Multi-Objective Deterministic Ambulance Routing: Exact MILP Benchmarking and Metaheuristic Optimization

1st Amal Mansour
Institut Supérieur de Gestion (ISG)
Université de Tunis, LARODEC Laboratory
Tunis, Tunisia

2nd Marouene Chaieb
Institut Supérieur de Gestion (ISG)
Université de Tunis, LARODEC Laboratory
École Nationale des Sciences de l'Informatique (ENSI)
Tunis, Tunisia

3rd Issam Nouaouri
University of Artois, UR 3926,
Laboratoire de Génie Informatique et d'Automatique
de l'Artois, 62400, Béthune, France

4th Saoussen Krichen
Institut Supérieur de Gestion (ISG)
Université de Tunis, LARODEC Laboratory
Tunis, Tunisia

***Abstract*—Efficient ambulance routing is a critical component of healthcare logistics, where minimizing response times and ensuring timely service to high-priority patients directly impact clinical outcomes. This paper investigates the deterministic Ambulance Routing Problem (ARP) through a unified multi-objective optimization framework that jointly considers total travel distance, average response time, and critical coverage. Exact Mixed-Integer Linear Programming (MILP) formulations solved with IBM ILOG CPLEX are employed to obtain optimal benchmark solutions for small-scale instances, while population-based metaheuristics are used to address medium- and large-scale realistic Emergency Medical Services (EMS) scenarios. Computational experiments on artificial and real EMS data sets demonstrate the trade-offs between solution quality, computational effort, and scalability. The results confirm that exact optimization provides valuable benchmarks for small instances, whereas metaheuristics offer effective and scalable alternatives for complex multi-objective ambulance routing problems. The source code for this work is publicly available at: Multi_ARP.**

***Index Terms*—Ambulance Routing Problem, Healthcare Logistics, Multi-Objective Optimization, Mixed-Integer Linear Programming, Metaheuristics, Emergency Medical Services.**

I. Introduction

Emergency Medical Services are very important in lives saving, especially where time is very crucial like in cases of cardiac arrest, stroke, and severe trauma where speed is the key issue in saving a patient. It has been demonstrated that ambulance response delays substantially decrease chances of survival in a number of studies presented in the field of ARP survey research and EMS survey research [1], [2].

Due to various limitations, such as the availability of ambulances, the time of service at the site of emergencies, the priority of patients, and traffic, optimization of ambulance routes has become a key operational issue in the healthcare logistics field to date [1], [3]. The Ambulance Routing Problem is a complicated combinatorial optimization problem in the light of these constraints.

Exact optimization methods, most prominently MILP optimized with IBM ILOG CPLEX, can yield globally optimal solutions for deterministic ARP instances. However, their applicability is limited to small-scale cases due to the combinatorial explosion of computations [4], [5]. To handle medium- and large-scale cases, metaheuristic methods such as Genetic Algorithm (GA), Tabu Search (TS), Ant Colony Optimization (ACO), and Simulated Annealing (SA) are used, providing high-quality near-optimal solutions within reasonable computation times [6], [7], [8]. This paper explores the trade-offs between exact MILP solutions and metaheuristics on real EMS instances, highlighting operational implications.

The rest of this paper is structured in the following manner. In Section II, the related literature is conducted regarding ambulance routing and hybrid optimization. The mathematical formulation is found in Section III. The proposed hybrid methodology is described in section IV. Section V presents experimental findings, Section VI addresses implications and Section VII gives a conclusion to the paper with future research directions.

II. Literature Review

The literature is organized according to three complementary dimensions: (i) exact optimization approaches for deterministic ambulance routing, (ii) metaheuristic and hybrid approaches for large-scale or dynamic routing problems, and (iii) multi-objective optimization strategies balancing response time, travel efficiency, and emergency coverage.

ARP has been widely explored with diverse optimization methods in healthcare logistics, including both precise, and metaheuristic and learning-based algorithms. As summarized in Table I, prior studies provide insights on MILP, metaheuristics, and hybrid approaches, but gaps remain in benchmarking and large-scale applicability.

