Variable Neighborhood Search Based on Reinforcement Learning for Green Vehicle Routing Problem

1st Malek Alrashidi
department of computer science, Applied College
University of Tabuk
Tabuk, Saudi Arabia
mqalrashidi@ut.edu.sa

2nd Mansoor Al Ghamdi
department of computer science, Applied College
University of Tabuk
Tabuk, Saudi Arabia
malghamdi@ut.edu.sa

Abstract*—The green vehicle routing problem (GVRP) is a trendy variant of the well-known vehicle routing problem that incorporates environmental considerations such as minimized fuel consumption or emissions. This study introduces a new hybrid approach combining variable neighborhood search (VNS) with the reinforcement learning (RL) paradigm to effectively resolve the GVRP. VNS is a metaheuristic optimization technique that explores numerous neighborhoods of a solution to improve it, whereas RL is an agent-based machine learning algorithm. Thus, integrated with VNS, the RL may find competitive solutions for the GVRP. Our numerical results and analysis prove the effectiveness of the proposed hybrid methodology for resolving the GVRP. This new hybrid strategy benefits from the combination of VNS as an evolutionary optimizer and RL as a machine learning methodology to effectively resolve the GVRP.

***Keywords*—variable neighborhood search, reinforcement learning, machine learning, green vehicle routing problem, hybridization**

I. INTRODUCTION

The vehicle routing problem (VRP) represents a combinatorial problem aiming to determine the optimal configuration for distributing merchandise or services from a central location (depot) to numerous geographically dispersed locations via a fleet of vehicles [1]. The main aim of the VRP is to reduce the overall cost, which can be achieved in various ways, such as minimizing the total traveled distance, the number of involved vehicles, and/or the overall required time.

The problem typically includes the following key elements:

- **Depot**: The central location where all vehicles start and return from deliveries. It serves as the point of origin for all deliveries.
- **Customers** (or nodes): The positions to which merchandise or services should be delivered. Each customer has a specific demand that needs to be met.
- **Vehicles**: The fleet of vehicles available for making deliveries. Each one has a limited quantity of items to transport.

The VRP mainly determines which customers each vehicle should visit, in what sequence, and how much to deliver to each customer while respecting capacity constraints and minimizing the total cost.

The following are the relevant categories of the VRP:

- **Capacitated VRP** (**CVRP**) [2]: It is the most common form of the VRP, where vehicles have limited capacities, and the goal is to meet the demands of all customers while not exceeding these capacities.
- **VRP including time windows** (**VRPTW**) [3]: In addition to capacity constraints, this version adds time windows within which each customer must be serviced.
- **VRP with pickup and delivery** (**VRPPD**) [4]: This category involves picking up items from certain customers and delivering them to others. It is common in applications such as waste collection and public transportation.
- **Green VRP (GVRP)** [5]: The GVRP focuses on optimizing routes while considering environmental factors such as fuel consumption and emissions.

The GVRP focuses on making the routing and distribution process more environmentally friendly. This process may involve reducing fuel consumption, minimizing emissions, or using alternative energy sources, such as electric vehicles. GVRP is an important research area owing to its practical applications, as companies and governments seek to reduce their carbon footprint and ensure sustainable operations.

Solving the GVRP may involve optimizing routes and schedules to minimize greenhouse gas emissions by considering factors such as vehicle type, fuel consumption, and alternative energy sources in addition to factors related to traffic and congestion that affect fuel efficiency. Considering these factors aligns with the broader goals of sustainable logistics and transportation.

The GVRP is a challenging problem with numerous real-world applications in logistics, such as package delivery, garbage collection, and school bus routing. Finding optimal solutions to VRPs can be computationally intensive, and many advanced algorithms and heuristics have been proposed to efficiently address these problems.

Variable neighborhood search (VNS) [6] is a metaheuristic optimization algorithm introduced by Mladenovic and Hansen in 1997. It is a local search–based approach that aims to efficiently explore different neighborhoods or regions in the search space of a problem to find high-quality solutions.

Reinforcement learning (RL) [7] is another methodology for the VRP. It is a machine learning (ML) paradigm in which an actor, called an agent, learns to make sequences of decisions by interacting with an environment. RL is a subfield of artificial intelligence that is particularly well-suited for solving problems where the optimal decision-making strategy is not known in advance. In RL, an agent learns through a process of trial and error—the same as how humans learn.

Hybridization can be a powerful approach to solving complex optimization problems in different research fields, such as Internet of Things (IoT) networking [8] [9] [10], pattern recognition [11], and healthcare monitoring [12][13], especially those that involve sequential decision-making, such as augmented reality [14] or financial prediction [15] problems. Hybridizing VNS with RL for efficiently and effectively solving the GVRP is an innovative approach that leverages the strengths of both methods.

The following research questions about the GVRP are addressed here: What novel hybrid optimization algorithms and heuristics can be developed to efficiently solve GVRPs considering different factors? How can environmental (e.g., emission reduction and improved fuel efficiency) and economic objectives (e.g., cost minimization) be simultaneously achieved, and how can they be balanced in GVRPs?

Indeed, in the optimization field, some algorithms may look similar to each other, but each one has its own innovations and differences. New scientific research draws from extant literature to advance research progress and overcome existing drawbacks. Therefore, regardless of whether the names and mechanisms of the algorithms are similar or different, the final goal is to efficiently solve various optimization problems, especially transport problems.

II. CONTRIBUTION OF THE STUDY

In this study, a novel hybrid optimization strategy is established to resolve the GVRP while mitigating environmental impacts and considering conflictual multi-objectives such as maximizing the quantity of delivered products and reducing the distance traveled, emissions, and fuel consumption.

This study does not represent a collection of existing works. Indeed, several hybridizations have been proposed in the extant literature for different variants of the VRP. Our hybrid model is well justified and leverages the benefits of both VNS and RL for enhancing, as evidenced by our study results via different metrics, the quality of the VRP solution. Thus, our study makes a significant contribution to the VRP literature by presenting a hybrid model that enhances the VRP solution, especially considering that the recent variants of the problem, such as the GVRP, are still not well studied.

In the remainder of this paper, Section III investigates the recent related state-of-the-art studies. Section IV details the hybrid RL–VNS. Section V presents the numerical tests. Section VI presents the deductions, and Section VII concludes the study.

III. STATE-OF-THE-ART RESEARCH

This section discusses recent studies on the multi-objective and VNS paradigms for VRP. Next, RL and LSTM-based studies for VRP are investigated. Then, hybrid RL–VNS-based studies and other related methods are discussed. Finally, Table 1 presents the surveys of VRP variants.

A. Optimization-based Studies for VRP

In [16], the CVRP is resolved by combining the ant colony optimizer (ACO) with the fireworks algorithm for better diversity. In this ACO, the local search process is enhanced by an ant strategy. Test results of execution on the Augerat benchmark instances highlight the performance of the suggested algorithm compared with the best-found results in the literature for five new solutions. The authors in [17] propose a two-stage algorithm (TSA) combined with a multi-objective evolutionary strategy (HMOEA) to resolve the time window multi-depot heterogeneous VRP. The results of execution on Solomon instances illustrate that HMOEA has better performance in terms of diversity and convergence than known optimizers, such as SPEA2 and NSGA-II.

B. *RL and LSTM for VRP*

In [18], a deep RL platform was introduced for the dynamic uncertain VRP (DU-VRP). Then, an RL cutting-edge–based method was proposed to train the uncertainty of the routing process. The authors in [19] introduce a multi-agent–based RL system for the scheduling and routing of electric vehicles. In [20], an attention-based deep RL learning strategy for electrical time window VRP has been proposed. This model is efficient for large-size instances of VRP. Other studies have suggested solutions for the VRP using neural networks [21] and RL [22].

TABLE I
RECENT SURVEYS OF VRP VARIANTS

Survey	Previously proposed VRP variant(s) and its constraint(s)	Resolution methodologies	Recommendations
[34]	Dynamic VRP; CVRP considering the capacity of vehicles; VRPTW considering time durations for collecting and delivering items; Stochastic VRP (probabilistic); Workload balance VRP	ML; Local search; Heuristics; Metaheuristics	Categorize vehicle routing heuristics into metaheuristics, improvement heuristics, and constructive heuristics. Recent related research topics include ML-assisted heuristics, unified heuristics, and automatic heuristic design.
[35]	Standard VRP; Urban vehicle routing (in the city); Multi-objective routing problem (having multiple objectives)	ML Multi-objective algorithms (MOA)	Routing services are considered (green routing, transport costs, time window, fleet management, travel time, and travel distance)
[36]	Standard VRP; Electric vehicles; VRPTW; Homogenous and heterogeneous vehicles; Single depot / multiple depot	Neighborhood search; Push-forward insertion heuristic; Improved artificial fish swarm; Tabu search	Heuristics and meta-heuristics remain the mainstream strategies for VRP. Complex VRP instances will be increasingly resolved owing to the rapid advances in hardware resources.
[37]	GVRP	Data-driven ML and RL strategies (forecasting methods) were not appropriately studied for GVRP. Owing to their acceptable time and precision, these optimization algorithms are the most used for the GVRP.	The GVRP considers environmental sustainability issues in logistics and transportation. The application of emergent paradigms for the VRP, such as RL, quantum computing, deep Q-learning, and chaos theory, remains under-researched.
[38]	GVRP	Optimizers (GA, NSGA-II, ACO, PSO, DEA, SA)	The study highlights future research directions for GVRPs: dynamic electric vehicle charging, variation in energy consumption, and real-time transportation.

C. *Hybrid RL–VNS-based Studies for VRP*

Stemming from the multi-armed bandit, a single-state RL paradigm, a heuristic named bandit for the VNS has been proposed in [23]. In another study [24], the authors have proposed an approximate dynamic programming strategy with a Markov decision process for the multi-depot stochastic road capacity dynamic VRP. [25] shows the application of the multi-agent deep RL to the dynamic and stochastic variant of the VRP considering the numerous operational specifications of this problem. Another combination of RL and VNS was introduced in [26] for the time window OPVRP, Wherein RL was employed during the local search phase to control the search by adjusting the probabilities of adaptive operators. Moreover, the thesis in [27] suggests an adaptive heuristic using an offline learning algorithm with a local search method.

D. *Other Paradigms for VRP*

This subsection discusses previous studies that use other paradigms to resolve the VRP. Multi-agent systems (MAS) and heuristic-based algorithms [28]–[30] are widely used in different optimization problems.

In [31], a sampling strategy is suggested for the capacitated DSVRP modeled as a two-stage stochastic program. In contrast, [32] illustrates the resolution of the time-dependent VRP in which a weight function is employed. Further, in [33], an asynchronous multi-agent tool (A-teams) is used to resolve the heterogeneous cooperative VRP. Agents in A-teams were used locally and globally to create, enhance, or remove solutions.

Table 1 below presents an investigation of the numerous recent surveys of VRP variants.

IV. METHODOLOGY

This section introduces the RL, VNS, and hybrid RL–VNS strategy.

A. *RL for GVRP*

RL is widely used for solving sequential decision-making problems and transport problems, such as the VRP. It can be applied to solve the GVRP as an ML approach where an agent learns to make a sequence of decisions by interacting with an environment.

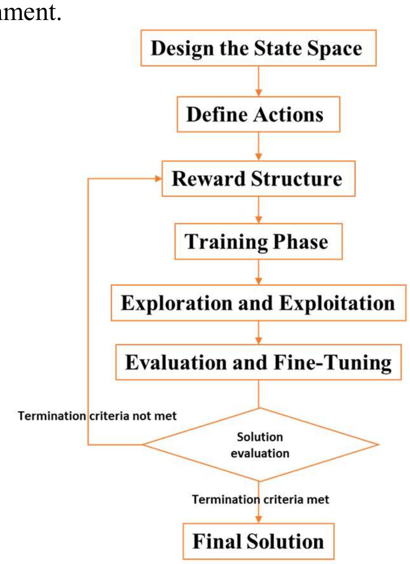

Fig. 1. The application of the RL algorithm to the GVRP.

This agent aims to maximize a cumulative reward signal by selecting actions that lead to favorable outcomes. Fig. 1 illustrates the application of RL for transport problems, such as the GVRP.

First, the problem must be formulated as an RL task. In the case of the GVRP, the state space determines the current route configuration; the action space represents the possible modifications to the routes; and the reward signal identifies the cost reduction achieved by the agent.

Then, the state space is designed. In the GVRP, the state space comprises information about the positions/locations of unvisited cities/customers, current routes, remaining capacity of the vehicles, and distance matrix between locations.

Using RL for transport problems, such as the GVRP, allows the development of adaptive, intelligent, and data-driven decision-making systems.

B. *VNS for GVRP*

VNS is a metaheuristic optimization technique applicable to complex optimization problems, such as the VRP. As illustrated in Fig. 2, VNS iteratively explores several *neighborhoods* or regions in the search space of a problem to find high-quality solutions. This algorithm first generates an initial VRP solution either randomly or using heuristics. Then, the neighborhoods are defined. In the context of the VRP, neighborhoods represent the different ways to modify or rearrange the routes or assignments of customers to vehicles. Common neighborhood structures include swap (swapping customers between different routes), 2-opt (reversing the order of customers within a single route), and Or-opt (reassigning a customer from one route to another).

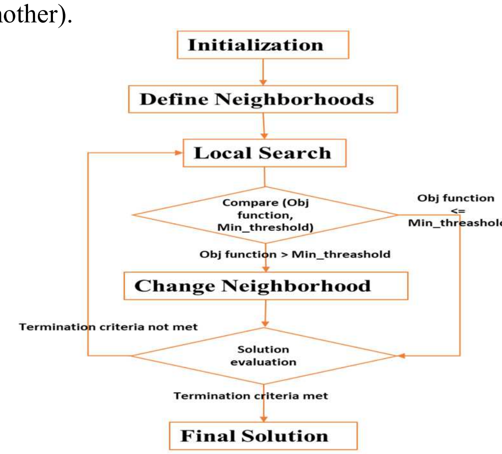

Fig. 2. The application of the VNS algorithm to the GVRP.

Therefore, VNS can be used as a standalone optimization method or as part of a hybrid approach—such as combining it with RL or any other metaheuristic algorithm—to enhance the efficiency of the solution.

C. *VNS–RL for GVRP*

Combining VNS with RL for solving the GVRP can be an efficient method to address sustainability issues in logistics and transportation.

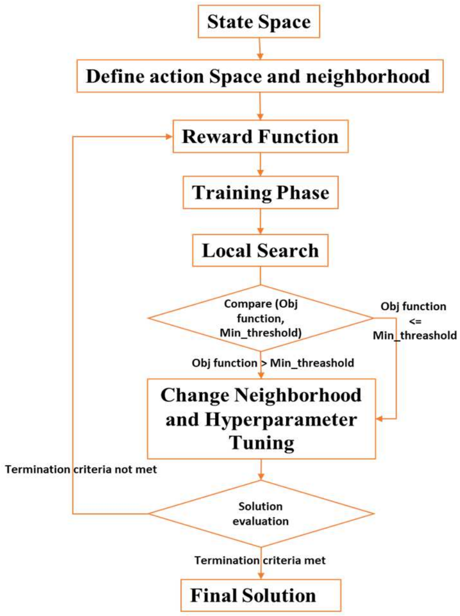

Fig. 3. The application of the hybrid RL–VNS to the GVRP.

Fig. 3 illustrates the application of the RL–VNS for solving the GVRP by leveraging the efficient VNS generation of initial solutions and exploiting the RL capacity for learning adaptive and environmentally conscious policies. This hybrid approach offers the potential for additional sustainable and efficient routing solutions.

V. NUMERICAL RESULTS

This section presents the performance assessment of the proposed RL–VNS on known datasets. The experimental tests were conducted using an Intel© CoreTM i9 CPU, 3.2 GHz, 24-GB memory under a Windows 11 operating system.

The parameter settings of the VNS algorithm are as follows: $r_n = 0.1$, $r_m = 0.25$, MaxTolerance = 0.04, and iterations = 1000. Here, r_n and r_m indicate the minimum and maximum rate of changing the neighborhood of the solutions, respectively. MaxTolerance indicates the limit of the tolerance dimension in the VNS algorithm.

The GA algorithm relies on the following probabilities: 0.8 and 0.1 for the crossover and mutation, respectively.

