2. Vehicle Routing Problem (VRP) optimization – a classical approach to optimizing route distribution among couriers.

3. Adaptive method (LSTM + RL) – the proposed algorithm, which integrates demand forecasting and RL-based adaptive learning for dynamic route adjustment.

The efficiency of the algorithms was assessed based on several key metrics. In particular, the average route length was analyzed, the reference value of which was taken as 12 km, which was obtained from the optimal solution of the VRP (Vehicle Routing Problem) and was consistent with real-world urban logistics data.

The experimental results are presented in Table 1.

**Table 1**  
**Results of Experiments in a Simulation Environment**

Method	Delivery time (s)	Average route computation time (ms)	Number of route recalculations	Route length (km)	Deviation from the reference route (%)
Traditional (Heuristic)	1200	15	2	15.8	31.7%
Traditional (VRP)	1100	35	3	14.2	18.3%
Adaptive (LSTM + RL)	950	50	5	13.1	9.2%

The following metrics were used to assess efficiency: delivery time (s) – the average time required to complete an order; average route computation time (ms) – the speed of route recalculations; number of route recalculations – the frequency of route updates during delivery execution; route length (km) – the total distance traveled by the courier while completing an order; deviation from the reference route (%) – the extent to which the actual route deviates from the theoretically optimal route.

### 3.2.2. Analysis of the Obtained Results

The obtained results demonstrate that the adaptive approach significantly outperforms traditional routing methods across all key performance indicators. In particular, using LSTM for demand forecasting reduced the average delivery time by 20–25% compared to the heuristic method and by 13–15% compared to VRP. Additionally, the lowest deviation from the reference route was achieved—ranging from 9.2% in the initial scenario to 3.3% with a higher number of route updates.

However, as noted in [13], using RL in urban logistics tasks incurs significant computational costs. In our experiment, the average route computation time for the adaptive approach was 50–60 ms, which is 2–3 times higher than that of traditional methods. Furthermore, the number of route recalculations increased, reaching 18 updates in complex scenarios, exceeding the acceptable threshold of 15 updates. This may indicate a risk of excessive route recomputation in high-demand scenarios, which is also confirmed in [14].

These findings highlight the potential of the adaptive approach for real-world urban logistics systems. However, to enable full-scale implementation, route recomputation processes must be optimized, and computational overhead must be reduced, which will be the focus of future research. The next step is experimental testing of the algorithm in real-world conditions, allowing for an assessment of its resilience to unpredictable changes in the urban transportation network.

## 4. Experimental Validation in a Real-World Environment

Following the positive results obtained in the simulation environment, the next stage involved testing the algorithm under real-world urban logistics conditions. Field experiments are crucial, as

real environments introduce additional factors that are difficult to simulate, such as incomplete or noisy data, unpredictable traffic variations, and fluctuating courier behavior [15].

Previous research has shown that reinforcement learning (RL) methods can be effectively applied to real transportation systems, but their performance heavily depends on the quality of input data. In [16], it was demonstrated that dynamic route adaptation in real-time can reduce average delivery time by 10–15%, even in cases of inaccurate demand predictions.

#### 4.1. Experimental Validation Methodology

The real-world trials were conducted using an operational C2C logistics platform in a major city. The algorithm was integrated into the courier management system, where couriers received updated routes through a mobile application. To monitor the algorithm's performance, GPS trackers were used to track courier movements, along with an analytics system that collected real-time delivery execution data.

During the experiment, the algorithm was tested in two modes:

1. Static routing (baseline approach) – routes were generated at the beginning of the workday and remained unchanged.
2. Adaptive routing (LSTM + RL) – the algorithm dynamically adjusted routes based on environmental changes and predicted demand.

The following metrics were used to evaluate performance:

- average delivery time – time from order acceptance to final delivery;
- number of deviations from the planned route – measures how often routes were modified during order execution;
- resource utilization rate – indicates how evenly workloads were distributed among couriers.

#### 4.2. Preliminary Testing Results

It is expected that the results of real-world experiments will follow similar trends observed in simulation-based trials, where the adaptive algorithm demonstrated reduced average delivery time and route deviation. However, as noted in [17], real-world traffic conditions can introduce significant performance variations due to external factors such as weather conditions, traffic accidents, or other unforeseen disruptions

In previous studies [15], it was highlighted that the efficiency of RL-based algorithms in real-world environments depends on model update speed, which can become a critical factor in large-scale systems with high request volumes. In our case, the experiment focuses on balancing routing adaptation accuracy and computational costs, as excessive route recalculations in real-world scenarios may slow down logistics system operations.

The next step after data collection and analysis will be algorithm parameter optimization and the development of a generalized methodology for implementing the approach in scalable C2C logistics systems.

### 5. Formalization of the Methodology

After conducting real-world experimental trials, the next step is to develop a generalized adaptive routing methodology for C2C logistics. The formalization of this methodology will be based on analyzing both simulation and field test results, particularly in terms of demand prediction accuracy, algorithm stability, and computational efficiency.

One of the key challenges is finding the right balance between routing accuracy and computational overhead. High algorithm adaptability, which minimizes deviation from the reference route, is associated with a significant increase in the number of recalculations. This

highlights the need to develop criteria for dynamically adjusting route update frequency, which will reduce computational load without significantly compromising efficiency.

Data quality dependency and sensitivity to training parameters also affect the overall performance of the system. To overcome these challenges, it is advisable to implement cloud computing and sensor networks to collect relevant data in real time.

Further development of the methodology may include the implementation of multi-agent systems for courier coordination, the integration of hybrid forecasting models, and the development of lightweight RL algorithms aimed at small platforms.

The methodology will also account for possible model parameter variations based on request density, urban infrastructure characteristics, and technological constraints of logistics platforms. To address this, a flexible algorithm configuration system will be designed, allowing the model to adapt to specific operational conditions.

The results of this formalization will be used for further implementation of the methodology in real logistics systems, as well as for developing recommendations on scalability and optimization in high-load transportation networks.

## 6. Conclusions

This study explored an adaptive routing approach for C2C logistics based on reinforcement learning (RL) and demand forecasting. The developed algorithm was tested in a simulation environment, allowing an evaluation of its effectiveness compared to traditional methods. The results demonstrated a significant reduction in deviation from the reference route, indicating high adaptability of the approach. However, an increase in computational costs and the number of route recalculations was observed, which could be a critical factor in system scalability.

Currently, the experimental validation of the algorithm in a real-world environment is ongoing. Field trials are expected to assess the actual impact of the algorithm on delivery time, routing stability, and resource utilization efficiency. The obtained data will be used to further optimize the algorithm and formalize a generalized adaptive routing methodology.

Future research will focus on reducing the algorithm's computational costs, improving route update strategies, and testing the approach on various logistics platforms. Additionally, a dynamic parameter tuning mechanism will be developed to adjust the algorithm's configuration based on system load and external factors.

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

- [1] Y. Yan, A. H. Chow, C. P. Ho, Y. H. Kuo, Q. Wu, C. Ying, Reinforcement Learning for Logistics and Supply Chain Management: Methodologies, State of the Art, and Future Opportunities. *Transportation Research Part E: Logistics and Transportation Review* (2022). doi:10.1016/j.tre.2022.102712.
- [2] H. Liu, J. Zhang, Z. Zhou, Y. Dai, L. Qin, A Deep Reinforcement Learning-Based Algorithm for Multi-Objective Agricultural Site Selection and Logistics Optimization Problem. *Applied Sciences* 14 (2024) 8479. doi:10.3390/app14188479.
- [3] H. Cai, P. Xu, X. Tang, G. Lin, Solving the Vehicle Routing Problem with Stochastic Travel Cost Using Deep Reinforcement Learning. *Electronics* 13 (2024) 3242. doi:10.3390/electronics13163242.
- [4] O. I. Akinola, Adaptive location-based routing protocols for dynamic wireless sensor networks in urban cyber-physical systems. *Journal of Engineering Research and Reports*, 26(7) (2024) 424-443. doi:10.9734/jerr/2024/v26i71220.

- [5] M. Gen, L. Lin, Y. Yun, H. Inoue, Recent advances in hybrid priority-based genetic algorithms for logistics and SCM network design. *Computers & Industrial Engineering* (2018). doi:10.1016/j.cie.2018.08.025.
- [6] A. Singh, M. Wiktorsson, J. B. Hauge, Trends In Machine Learning To Solve Problems In Logistics. *Procedia CIRP* (2021). doi:10.1016/j.procir.2021.10.010.
- [7] M. E. Bruni, E. Fadda, S. Fedorov, G. Perboli, A machine learning optimization approach for last-mile delivery and third-party logistics. *Computers & Operations Research* (2023). doi:10.1016/j.cor.2023.106262.
- [8] D. Knoll, M. Prügmeier, G. Reinhart, Predicting Future Inbound Logistics Processes Using Machine Learning. *Procedia CIRP* (2016). doi:10.1016/j.procir.2016.07.078.
- [9] S. Bosse, Self-adaptive Traffic and Logistics Flow Control using Learning Agents and Ubiquitous Sensors. *Procedia Manufacturing* (2020). doi: 10.1016/j.promfg.2020.11.013.
- [10] A. Mehra, S. Saha, V. Raychoudhury, A. Mathur, DeliverAI: Reinforcement Learning Based Distributed Path-Sharing Network for Food Deliveries. *International Joint Conference on Neural Networks (IJCNN)* (2024). doi:10.1109/IJCNN60899.2024.10651403.
- [11] G. Ghiani, G. Laporte, R. Musmanno, *Introduction to Logistics Systems Planning and Control*. John Wiley & Sons, 2004. ISBN 0-470-84916-9.
- [12] V.R. Kannan, K.C. Tan, Just-in-time, total quality management, and supply chain management: understanding their linkages and impact on business performance. *Omega*, 33(2) (2005) 153-162. doi:10.1016/j.omega.2004.03.012.
- [13] H. Zeng, Y. Liu, D. Zhang, K. Han, H. Hu, A Deep Reinforcement Learning Approach for Constrained Online Logistics Route Assignment, *arXiv: Learning* (2021). URL: <https://dblp.uni-trier.de/db/journals/corr/corr2109.html#abs-2109-03467>.
- [14] T. Phiboonbanakit, T. Horanont, V.N. Huynh, T. Supnithi, A Hybrid Reinforcement Learning-Based Model for the Vehicle Routing Problem in Transportation Logistics, *IEEE Access*, vol. 9 (2021), 163325-163347. doi: 10.1109/ACCESS.2021.3131799.
- [15] S. Kontogiannis, A. Paraskevopoulos, Online Vehicle Routing with Pickups and Deliveries under Time-Dependent Travel-Time Constraints (2024). doi: 10.48550/arxiv.2408.06324.
- [16] Y. Wang, Y. Yuan, X. Guan, H. Wang, L. Young, M. Xu, Collaborative Mechanism for Pickup and Delivery Problems with Heterogeneous Vehicles under Time Windows. *Sustainability*, 11(12) (2019) 3492. doi:10.3390/su11123492.
- [17] A. Jaiswal, S. K. Singh, S. Kumar, B. B. Gupta, K. T. Chui, Advanced Evaluation of Propagation Models and Routing Protocols in Vehicular Ad-Hoc Networks, pp. 406–411, Nov. 2024, doi: 10.1109/cyberscitech64112.2024.00070.