A. Exact Optimization Approaches

Deterministic Ambulance Routing Problems are often solved using exact Mixed-Integer Linear Programming (MILP) formulations, which model routing, ambulance availability, service times, and patient priorities. Solvers like CPLEX provide optimal solutions for small instances [9], e.g., a robust allocation–routing MILP integrating hospital location, casualty transport, and resource distribution while accounting for patient deterioration.

Although MILP models provide exact solutions, they become computationally intractable as problem size increases. Tlili et al. (2018) [4] highlighted this limitation in disaster-response ARP, where congestion and time constraints complicate optimization. Similarly, Chew (2025) [5] showed that exact methods struggle in dynamic urban environments with traffic uncertainty, motivating hybrid metaheuristics like CaputoGA [5] to deliver high-quality solutions within reasonable computation times. Exact MILP models remain essential as reference solutions for evaluating heuristic and metaheuristic methods. For instance, Orazani et al. (2025) [10] used an Augmented Epsilon Constraint approach to solve small-scale multi-objective EMS routing instances and benchmark NSGA-II and PESA-II on larger problems. While MILP provides optimal solutions, it is generally not scalable to medium- or large-scale real-world ARP cases.

B. Metaheuristic Methods

Metaheuristics are preferred for large-scale, dynamic ARP, as MILP are computationally limited. They provide high-quality solutions while handling multiple objectives, traffic dynamics, heterogeneous fleets, and patient prioritization.

Yan et al. (2023) [6] proposed a post-disaster location-routing model accounting for relief shortages and urgent demand, solved using a hybrid DPSO-HHO metaheuristic [6]. A Covid-19 case study in Wuhan demonstrated superior solution quality and robustness compared to classical metaheuristics.

CaputoGA [5] is a hybrid genetic algorithm for ambulance routing using Caputo fractional derivatives to model traffic. It integrates memory-based penalties, adaptive mutation, and hybrid local search (2-opt + swap), outperforming standard GA and other hybrids in route cost, robustness, and execution time, with a minimum deviation of 3.98%.

Xia et al. (2018) [11] studied the distance-constrained VRP with split deliveries using a bi-objective model minimizing travel distance and fleet size. Their adaptive tabu search with multi-neighborhood exploration, adaptive penalties, and infeasible-solution acceptance outperformed classical methods in solution quality and global search.

Tlili et al. (2017, 2024) [12], [13] applied hybrid metaheuristics to complex ARP. In mass-casualty scenarios, they combined the Petal Algorithm with PSO, achieving deviations lower than 5.5%. For heterogeneous ARP with time windows, Hybrid Tabu Search and a Tabu Search-based Hyper-Heuristic (TSHH) distinguished green/red cases and considered fleet heterogeneity, with TSHH yielding the best solutions in 83.3% of instances.

Javidaneh et al. (2025) [8] used ARP as a VRP, with the locations of injuries and the capacities of hospitals taken into account, as well as travel time. Implementation in MATLAB has shown that ACO is an effective algorithm to find close optimal ambulance routes to real-life NP-hard problems.

As reported by Schjolberg et al. (2023) [3], the article is based on comparing genetic, stochastic local search, and memetic algorithms to allocate ambulances in Oslo and Akershus (2015-2019). Metaheuristics performed better than basic baseline models, where the genetic algorithm worked better where high-demand conditions were encountered.

Genetic Programming Guided Local Search (GPGLS) was suggested by Liu et al. (2024) [14] to solve large scale vehicle routing problems involving more than 200 customers. GPGLS extends utility functions to provide local search, which enhance the ability to leave local optima, and performs better than the traditional Knowledge Guided Local Search (KGLS) on benchmark problems.

Existing studies use diverse instances, including synthetic datasets, simulated environments, and real EMS data, which hinders direct comparison of optimization methods due to differences in objectives, fleet settings, and demand distributions.

C. Multi-Objective Optimization

Rabbani et al. (2022) [15] modeled post-disaster ARP as a multi-objective MILP with Red/Green triage, deterioration rules, and hospital capacity. NSGA-II and MOPSO generated non-dominated solutions, capturing congestion and resource scarcity, though scalability is limited.

In this direction of research,Yan et al. (2023) [6] proposed a post-disaster location-routing model accounting for relief shortages and demand urgency. They solved this NP-hard problem with a hybrid DPSO-HHO metaheuristic using learning and exploration strategies. A COVID-19 case study in Wuhan showed it outperforms classical metaheuristics in solution quality and robustness.

Ambulance Routing Problems (ARP) are complex, involving dispatching, routing, relocation, and hospital assignment. Chaieb et al. (2015) [16] showed that treating these decisions in isolation can be suboptimal, motivating unified optimization models that capture interdependencies in EMS routing.

These papers collectively reveal the usefulness of population-based metaheuristics including NSGA-II, MOPSO, and hybrid DPSO-HHO when applied to complex multi-objective emergency routing problems in generating Pareto-optimal solutions.

D. Research Gap

Despite extensive research on MILP and metaheuristics for ARP, gaps remain. Exact methods like CPLEX provide small-instance benchmarks [9], [10] but are intractable for larger cases, while metaheuristics handle large, dynamic problems [5], [6], [13] without systematic benchmarking. This study adopts a hybrid approach, using a unified MILP as a reference while applying metaheuristic and learning-based methods to large-scale ARP.

TABLE I
DETAILED SUMMARY OF ARP LITERATURE AND OUR CONTRIBUTION

Study / Author	**Method**	**Key Contributions**
Zhang & Liu (2025) [9]	MILP	Robust allocation-routing MILP under uncertain demand; integrates hospital location, casualty transport, and trauma deterioration
Orazani et al. (2025) [10]	MILP (Augmented Epsilon Constraint)	Solves small multi-objective EMS routing; benchmarks for metaheuristics NSGA-II and PESA-II
Tlili et al. (2018) [4]	MILP	Disaster-response ARP modeling considering traffic congestion
Chew (2025) [5]	Hybrid GA (CaputoGA)	Memory-aware penalty, adaptive mutation, hybrid local search; handles urban traffic
Xia et al. (2018) [11]	Adaptive Tabu Search	Multi-neighborhood search; bi-objective VRP with split deliveries; better global search
Tlili et al. (2017) [12]	PA-PSO	Mass-casualty ARP; hybrid cluster-first route-second framework
Tlili et al. (2024) [13]	Hybrid Tabu Search + Hyper-Heuristic	Handles Heterogeneous Ambulance Routing with time windows; distinguishes on-site vs hospital transport
Javidaneh et al. (2025) [8]	Ant Colony Optimization (ACO)	Efficient generation of near-optimal ambulance routes in NP-hard problems
Schjolberg et al. (2023) [3]	GA, SLS, Memetic	Compared metaheuristics for ambulance allocation; GA performs best under high demand
Liu et al. (2024) [14]	GPGLS	Evolves utility functions for local search; escapes local optima; improves KGLS
Rabbani et al. (2021) [15]	Multi-Objective MILP	Post-disaster ARP with patient deterioration rules; NSGA-II and MOPSO used for non-dominated solutions
Yan et al. (2023) [6]	DPSO-HHO	Multi-objective emergency logistics; considers relief shortages and fairness; hybrid metaheuristic

III. MATHEMATICAL MODELING

In this section, the mathematical formulation of the problem is proposed. It is a common paradigm used across all resolution models, such as exact optimization, metaheuristics and learning-based models. The proposed MILP formulation is inspired by classical vehicle routing and EMS optimization models from [9], [10], [15], while integrating multiple operational objectives within a unified deterministic framework.

A. Problem Definition

The considered deterministic ARP assumes a single central depot from which all ambulances depart and to which they return after completing their assigned routes. Each ambulance may serve multiple emergency requests within a single route while respecting route duration and response-time constraints. The Deterministic Ambulance Routing Problem (DARP) seeks optimal ambulance routes to serve a set of emergency calls, with all ambulances starting and ending at depots. Travel and service times are deterministic and known in advance.

The multi-objective formulation aims to:

- Minimize total travel time,
- Minimize average (or maximum) response time,
- Maximize coverage within critical time windows.

These objectives conflict: faster response and higher coverage can require longer routes, while minimizing travel may delay high-priority emergencies, motivating a multi-objective approach balancing efficiency, responsiveness, and patient prioritization.

B. Notation

- V : set of nodes (depots, emergency request locations, hospitals)
- $E \subset V$: set of emergency request locations
- K : set of available ambulances
- d_{ij} : deterministic travel time between nodes i and j
- s_i : service time at node i
- $x_{ij}^k \in \{0, 1\}$: 1 if ambulance k travels from i to j, 0 otherwise
- y_j : arrival time of the ambulance at emergency j
- z_k : total route duration of ambulance k
- $u_j \in \{0, 1\}$: 1 if emergency j is served within a critical time threshold, 0 otherwise

C. Objective Functions

The ARP is formulated as a multi-objective MILP with the following goals:

$$\text{Minimize } Z_1 = \sum_{k \in K} \sum_{i \in V} \sum_{\substack{j \in V \\ i \neq j}} d_{ij}\, x_{ij}^k, \tag{1}$$

$$\text{Minimize } Z_2 = \frac{1}{|E|} \sum_{j \in E} y_j, \tag{2}$$

$$\text{Maximize } Z_3 = \sum_{j \in E} u_j. \tag{3}$$

Z_1 minimizes total ambulance travel, Z_2 maximizes service coverage, and Z_3 prioritizes urgent patients within critical time windows. Together, they form the multi-objective MILP for ARP, balancing efficiency, coverage, and patient urgency.

To solve the multi-objective formulation using both exact optimization and metaheuristic approaches, the three objectives are aggregated into a single weighted objective function using a normalized weighted-sum strategy:

$$Z = w_1 \frac{Z_1}{Z_1^{max}} + w_2 \frac{Z_2}{Z_2^{max}} - w_3 \frac{Z_3}{Z_3^{max}} \tag{4}$$

where w_1, w_2, and w_3 are weighting coefficients satisfying $w_1 + w_2 + w_3 = 1$. In this study, equal weights were initially adopted ($w_1 = w_2 = w_3 = \frac{1}{3}$) to ensure balanced consideration of operational efficiency, response rapidity, and emergency coverage. Normalization is applied to avoid scale dominance between objectives with different magnitudes.

The proposed optimization framework adopts a scalar aggregation strategy based on the weighted-sum approach rather than Pareto dominance optimization. This choice allows unified evaluation of exact optimization and metaheuristic algorithms under a common fitness structure while maintaining computational tractability for large-scale EMS instances.

D. Constraints

1) **Each emergency visited exactly once:**

$$\sum_{k\in K}\sum_{i\in V, i\neq j} x_{ij}^k = 1, \quad \forall j \in E \tag{5}$$

2) **Flow conservation for each ambulance:**

$$\sum_{j\in V, j\neq i} x_{ij}^k = \sum_{j\in V, j\neq i} x_{ji}^k, \quad \forall i \in V, k \in K \tag{6}$$

3) **Start and end at depot:**

$$\sum_{j\in V, j\neq depot} x_{depot,j}^k = 1, \quad \forall k \in K \tag{7}$$

$$\sum_{i\in V, i\neq depot} x_{i,depot}^k = 1, \quad \forall k \in K \tag{8}$$

4) **Maximum route duration per ambulance:**

$$z_k = \sum_{i\in V}\sum_{j\in V, j\neq i} d_{ij} x_{ij}^k \leq \text{MaxTime}_k, \quad \forall k \in K \tag{9}$$

5) **Arrival times / time windows including service time:**

$$y_j \geq y_i + s_i + d_{ij} - M(1 - x_{ij}^k), \quad \forall i \in V, j \in E, k \in K \tag{10}$$

6) **Coverage of emergencies within critical time:**

$$u_j = \begin{cases} 1 & \text{if } y_j \leq \text{CriticalTime}_j \\ 0 & \text{otherwise} \end{cases}, \quad \forall j \in E \tag{11}$$

7) **Binary / integer restrictions:**

$$x_{ij}^k \in \{0, 1\}, \quad \forall i, j \in V, k \in K \tag{12}$$

$$u_j \in \{0, 1\}, \quad \forall j \in E \tag{13}$$

E. Model-Based Integration of Metaheuristic Approaches

All of the solution paradigms have a structural basis in the MILP formulation. Candidate solutions represent feasible (or near-feasible) assignments of x_{ij}^k. The objective functions are considered in line with the MILP and the constraints are imposed through feasibility-preserving representations, penalty functions, or repair functions. For all metaheuristics, a solution represents a complete set of ambulance routes, where each chromosome, state, or path encodes feasible assignments of emergencies to ambulances. In GA, chromosomes define ordered sequences and are evolved via crossover and mutation. TS uses swap and relocation moves under tabu restrictions, while SA accepts worse solutions probabilistically to escape local optima. ACO builds routes incrementally using pheromone levels and heuristic information based on distance and priority. Constraint violations are managed through penalty and repair mechanisms within the fitness evaluation.

F. Solving Approach

Small-scale instances are solved optimally using CPLEX, providing benchmark solutions. Metaheuristics: Medium- and large-scale instances are solved with GA, TS, PSO, and ACO. Solutions respect MILP constraints and objectives. The strategy guarantees the high quality of solutions and flexibility in operation in various ARP situations, which enables systematic comparisons between solution paradigms.

IV. METHODOLOGY

This section presents the solution paradigms for the unified model in Section III, considering problem size, computational requirements, and operational constraints.

We adopt a two-phase optimization for the deterministic ambulance routing problem (Figure 1):

- Phase 1 – Exact Optimization: MILP solved with IBM ILOG CPLEX on small-scale instances (≤ 10 emergencies) to obtain global optimal solutions as benchmarks.
- Phase 2 – Metaheuristic Optimization: Medium- and large-scale instances from real EMS databases (e.g., Mashhad, NEMSIS) are solved to evaluate solution performance and realism.

This approach ensures both solution quality and computational tractability across diverse problem sizes.

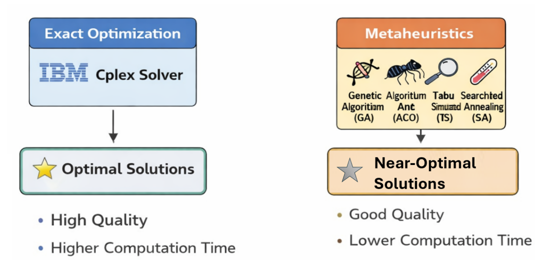

Fig. 1. Two-phase optimization strategy for the deterministic ARP

A. Exact Optimization Using CPLEX

For small-scale ARP instances, the MILP model is solved optimally with IBM ILOG CPLEX, producing global solutions in negligible time (0.06 s) with low average response times and full critical coverage. Due to the CPLEX Community Edition limits ($\leq 1,000$ variables/constraints), exact optimization becomes infeasible for medium- and large-scale instances, motivating metaheuristic approaches.

B. Metaheuristic Approaches

For medium- and large-scale ARP, metaheuristics efficiently explore feasible solutions while respecting MILP constraints, including service windows, maximum route time, and exclusive assignment of emergencies.

The metaheuristic framework includes:

1) **Dataset preparation:** Instances are built from a real EMS dataset of a large teaching hospital in Mashhad, Iran [17], including triage grades, critical status, and service times. Spatial locations are estimated, and distances computed using Euclidean metrics.
2) **Solution representation:** Each candidate encodes routes for all ambulances, visiting each emergency exactly once, starting and ending at the depot.
3) **Objective evaluation:** Multi-objective fitness function considers total travel time, average response time, and critical coverage; constraint violations are penalized.
4) **Operators and neighborhood moves:** Includes intra-route swaps, inter-route relocations, and exchanges depending on the metaheuristic.
5) **Metaheuristics implemented:**

- Genetic Algorithm (GA): Population of solutions evolves by means of selection, crossover and mutation with the help of the fitness function [5]–[7].
- Tabu Search (TS): Uses neighbors to optimize a solution by generating a solution and avoiding repetition by some sort of tabu list [11], [13], [18].
- Simulated Annealing (SA): Accepts worse solutions probabilistically to escape local optima [6].
- Ant Colony Optimization (ACO): Constructs solutions using pheromone trails and heuristic information on distances and emergency priority [8], [12].

6) **Stopping criteria:** Termination occurs when maximum iteration, time limit, solution convergence is reached.

This framework enables effective solution of large EMS and provides near-optimal routing for real-world operations.

V. Experimental Results

A. Experimental Setup

The suggested techniques are experimented on artificial and real EMS networks. The cases differ in terms of calling number of emergency requests, the number of ambulances, and size of the network. Exact optimization used small-scale instances (up to 10-15 calls) whereas medium- and large-scale instances (up to 3,000 calls) are used to measure metaheuristic performance. Exact optimization is applied only to small-scale instances due to the limitations of the CPLEX Community Edition (restricted to 1,000 variables and constraints), while metaheuristic methods are employed for larger-scale cases.

B. Instance Characterization

The experiments include synthetic and real EMS-inspired instances. Small cases (10–15 requests) are used for exact benchmarking, while medium and large cases (up to 3,000 requests) are derived from the Mashhad EMS dataset with heterogeneous urgencies and service times. Distances are computed using Euclidean measures, and results are averaged over 10 runs per metaheuristic.

C. Performance Metrics

Performance is evaluated using the following metrics:

- Total Distance (km): Sum of travel distances of all ambulances.
- Average Response Time (min): Mean arrival time.
- Critical Coverage (calls): Number of emergencies served within the critical time threshold.
- Compute Time (s): Time required to generate solutions.

D. Results and Analysis

1) Exact Optimization Results: ARP is a multi-objective MILP which is solved precisely on IBM ILOG CPLEX. Ambulances have a central depot as starting and ending point and emergency points as demand node with service time, and critical response limitations.

The optimization is obtained by minimizing the total travel and the average response time and maximizing the critical coverage. In the case of small-scale problems, CPLEX would ensure global optimality at extremely short run times.

2) Metaheuristic Results on Real EMS Instances: For small instances, TS provides the best compromise, achieving full critical coverage (10 calls), low total distance (24.74 km), and a low average response time (12.61 min), as shown in Table II. GA and SA yield higher response times (26.91 and 30.33 min, respectively) but require negligible computation time (≤ 0.01 s). ACO also ensures full coverage, though with moderate distance (51.49 km) and response time (18.96 min).

For medium and large instances (3000 calls), Table III shows that GA performs best in terms of critical coverage (172 calls) while maintaining reasonable distance (24,579.27 km) and response time (508.35 min). TS and ACO incur slightly higher distances with reduced coverage, whereas SA achieves the lowest computation time (0.13 s) but serves fewer critical calls. Overall, metaheuristics are well suited for large-scale EMS routing, while small instances can be optimally solved by CPLEX to provide benchmark solutions.

TABLE II
Performance Metrics of Metaheuristics on Small Instances

Instance	Method	Total Distance (km)	Avg Response (min)	Critical Coverage (calls)	Comp.Time (s)
Small	GA	67.38	26.91	9	0.0024
	TS	24.74	12.61	10	0.0186
	SA	63.16	30.33	8	0.0071
	ACO	51.49	18.96	10	0.0093

TABLE III
Performance Metrics of Metaheuristics on EMS Medium/Large Instances

Instance	Method	Total Distance (km)	Avg Response (min)	Critical Coverage (calls)	Comp.Time (s)
Medium/Large	GA	24,579.27	508.35	172	0.91
	TS	25,035.36	516.23	154	2.66
	SA	25,338.36	520.79	153	0.13
	ACO	24,888.31	525.17	155	0.93

VI. Discussion

A. Internal Analysis

This study compares MILP and metaheuristics within a unified framework, using CPLEX for small instances and metaheuristics for medium and large real EMS cases. For small problems, CPLEX provides optimal solutions in 0.06 s with low response time (11.17 min) and full coverage, but becomes infeasible for larger instances, motivating heuristic approaches. For larger cases, metaheuristics efficiently manage complexity but show trade-offs between distance, response time, and coverage, as shown in Table IV. Overall, results emphasize the multi-objective nature of ARP and the need to balance efficiency and clinical priorities.

Several insights emerge:

- