Kashf Journal of Multidisciplinary Research
Vol: 02 - Issue 2 (2025)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

COST EFFECTIVE ROUTE OPTIMIZATION FOR DAIRY PRODUCT DELIVERY

Nasir Khan
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan.
Shahbaz Ahmad*
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan.
Salman Raza
Department of Computer Science, National Textile University, Pakistan
Dr. Ahmad Khan
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan.
Muhammad Younas
Faculty of Computer Science & Information Technology, The Superior University Lahore, Pakistan.

****Corresponding author: shahbaz.ahmad.fsd@superior.edu.pk***

Article Info

Abstract

Milk is an essential component of human nutrition, particularly for the growth and cognitive development of children in undernourished populations. In Pakistan, milk production has more than doubled, with the country now the third-largest milk producer in the world. Most milk in Pakistan, however, is consumed in the form of unpackaged loose milk not only making hygiene an issue but logistical as well. An optimized route planner caters to the cost-effective and environmentally friendly delivery of dairy products. Vehicle routing has only played a minor role in large-scale, dimension-driven, and technology-defined approaches in developing traditional solutions. To address these limitations, sophisticated optimization methods like metaheuristics, machine learning, and deep reinforcement learning (DRL), are studied. In this paper, we proposed a DRL-based delivery model that optimizes routes of dairy products in both capacity and time windows constraint using Proximal Policy Optimization (PPO). By integrating vehicle, customer, and global states, the framework makes informed decisions while satisfying constraints including vehicle capacity and delivery deadlines. We analyze performance measurements such as total route length, inter-customer gaps, time constraints, and capacity utilization to assess efficiency. Experimental results demonstrate that the DRL-PPO model significantly outperforms traditional benchmarks such as Deep Q-Network (DQN) and Advantage Actor-Critic (A2C). Specifically, DRL-PPO yielded advantages in several key areas: it achieved reduced route lengths, shorter delivery times, minimized inter-customer gaps, increased vehicle capacity utilization, and higher success rates across varying levels of task complexity. These findings indicate the potential of Deep Reinforcement Learning (DRL)-based optimization methods to effectively address logistical inefficiencies prevalent in Pakistan's dairy industry. The implications of adopting DRL-PPO are substantial, offering cost-effective solutions that can enhance profitability while concurrently reducing operational expenses. By leveraging advanced machine learning techniques, businesses can pave the way toward sustainable logistics systems for perishable goods. This not only boosts overall supply chain efficiency but also ensures better access to high-quality dairy products for consumers. As such, the use of DRL-PPO represents a strategic opportunity to transform logistics practices within the dairy sector, leading to improved outcomes for both producers and consumers alike.

Keywords: *Route optimization, Dairy product delivery, Deep learning, Reinforcement learning*

Introduction

Milk is very important for human diet due to its rich components.(Pereira, 2014). It plays a very important role in the growth and cognitive development of children, particularly in undernourished populations.(Miller et al., 2022).Globally, milk production has grown substantially. Over the past few decades' production of raw milk has increased from 1.9 billion litters in 2014 to 2.8 billion in 2021. Countries like India, the United States, Pakistan, China, and Brazil are the top milk producers, with the European Union leading in cow milk production (Daneshmand & Shahidi, 2023).Pakistan is the third-largest milk producer country which produce 57 million tons milk in a year (Sattar, 2022), In Pakistan 80% milk distribute in rural ,15% in pre urban and 5% in urban areas. In Pakistan, there is low price of unpacked milk as compare to packed milk (Sid, Mor, Kishore, & Sharan Agat, 2021). There exists a variability in the costs associated with loose milk in urban and rural areas. This discrepancy can be mitigated through the adaptation of route optimization strategies.

To minimize the delivery cost route optimization is very important, it cannot be manual manage. A strong route optimization system reduces the cost of product and decrease the delivery time.

In logistics, route optimization for product delivery has advanced significantly, focusing on reducing costs, delivery times, and fuel consumption while maximizing customer satisfaction(Arishi et al., 2022).

In route optimization, traditional methods often use heuristics, which are particularly valuable in scenarios where computational resources are limited or information is incomplete. These methods expedite computations by focusing on a subset of the solution space, making them practical for complex optimization problems like the traveling salesman problem. However, heuristics do not always guarantee optimal solutions, leading to the exploration of more advanced techniques. Metaheuristic optimization methods, inspired by natural processes, have proven effective in solving Vehicle Routing Problems (VRP) by balancing convergence rates and diversity in solution search spaces, making them suitable for complex real-world engineering challenges (Xia et al., 2021).

Modern approaches leverage machine learning algorithms like metaheuristics, reinforcement learning, and K-means clustering to efficiently optimize delivery routes. Linear programming models are also used to minimize transportation costs by optimizing distances between branches and customer locations, considering factors like vehicle capacity, delivery time windows, road networks, and customer demand (Etemadnia et al., 2015). These methods help companies streamline their delivery processes and improve logistics efficiency.

This study proposed a comprehensive route optimization model that addresses both vehicle capacity and time window constraints. By leveraging advanced algorithms and optimization techniques, the goal is to streamline delivery routes, minimize travel time and costs, and ultimately enhance customer satisfaction. The findings from this research aim to provide actionable insights and practical solutions for dairy products, logistics companies handling perishable goods, contributing to improved efficiency and sustainability in the supply chain.

In order to evaluate the proposed approach, this research used the comprehensive dataset collected from [ref]. Further, this research used key metrics: total route length, inter-customer gap, time constraints, and capacity constraints. Total route length reflects the overall distance traveled, while the inter-customer gap measures the distance between consecutive customer visits. Time constraints assess the total duration required for route completion, and capacity constraints represent customer demand satisfaction.

Literature review

Traditional approaches in route optimization

VRP, or Vehicle Routing Problem, is an important optimization challenge in various domains like Intelligent Transportation and airline operations. Traditional VRP approaches face restrictions in handling complex routing problems and inter-vehicle conflicts (Prins, 2004; Yeh & Tan, 2021). Therefore, to address these issues, innovative solutions like the Spatiotemporal VRP algorithm have been suggested, capable of managing large graphs efficiently and providing collision-free routes for multiple UAVs during infrastructure inspections. Moreover, VRP is closely related to vehicle characteristics, necessitating exact consideration in problem-solving (Mańdziuk, 2018). In the context of Intelligent Transportation Systems, distributed applications based on Vehicular Ad Hoc Networks need robust coordination mechanisms like the Vehicular Causal Block Protocol to ensure reliable communication and coordination among vehicles despite network challenges. Furthermore, VRP has been also used in crew pairing problems within airline operations, where new VRP-based models have shown efficiency and effectiveness in optimizing flight pairings and crew costs (Campbell, 2013).

Capacity Vehicle Routing Problems (CVRP) are addressed through numerous optimization approaches in logistics and transportation. One of the most used common methods is the use of algorithms like the Sweep algorithm and Guided Local Search (Avdoshin & Beresneva, 2019). These algorithms aim to find the most effective routes for delivering goods while considering capacity constraints and minimizing costs.

VRPTW is skillfully addressed through various innovative approaches. Researchers have proposed hyper-heuristic algorithms based on reinforcement learning. Additionally, studies have focused on utilizing Simulated Annealing (SA) algorithms to optimize VRPTW, achieving impressive results in minimizing total distance traveled and adhering to specific timetables for customer service . These approaches showcase the advantage of combining advanced algorithms with optimization techniques to tackle the complexities of VRPTW and enhance the efficiency of vehicle routing in various real-world scenarios.

The study(Guo & Wang, 2023) investigates a previously neglected aspect of the vehicle routing problem with concurrent pickup and delivery considering the total number of collected goods. Based on the postulates of considering the number of collected goods, a bi-objective vehicle routing model decreasing the total travel time and maximizing the total number of collected goods simultaneously is developed. A polynomial time approximation algorithm based on the ε-constraint method is designed to address this problem, and the approximation ratio of the algorithm is examined.

The research (Lai et al., 2022) proposed a heuristic backbone-based origin-destination insertion algorithm for vehicle route optimization in a data-driven flexible transit system, enhancing delivery ratio and reducing passenger waiting time.

Metaheuristic optimization methods, such as the Artificial Hummingbird Optimization Algorithm (AHA) (Yang et al., 2023), Search and Rescue Algorithm (SAR) , and spider colony simulation optimizer (Xia et al., 2021), have proven to be highly effective in optimizing routes compared to classical and heuristic algorithms. These metaheuristic algorithms are inspired by natural processes and excel in exploring complex search spaces to identify global optima, making them particularly suitable for real-world engineering problems where mathematical models may be challenging to map out. Metaheuristic algorithms have been largely utilized in various domains like engineering, finance, and computer science due to their ability to provide superior solutions by balancing merging rates and solution search space diverseness (Agrawal et al., 2021).
Traditional approaches in route optimization face several limitations. Firstly, traditional optimization techniques lack robustness, requiring a change in algorithms whenever the problem changes, leading to

ineffectiveness (Kilby et al., 2000). Therefore, the rigid and static nature of traditional approaches hinders adaptability and real-time responsiveness in route optimization tasks, necessitating the exploration of more flexible and dynamic solutions to address these limitations.

2.1 Advanced approaches in route optimization

Advanced approaches in route optimization involve the integration of machine learning and operational research algorithms to enhance solution quality. One such approach, "Learning to Guide Local Search" (L2GLS), combines reinforcement learning with penalty terms to adaptively adjust search efforts, effectively escaping local optima and achieving state-of-the-art results in larger Traveling Salesman Problem (TSP) and Capacitated Vehicle Routing Problem (CVRP) instances(Sultana et al., 2021). Additionally, the field of Combinatorial Optimization offers a generalized framework for formulating hard combinatorial optimization problems as linear programs, demonstrating advancements in solving NP-complete problems directly without reduction, thus contributing to the theory and application of extended formulations(Sánchez et al., 2020). Furthermore, optimization frameworks utilizing Evolutionary Computation aid in setting optimal routing weights for network protocols, supporting complex network planning tasks and enhancing service quality through multi-objective optimization approaches (Sánchez et al., 2020). These approaches collectively showcase the progress in leveraging diverse methodologies to address routing optimization challenges effectively.

Geographic Information Systems (GIS) play a Vitol role in route optimization by enabling the analysis of geospatial data for efficient planning and decision-making in transportation systems (Tao, 2013). GIS technology allows for the collection, processing, and utilization of geographic location factors to optimize routing paths, minimize delays, and enhance the overall efficiency of transportation networks (Abousaeidi et al., 2016).

The research conducted by (Deshmukh et al., 2019) centered around Geographic Information Systems and remote sensing aid in optimizing routes by utilizing satellite data for shortest and optimal paths, reducing travel time and fuel consumption in transportation planning. Geographic Information Systems (GIS) and remote sensing play pivotal roles in route optimization by providing essential data for efficient planning and decision-making processes. GIS enables the analysis of geospatial and non-spatial data, aiding in managing complex transportation networks (Deshmukh et al., 2019). Remote sensing technology, coupled with GIS, allows for the collection of valuable spatial data from satellites, enhancing the accuracy of route planning and optimization (Ouellette & Getinet, 2016). By utilizing GIS, transportation systems can benefit from network analysis to determine optimal routes, shortest paths, and closest facilities, ultimately reducing travel time and costs (Abousaeidi et al., 2016).

In (Hassine et al., 2023)the researchers use machine learning approach for solution and they use two algorithms first algorithms use K-means method for customers and second find the optimal delivery route. Route optimization using machine learning involves leveraging advanced algorithms to enhance traditional approaches for solving complex optimization problems in various domains(Tiwari & Sharma, 2023).

The authors emphasized a novel approach in the context of an artificial neural network to predict the fuel consumption according to the weather environment and the optimal route was determined through a genetic algorithm, which was confirmed to be a useful method for determining the optimal path (Jong-Kyu, et al).

Route optimization using AI involves leveraging advanced technologies like machine learning and deep learning to find the most efficient and cost-effective routes for vehicles. Various methods have been proposed, such as reinforcement learning combined with nonparametric clustering and Dijkstra's algorithm , deep neural networks with weighting methods like Ratio estimation and Rank sum method ,

and the development of a new machine learning method for optimal route determination(Migel et al., 2024). For Electric Vehicles (EVs), optimization algorithms consider factors like energy consumption, battery choice, and topography, utilizing metaheuristic algorithms like the Artificial Hummingbird Algorithm (AHA) for energy-efficient route planning (Lokhannadh et al., 2023).

Research methodology

3.1 Research Design

The proposed methodology is based on the Deep Reinforcement Learning (DRL) which is considered to be the state-of-the-art in the VRP. Therefore, this research proposed a modified version of DRL.

DRL is a branch of Machine Learning (ML) the combination of Reinforcement Learning (RL) and Deep Learning (DL). RL involves an agent learning to act in a way that maximizes a reward, while DL is a kind of machine learning which uses multi-layered neural networks to learn complex patterns in data.

In DRL, the agent uses a deep multi-layered neural network to estimate the value function or policy, allowing it to learn complex behaviors and make decisions in multi-dimensional state and action spaces. In figure 1 explain all the process.

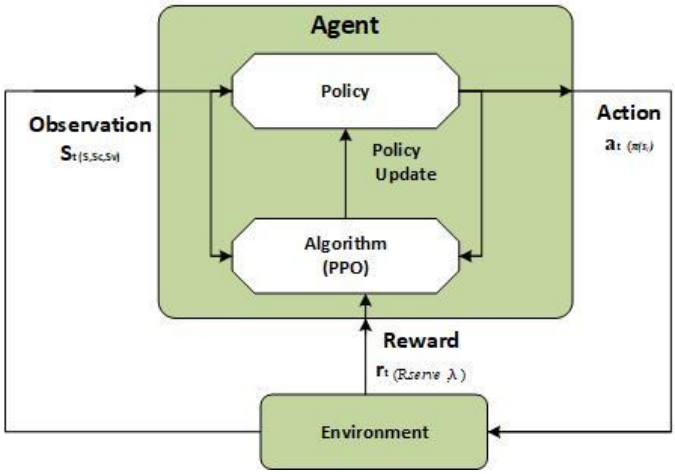

Figure 3.1: Loop recurring in reinforcement learning algorithms

3.2 Objective

The goal is to minimize the travel time or distance for a vehicles to service customers, minimizing travel distance or time while respecting vehicle capacity and customer time window constraints with capacity and time limits. All vehicles start and end at a central depot. Key constraints include not exceeding vehicle capacity, visiting customers within their time windows, and minimizing total travel distance or time.

The goal is to minimize the total travel distance or time while respecting vehicle capacity and customer time windows. All the above scenario we represent in mathematically in Equation 1.

$$min \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij} \cdot x_{ij} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \textbf{(1)}$$

In the Equation1,$\boldsymbol{N}$ is the total number of customers and depot locations, $\boldsymbol{d_{ij}}$ is the distance or time between location $\boldsymbol{i}$ and location $\boldsymbol{j}$ and $\boldsymbol{x_{ij} = 1}$ if vehicle travels from location i ***to*** location $\boldsymbol{j}$ otherwise $\boldsymbol{x_{ij} = 0}$

In capacity constraint the total demand served by a vehicle should not exceed its capacity. We formulate it in Equation 2.

$$\sum_{j=1}^{N} q_j \cdot x_{ij} \leq C_k \forall k \text{……. (2)}$$

In Equation 2,$q_j$is the Demond of customer j and $C_k$is the capacity of the vehicle k.

In time window constraint each customer must be visited within their specific time window. This constraint is mathematically formulated in Equation 3.

$\alpha_j \leq t_j \leq b_j \forall j$.. …… ……(3)

In Equation 3, α_j and b_j are the start and end times of the time window for customers j and t_j is the actual arrival time at customer j.

Each vehicle must leave a location it arrives at, except for the depot. The vehicle location mathematically formulates in Equation 4.

$$\sum_{i=1}^{N} x_{ij} = \sum_{k=1}^{N} x_{jk} \forall j \text{……..……..(4)}$$

3.3 Environment

Vehicle states, customer states, and a global state are the part of the environment. Vehicle states include the location, time and the left amount of capacity. Customer states include remaining demand, service status and time windows. This comprehensive view informs the decision-making process. The environment mathematically we show as under.

The above statement the vehicle state (s_v) = {location, remaining capacity, current time}, the Customer state (s_c) = {remaining demand, time window, service status} and Global state (S) = {current routes, remaining customers, elapsed time}

3.4 Action

Actions involve choosing the next customer for each vehicle, with the option to return to the depot if no feasible customer is available within constraints. The action procedure mathematically formulated in Equation 5.

The action α_t at time step t involves selecting the next customer or returning to depot.

$\alpha_t = \arg \max \pi(s_t)$……..(5)

In Equation5,$\pi(s_t)$ is the policy network output giving the probability of choosing each customer c or returning to the depot.

3.5 Reward Function

The reward function offers positive rewards for serving customers and negative rewards for travel costs or distances. Penalties are applied for exceeding time windows or exceeding capacity, ensuring adherence to constraints. The reward function we write mathematically in Equation 6.

The reward r_t at time step t is designed to encourage serving customers and penalize distance, time and constraint violations.

$r_t = \sum_{i=1}^{N} R_{\text{serve}} \cdot y_{ij} - \lambda \sum_{i=1}^{N} d_{ij} \cdot x_{ij} - \mu \cdot \text{penalties}$ **……..(6)**

In Equation6, R_{serve} is the positive reward for serving a customer, λ is a penalty factor for distance or time and μ is a penalty factor for constraint violations

3.6 Algorithm

Proximal Policy Optimization (PPO) is employed for its stability and effectiveness in handling complex environments. PPO provides a robust framework for training the Deep Reinforcement Learning (DRL) agent.

Using PPO, the agent updates its policy π_θ and value function V_ϕ based on the experience stored in replay buffer.The algorithm instructions mathematically formulated in Equation 7.

Max Et [min (rt(θ)$\hat{A}$t,clip (rt (θ) 1 - ϵ, 1 +ϵ) $\hat{A}$t)]**……(7)**
In Equation 7, the $r_t(\theta)$ is the probability ratio under the new and old policies , $\hat{A}_t$ is the advantages function at time t and ϵ is a small hyperparameter for clipping.

3.7 Agent

The DRL agent consists of a policy network, value network, and experience replay buffer. The policy network outputs probabilities for selecting the next customer, the value network estimates the expected return from the current state, and the experience replay buffer stores past experiences to stabilize training.

3.8 Training Process

The training process starts by initializing the environment at the beginning of each episode. During each step, the policy network selects the next customer based on the current state. The action is executed, and the next state and reward are observed. The (state, action, reward, next state) tuple is stored in the replay buffer. This iterative process allows the agent to learn and improve its performance in optimizing vehicle routes under given constraints.

Using proximal Policy Optimization (PPO) the agent updates its policy π_θ and value function V_ϕ based on the experience stored in replay buffer. In training processes the agent follow the algorithm instructions so, It will also mathematically formulate in the Equation 8.

Max E_t [min ($r_t(\theta)\hat{A}_t$, Clip ($r_t(\theta)$ 1 - ϵ, 1 +ϵ) $\hat{A}_t$)]**……(8)**
In Equation8, the $r_t(\theta)$ is the probability ratio under the new and old policies , $\hat{A}_t$ is the advantages function at time t and ϵ is a small hyperparameter for clipping.

NUMERICAL RESULTS

We compare our proposed DRL-PPO algorithm with two prominent DRL benchmark algorithms: Advantage Actor-Critic (A2C) and Deep Q-Network (DQN). DQN, a value-based DRL algorithm, approximates optimal Q-values to make vehicle routing decisions while optimizing costs and adhering to constraints such as delivery time and vehicle capacity. In contrast, A2C employs a hybrid approach with

separate networks for policy (actor) and value estimation (critic), enhancing decision-making efficiency in dynamic and constrained routing scenarios.

The performance evaluation focuses on key metrics: total route length, inter-customer gap, time constraints, and capacity constraints. Total route length reflects the overall distance traveled, while the inter-customer gap measures the distance between consecutive customer visits. Time constraints assess the total duration required for route completion, and capacity constraints represent customer demand satisfaction.

This comprehensive analysis provides a detailed understanding of each algorithm's performance across these critical parameters.

4.1 Experimental Results

This section provides a detailed analysis of the experimental results to evaluate the routing and delivery efficiency of A2C, DQN, and DRL-PPO.

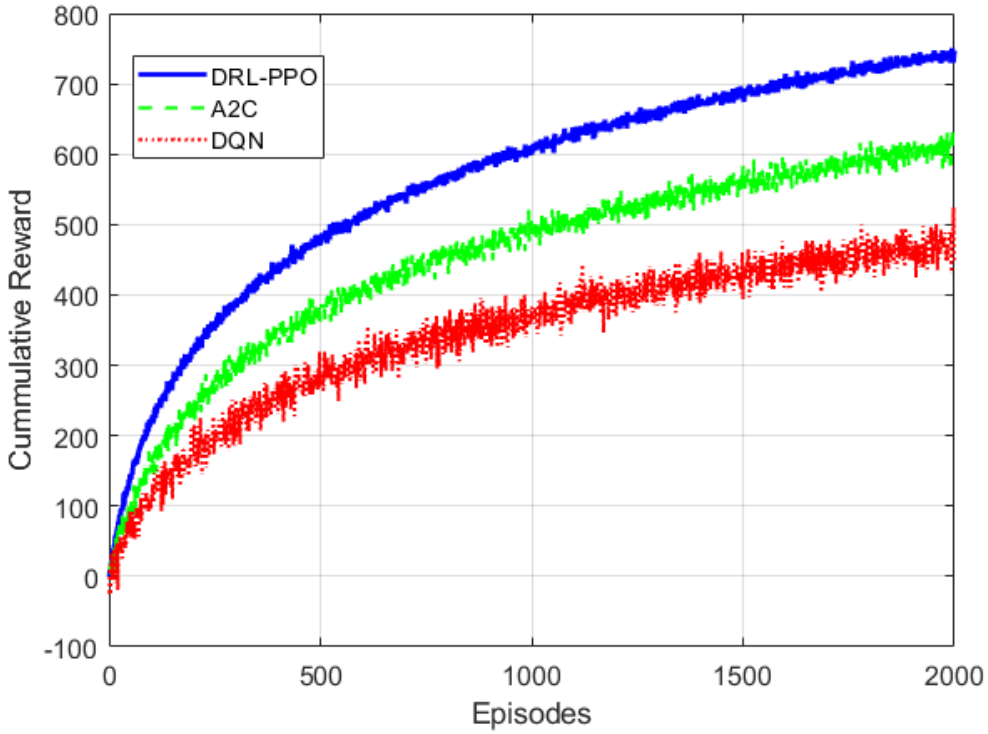

Fig. 1 Cumulative Rewards vs No. of Episodes

Fig. 1 showcases the performance comparison of the proposed DRL-PPO algorithm with DRL-based benchmarks, A2C and DQN, across training episodes. The results highlight the superior performance of DRL-PPO in terms of cumulative reward, underscoring its effectiveness in learning and implementing optimal delivery strategies. This notable performance advantage positions DRL-PPO as a promising solution for challenges in the Pakistani dairy sector, such as high delivery costs and inefficiencies. By optimizing delivery routes, DRL-PPO can significantly reduce delivery times, fuel consumption, and overall operational expenses, contributing to enhanced profitability for the dairy industry and more affordable milk prices for consumers.

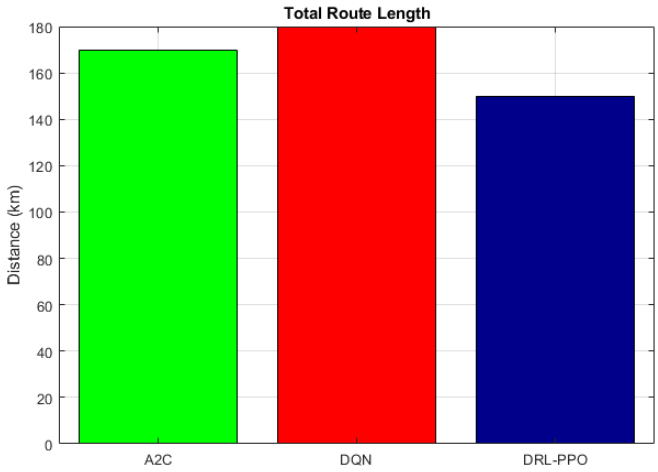

Fig. 2 Total Route Length

Fig. 2 illustrates the total route length (in kilometers) traveled by a delivery vehicle under different routing strategies learned by A2C, DQN, and DRL-PPO. Among these algorithms, DRL-PPO demonstrates the most efficient performance, achieving the shortest route length of approximately 150 km, indicating its effectiveness in minimizing travel distances. In contrast, A2C and DQN yield longer routes, measuring around 170 km and 180 km, respectively, reflecting their relatively less efficient exploration and decision-making capabilities. This comparison highlights the superiority of DRL-PPO in optimizing vehicle routing, making it the most suitable approach for scenarios requiring reduced travel distances and improved route efficiency.

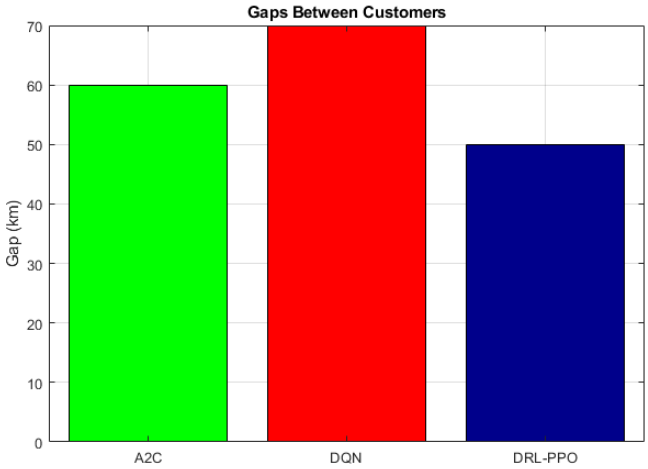

Fig. 3 Gap Between Customers

Fig. 3 presents the cumulative distances (kms) between consecutive customer visits for the routing strategies learned by A2C, DQN, and DRL-PPO. Among the algorithms, DRL-PPO demonstrates the most optimized transitions, achieving a reduced cumulative gap of approximately 50 km, indicating smoother and more efficient sequencing of deliveries. In comparison, A2C and DQN result in larger cumulative

gaps of approximately 60 km and 70 km, respectively, reflecting less efficient transitions between customers. This analysis highlights the effectiveness of DRL-PPO in minimizing inter-customer gaps, leading to shorter overall routes and improved delivery efficiency.

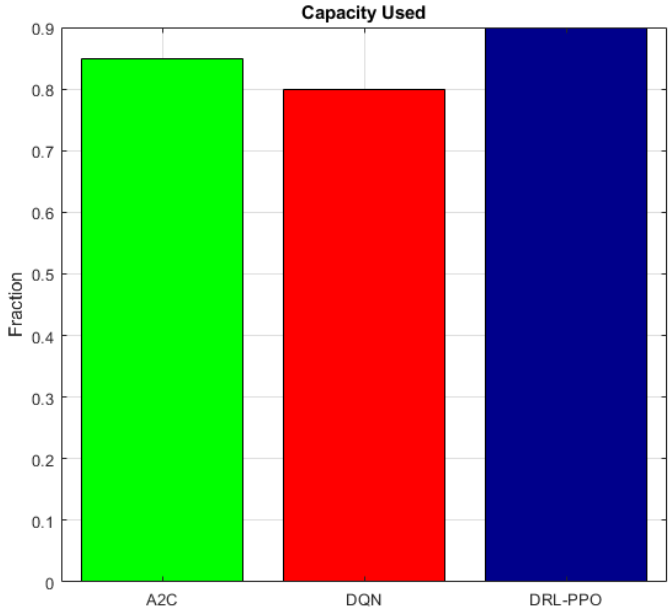

Fig. 4 Capacity Utilization

Fig. 4 compares the fraction of vehicle capacity utilized during deliveries for A2C, DQN, and DRL-PPO. Our proposed algorithm DRL-PPO achieves the highest capacity utilization at approximately 90%, demonstrating superior management of delivery constraints while maximizing resource usage. A2C follows with a utilization of about 85%, and DQN exhibits the lowest utilization at approximately 80%. The enhanced performance of DRL-PPO highlights its ability to optimize vehicle loading, which is crucial for ensuring cost-effective and efficient logistics operations.

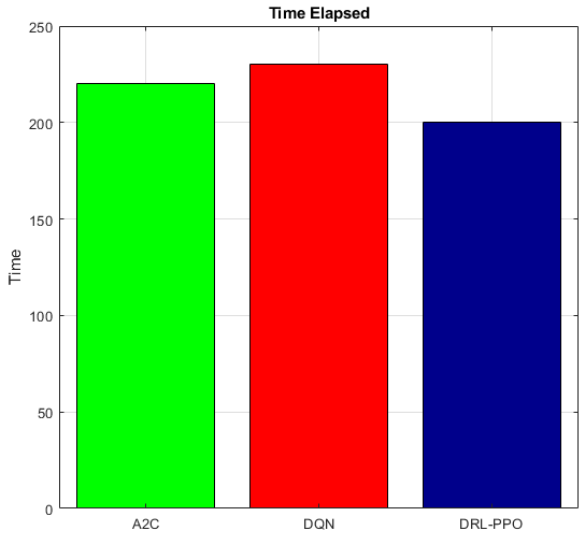

Fig. 5 Time Elapsed

Fig. 5 illustrates the total time taken to complete the delivery route for each algorithm. DRL-PPO achieves the shortest completion time of 200 units, showcasing its efficiency in identifying time-optimal routes. In contrast, A2C and DQN require 220 and 230 units, respectively, indicating longer route completion times due to less effective policy learning. The results emphasize DRL-PPO's superiority in minimizing delivery time, which is critical for time-sensitive operations.

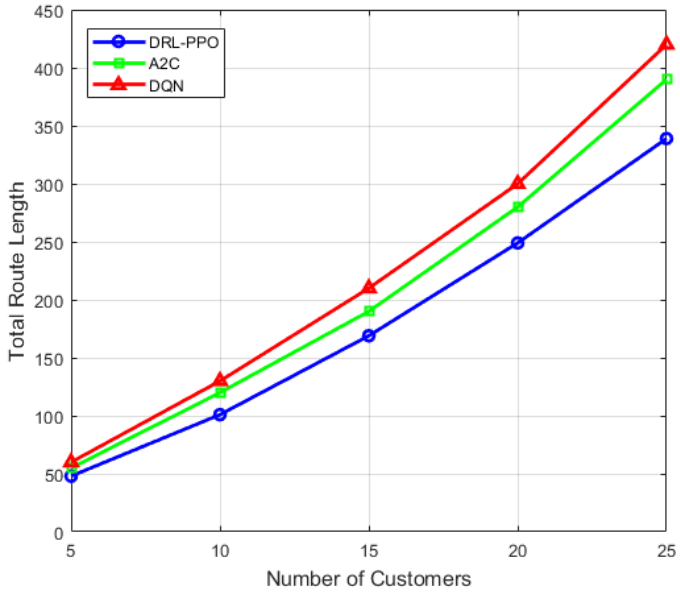

Fig. 6 Impact of Task Complexity on Route Length

Fig. 6 illustrates the relationship between the total route length and the number of customers for three algorithms: DRL-PPO, A2C, and DQN. As the number of customers increases, the total route length consistently grows for all algorithms, reflecting the inherent complexity of solving vehicle routing problems with a larger customer base. Notably, DRL-PPO consistently achieves shorter route lengths compared to A2C and DQN, highlighting its superior optimization capabilities. The linear growth pattern across all methods suggests that DRL-PPO is more efficient in handling task complexity, whereas DQN exhibits the highest route lengths, underscoring its relative inefficiency. The results showcase DRL-PPO's ability to minimize travel distance while maintaining operational efficiency

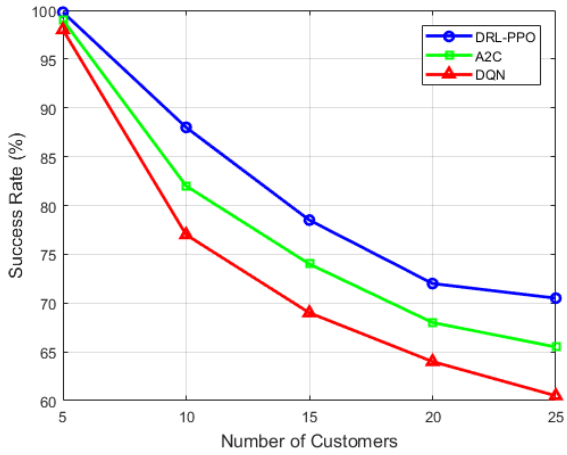

Fig. 7 Impact of Task Complexity on Success Rate

Fig. 7 depicts the success rate of the algorithms against the number of customers, defined as the percentage of tasks completed within given constraints (e.g., time and capacity). DRL-PPO exhibits the highest success rates across all complexity levels, starting at 100% and declining gradually to 71% as the number of customers increases. A2C follows with moderate performance, while DQN shows the steepest decline in success rates, dropping to 60% for 25 customers. This outcome suggests that DRL-PPO is more robust in adapting to increasing task complexity, maintaining higher efficiency in constraint satisfaction. The performance gap between the algorithms widens with a growing number of customers, emphasizing DRL-PPO's superior scalability.

Conclusion

This study clarifies the effectiveness of advanced route optimization techniques, particularly Deep Reinforcement Learning (DRL) using Proximal Policy Optimization (PPO), in reducing logistical challenges related to the distribution of dairy products. The suggested framework significantly reduces delivery costs, travel distances, and total time while making the best use of vehicle capacity by taking into account constraints such vehicle capacity, delivery windows, and customer demand. The empirical results demonstrate the model's ability to improve operational efficiencies in Pakistan's dairy industry and validate its superiority over traditional techniques like Deep Q-Network (DQN) and Advantage Actor-Critic (A2C). By lowering fuel usage and carbon emissions, DRL-PPO not only improves operational effectiveness but also promotes sustainability. Both the dairy industry and consumers gain from these advancements since they make it easier to distribute dairy products in a more cost-effective and hygienic manner.

REFERENCES

Abousaeidi, M., Fauzi, R., & Muhamad, R. (2016). Geographic Information System (GIS) modeling approach to determine the fastest delivery routes. *Saudi journal of biological sciences*, *23*(5), 555-564.

Agrawal, P., Abutarboush, H. F., Ganesh, T., & Mohamed, A. W. (2021). Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). *Ieee Access*, *9*, 26766-26791.

Arishi, A., Krishnan, K., & Arishi, M. (2022). Machine learning approach for truck-drones based last-mile delivery in the era of industry 4.0. *Engineering Applications of Artificial Intelligence*, *116*, 105439.

Avdoshin, S. M., & Beresneva, E. N. (2019). Local search metaheuristics for capacitated vehicle routing problem: a comparative study. *Труды института системного программирования РАН*, *31*(4), 121-138.

Campbell, I. M. D. (2013). *Construction heuristics for the airline taxi problem* University of the Witwatersrand, Faculty of Engineering and the Built …].

Daneshmand, R., & Shahidi, S. (2023). A review and analysis on fertility and milk production in commercial dairy farms with customized lactation length during the last ten years. *Journal of New Findings in Health and Educational Sciences (IJHES)*, *1*(3), 20-37.

Deshmukh, P., Rao, D., Botale, R., & Pwade, P. (2019). Remote Sensing and Geographic Information System-Based Route Planning. Smart Technologies for Energy, Environment and Sustainable Development: Select Proceedings of ICSTEESD 2018,

Etemadnia, H., Goetz, S. J., Canning, P., & Tavallali, M. S. (2015). Optimal wholesale facilities location within the fruit and vegetables supply chain with bimodal transportation options: An LP-MIP heuristic approach. *European Journal of Operational Research*, *244*(2), 648-661.

Guo, Q., & Wang, N. (2023). The Vehicle Routing Problem with Simultaneous Pickup and Delivery Considering the Total Number of Collected Goods. *Mathematics*, *11*(2), 311.

Hassine, M. B., Sakhri, M. S. A., & Tlili, M. (2023). Machine learning approach to solving the capacitated vehicle routing problem. 2023 IEEE Afro-Mediterranean Conference on Artificial Intelligence (AMCAI),

Kilby, P., Prosser, P., & Shaw, P. (2000). A comparison of traditional and constraint-based heuristic methods on vehicle routing problems with side constraints. *Constraints*, *5*(4), 389-414.

Lai, Y., Yang, F., Meng, G., & Lu, W. (2022). Data-driven flexible vehicle scheduling and route optimization. *IEEE Transactions on Intelligent Transportation Systems*, *23*(12), 23099-23113.

Lokhannadh, C., Rajesh, A., Manikandan, C., Rajamohan, J., Easwaran, M., & Uma, S. (2023). Electric Vehicles (EV) Route Optimization Using Artificial Hummingbird Algorithm. 2023 4th International Conference on Smart Electronics and Communication (ICOSEC),

Mańdziuk, J. (2018). New shades of the vehicle routing problem: Emerging problem formulations and computational intelligence solution methods. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *3*(3), 230-244.

Migel, S., Maloied, M., Zaliskyi, M., Lelechenko, A., Osipchuk, A., & Solomentsev, O. (2024). Optimal Pathfinding Based on Artificial Intelligence Tools. International Workshop on Advances in Civil Aviation Systems Development,

Miller, V., Reedy, J., Cudhea, F., Zhang, J., Shi, P., Erndt-Marino, J., Coates, J., Micha, R., Webb, P., & Mozaffarian, D. (2022). Global, regional, and national consumption of animal-source foods between 1990 and 2018: findings from the Global Dietary Database. *The Lancet Planetary Health*, *6*(3), e243-e256.

Ouellette, W., & Getinet, W. (2016). Remote sensing for marine spatial planning and integrated coastal areas management: Achievements, challenges, opportunities and future prospects. *Remote Sensing Applications: Society and Environment*, *4*, 138-157.

Pereira, P. C. (2014). Milk nutritional composition and its role in human health. *Nutrition*, *30*(6), 619-627.

Prins, C. (2004). A simple and effective evolutionary algorithm for the vehicle routing problem. *Computers & operations research*, *31*(12), 1985-2002.

Sánchez, M., Cruz-Duarte, J. M., carlos Ortíz-Bayliss, J., Ceballos, H., Terashima-Marin, H., & Amaya, I. (2020). A systematic review of hyper-heuristics on combinatorial optimization problems. *IEEE Access*, *8*, 128068-128095.

Sultana, N., Chan, J., Sarwar, T., Abbasi, B., & Qin, A. K. (2021). Learning enhanced optimisation for routing problems. *arXiv preprint arXiv:2109.08345*.

Tao, W. (2013). Interdisciplinary urban GIS for smart cities: advancements and opportunities. *Geo-spatial Information Science*, *16*(1), 25-34.

Tiwari, K. V., & Sharma, S. K. (2023). An optimization model for vehicle routing problem in last-mile delivery. *Expert Systems with Applications*, *222*, 119789.

Xia, X., Liao, W., Zhang, Y., & Peng, X. (2021). A discrete spider monkey optimization for the vehicle routing problem with stochastic demands. *Applied Soft Computing*, *111*, 107676.

Yang, F., Yue, L., & Zhang, X. (2023). Application of Diffusion Artificial Hummingbird Algorithm for Multimodal Transportation Logistics Distribution Routing Problem. 2023 China Automation Congress (CAC),

Yeh, W.-C., & Tan, S.-Y. (2021). Simplified swarm optimization for the heterogeneous fleet vehicle routing problem with time-varying continuous speed function. *Electronics*, *10*(15), 1775.