For the VRP, the instances of Solomon [39] are used, and the following setting is applied:

Number of generations = 350 Individuals per generation = 300 Number of cities = 100

After a certain period, the search becomes sensible if the search solutions do not improve. Using a set of neighborhood operators to define the action space (Table 2), the VNS algorithm can concentrate the search on a specific region to retrieve optimal solutions from the neighborhood. Then, the best solutions from the neighborhood are selected.

TABLE 2
NEIGHBORHOOD LOCAL SEARCH OPERATORS

	Complexity	Space index
Cross-insertion	$O(m^2n^3)$	0
2-opt	$O(mn^2)$	2
3-opt	$O(mn^2)$	1
3-cross-exchange	$O(m^2n^3)$	3

Table 3 compares the needed distance and the number of vehicles suggested by each algorithm (TS, GA, VNS, and RL–VNS) executed on test problem instances. The test results (Table 3) indicate the effectiveness of the RL–VNS approach.

TABLE 3
COMPARISON OF THE OBTAINED NUMBER OF VEHICLES AND DISTANCE FOR EACH ALGORITHM

	TS		**GA**		**VNS**		**RL–VNS**	
	No. of Vehicles	**Distance**	**No. of Vehicles**	**Distance**	**No. of Vehicles**	**Distance**	**No. of Vehicles**	**Distance**
R101	21	1692.39	16	1823.53	17	1835.23	14	1588.03
R103	19	1239.72	15	1739.34	16	1678.38	14	1448.48
R105	17	1428.56	15	1579.36	15	1542.21	13	1356.68
R107	16	1109.43	14	1452.14	14	1387.82	12	1128.44
R109	15	1087.15	12	1322.76	14	1235.60	11	1029.36
R111	12	1065.24	11	1236.17	13	1194.25	9	980.59

Fig. 4 presents the proposed traveled distances obtained by the algorithms. For non-electric vehicles, the emission values for each solution proposed by a specific algorithm provide an estimate of the pollution.

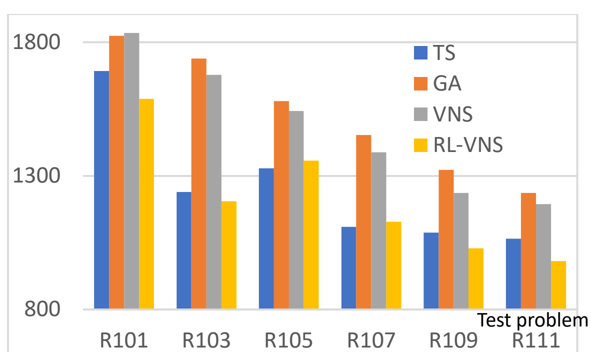

Fig. 4. Measure of traveled distances.

In contrast, Table 4 presents the needed execution time, in minutes, of the implemented algorithms. TS is executed once, Whereas GA, VNS, and RL–VNS are executed 20 times. Table 4 highlights that the

RL–VNS requires only slightly more time despite its efficiency. The results also show that TS and GA have comparable execution times.

TABLE 4
EXECUTION TIME OF THE ALGORITHMS

	TS	GA	VNS	RL–VNS
R101	77	78	82	86
R103	74	72	76	81
R105	69	69	68	73
R107	54	56	61	67
R109	43	48	53	53
R111	39	42	44	48

Regarding the convergence rate, VNS shows better objective values than the other algorithms for different numbers of iterations, as depicted in Fig. 5.

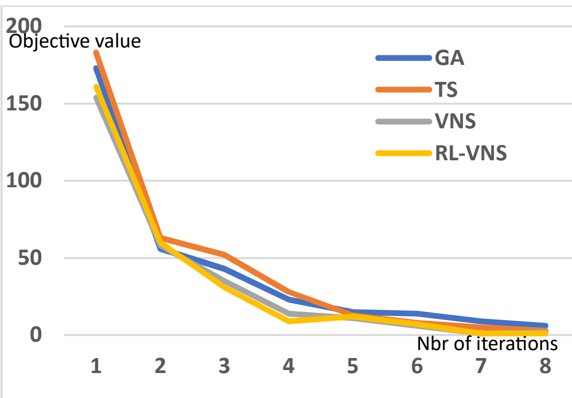

Fig. 5. Objective value of the tested algorithms for different numbers of iterations.

VI. DISCUSSIONS

The presented results and the analysis in the previous sections (Table 3, Table 4, Fig. 4, and Fig. 5) clearly demonstrate the advantages of hybridizing an optimization method such as VNS with a learning algorithm such as RL.

A hybrid approach combining VNS with RL for solving the VRP offers various potential benefits. VNS is a powerful optimization method for finding high-quality initial solutions in combinatorial problems such as the VRP. It can appropriately improve existing solutions and explore neighborhoods. RL is suitable for refining solutions through sequential decision-making and learning from interactions with the environment.

VNS immediately generates good initial solutions, which are used as a starting point for RL. Accordingly, the search space iteratively explored by the RL to refine the solutions is reduced.

Moreover, the dynamic aspect of the VNS, which allows the changing of neighborhoods during the search, enhances the capability of RL to avoid suboptimal solutions and local optima.

VII. CONCLUSION

Resolving the GVRP by simultaneously employing VNS and RL is a cutting-edge, complex research approach. This hybrid approach aims to optimize the routing of green vehicles, such as electric or hybrid vehicles, in a cost-effective and environmentally friendly manner.

In this regard, the success of a hybrid RL–VNS approach depends on various factors, such as the selected algorithms, the quality of the RL model, and the problem instance. Several perspectives and potential research directions exist for the use of RL and VNS for resolving the GVRP, such as their hybridization and complementarity with other paradigms and the scalability of large-scale transport problems.

VIII. REFERENCES

[1] P. Toth and D. Vigo, Eds. *The Vehicle Routing Problem*. Society for Industrial and Applied Mathematics, 2002.

[2] T. Ralphs, L. Kopman, W. Pulleyblank et al., "On the capacitated vehicle routing problem," *Math. Program., Ser. B*, vol. 94, pp. 343–359, 2003, doi: 10.1007/s10107-002-0323-0

[3] S. Mnasri, F. Abbes, K. Zidi, and K. Ghedira, "A multi-objective hybrid BCRC-NSGAII algorithm to solve the VRPTW," in *13th Int. Conf. Hybrid Intelligent Systems (HIS 2013)*, Gammarth, Tunisia, 2013, pp. 60–65, doi: 10.1109/HIS.2013.6920455

[4] PP. Ballesteros Silva and A. Escobar Zuluaga, "Review of state of the art vehicle routing problem with pickup and delivery (VRPPD)," *Ingeniería y Desarrollo*, vol. 34, no. 2, pp. 463–482, 2016.

[5] A. Oumachtaq, L. Ouzizi and M. Douimi, "Green Vehicle Routing Problem (GVRP): State-of-the-art," in *Advances in Integrated Design and Production II. CIP 2022. Lecture Notes in Mechanical Engineering*, Azrar, L. et al., Ed. Cham: Springer, 2023. doi: 10.1007/978-3-031-23615-0_42

[6] J. Brimberg, S. Salhi, R. Todosijević, and D. Urošević, "Variable neighborhood search: The power of change and simplicity," *Comput. Oper. Res.*, vol. 155, p. 106221, 2023.

[7] R. F. Prudencio, M. R. O. A. Maximo and E. L. Colombini, "A survey on offline reinforcement learning: Taxonomy, review, and open problems," in IEEE Trans. Neural Netw. Learn. Syst. doi: org/10.1109/TNNLS.2023.3250269

[8] S. Mnasri and M. Alrashidi, "Energy-efficient IoT routing based on a new optimizer," *Simul. Model.*

Pract. Theory, vol. 119, 2022, doi: org/10.1016/j.simpat.2022.102591

[9] W. Abdallah and T. Val, "Genetic-Voronoi algorithm for coverage of IoT data collection networks," in *2020 30th Int. Conf. Computer Theory and Applications (ICCTA)*, Alexandria, Egypt, 2020, pp. 16–22, doi: 10.1109/ICCTA52020.2020.9477675

[10] S. Mnasri, N. Nasri, and T. Val, "An overview of the deployment paradigms in the wireless sensor networks," in *Performance Evaluation and Modeling in Wireless Networks (PEMWN 2014)*, Sousse, Tunisia, Nov. 2014.

[11] A. S. Tarawneh, A. B. Hassanat, E. Alkafaween, B. Sarayrah, S. Mnasri, G. A. Altarawneh, M. Alrashidi, M. Alghamdi, and A. Almuhaimeed, "DeepKnuckle: Deep learning for finger knuckle print recognition," *Electronics*, vol. 11, no. 4, p. 513, 2022, doi: 10.3390/electronics11040513

[12] M. Aseeri, A. B. Hassanat, S. Mnasri, A. S. Tarawneh, K. Alhazmi, G. Altarawneh, M. Alrashidi et al., "Modelling-based simulator for forecasting the spread of COVID-19: A case study of Saudi Arabia," *Int. J. Comput. Sci. Netw. Secur.*, vol. 20, no. 10, Oct. 2020, doi: org/10.22937/IJCSNS.2020.20.10.16

[13] M. Alrashidi, "Social distancing in indoor spaces: An intelligent guide based on the Internet of Things: COVID-19 as a case study," *Computers*. vol. 9. No. 4, p. 91, 2020, doi: org/10.3390/computers9040091

[14] M. Alrashidi, K. Almohammadi, M. Gardner and V. Callaghan, "Making the invisible visible: Real-Time feedback for embedded computing learning activity using pedagogical virtual machine with augmented reality," in *Augmented Reality, Virtual Reality, and Computer Graphics. AVR 2017. Lecture Notes in Computer Science*, L. De Paolis, P. Bourdot, and A. Mongelli, Eds., vol. 10324, Cham: Springer, 2017, doi: 10.1007/978-3-319-60922-5_27

[15] G. A. Altarawneh, A. B. Hassanat, A. S. Tarawneh, A. Abadleh, M. Alrashidi, and M. Alghamdi, "Stock price forecasting for Jordan insurance companies amid the COVID-19 pandemic utilizing off-the-shelf technical analysis methods," *Economies*, vol. 10, no. 2, p. 43, 2022, doi: 10.3390/economies10020043

[16] Y. Gao, H. Wu, and W. Wang, "A hybrid ant colony optimization with fireworks algorithm to solve capacitated vehicle routing problem," *Appl. Intell.*, vol. 53, pp. 7326–7342, 2023, doi: 10.1007/s10489-022-03912-7

[17] J. Zhang, G. Wang, Q. Sheng, X. Jia, and P. Xie, "Multi-Depot heterogeneous vehicle routing optimization for hazardous materials transportation," *IEEE Access, IEEE Vehicular Technology Society Section*, 2023, doi: 10.1109/ACCESS.2023.3300041

[18] W. Pan and S. Q. Liu, "Deep reinforcement learning for the dynamic and uncertain vehicle routing problem," *Appl. Intell.*, vol. 53, pp. 405–422, 2023, doi: 10.1007/s10489-022-03456-w

[19] Y. Wang, D. Qiu, and G. Strbac, "Multi-agent reinforcement learning for electric vehicles joint routing and scheduling strategies," in *2022 IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Macau, China, Oct. 8–12, 2022.

[20] B. Lin, B. Ghaddar, and J. Nathwani, "Deep reinforcement learning for the electric vehicle routing problem with time windows," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, Aug. 2022.

[21] H. Ma, Y. Sheng, and W. Xia, "A pointer neural network for the vehicle routing problem with task priority and limited resources," *Inf. Technol. Control.*, vol. 49, no. 2, pp. 237–248, 2020, doi: 10.5755/j01.itc.49.2.24613

[22] Z. Iklassov, I. Sobirov, R. Solozabal and M. Takáč, "Reinforcement learning approach to stochastic vehicle routing problem with correlated demands," *IEEE Access*, vol. 11, pp. 87958–87969, 2023, doi: 10.1109/ACCESS.2023.3306076

[23] P. Kalatzantonakis, A. Sifaleras, and N. Samaras, "A reinforcement learning-variable neighborhood search method for the capacitated vehicle routing problem," *Expert Syst. Appl.*, vol. 213, Part A, 2023, doi: 10.1016/j.eswa.2022.118812

[24] W. K. Anuar, L. S. Lee, H.-V. Seow, and S. Pickl, "A multi-depot dynamic vehicle routing problem with stochastic road capacity: An MDP model and dynamic policy for post-decision state rollout algorithm in reinforcement learning," *Mathematics*, vol. 10, p. 2699, 2022, doi: 10.3390/math10152699

[25] G. Bono, "Deep multi-agent reinforcement learning for dynamic and stochastic vehicle routing problems," Ph.D. dissertation, Univ. Lyon, Lyon, France, 2020, NNT: 2020LYSEI096.

[26] B. Chen, R. Qu, R. Bai, and W. Laesanklang, "A reinforcement learning based variable neighborhood search algorithm for open periodic vehicle routing problem with time windows,"

[27] D. Ödling, "A metaheuristic for vehicle routing problems based on reinforcement learning," Degree project in mathematics, second cycle, 30 credits, Stockholm, Sweden, 2018.

[28] S. Mnasri, K. Zidi, and K. Ghedira. "A heuristic approach based on the multi-agents negotiation for the resolution of the DDBAP," in *4th Int. Conf. Metaheuristics and Nature Inspired Comput.*, Sousse, Tunisia, 2012.

[29] B. Ahmad, A. Hassanat, E. Alkafaween, N. A. Al-Nawaiseh, M. A. Abbadi, M. Alkasassbeh, and M. B. Alhasanat, "Enhancing genetic algorithms using multi mutations: Experimental results on the

travelling salesman problem," *Int. J. Comput. Sci. Inf. Secur.*, vol. 14, no. 7, Jul. 2016.

[30] S. Mnasri and M. Alrashidi, "A comprehensive modeling of the discrete and dynamic problem of berth allocation in maritime terminals," *Electronics*, vol. 10, no. 21, p. 2684, 2021, doi: 10.3390/electronics10212684

[31] M. Bernardo and J. Pannek, "Robust solution approach for the dynamic and stochastic vehicle routing problem," *J. Adv. Transp*, vol. 2018, doi: 10.1155/2018/9848104

[32] A. Aggarwal, F. Ho, and S. Nakadai, "Extended time dependent vehicle routing problem for joint task allocation and path planning in shared space," in *2022 IEEE/RSJ Int. Conf. Intell. Robots and Syst.*, Kyoto, Japan, Oct. 23–27, 2022, doi: 10.1109/IROS47612.2022.9981570

[33] S. Ramasamy, M. S. Mondal, J.-P. F. Reddinger, J. M. Dotterweich, J. D. Humann, M. A. Childers, and P. A. Bhounsule, "Solving vehicle routing problem for unmanned heterogeneous vehicle systems using asynchronous multi-agent architecture (A-teams)," in *2023 Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, Lazarski University, Warsaw, Poland, Jun. 6–9, 2023, doi: 10.1109/ICUAS57906.2023.10156585

[34] F. Liu, C. Lu, L. Gui, Q. Zhang, X. Tong, and M. Yuan, "Heuristics for vehicle routing problem: A survey and recent advances," 2023, arXiv: 2303.04147v1 [cs.AI]

[35] E. Boumpa, V. Tsoukas, V. Chioktour, M. Kalafati, G. Spathoulas, A. Kakarountas, P. Trivellas, P. Reklitis, G. Malindretos, "A review of the vehicle routing problem and the current routing services in smart cities," *Analytics*, vol. 2, pp. 1–16, 2023, doi: 10.3390/analytics2010001

[36] S.-Y. Tan, and W.-C. Yeh, "The vehicle routing problem: State-of-the-art classification and review. *Appl. Sci.* vol. 11, p. 10295, 2021, doi: 10.3390/app112110295

[37] S. Sabet and B. Farooq, "Green vehicle routing problem: State of the art and future directions," in *IEEE Access*, vol. 10, pp. 101622–101642, 2022, doi: 10.1109/ACCESS.2022.3208899

[38] M. Asghari, S. Mohammad, and J. M. Al-e-hashem, "Green vehicle routing problem: A state-of-the-art review," *Int. J. Prod. Economics*, vol. 231, 2021, doi: 10.1016/j.ijpe.2020.107899

[39] M. M. Solomon, "Algorithms for the vehicle routing and scheduling problems with time window constraints," *Oper. Res.*, vol. 35, pp. 254–265, 1987.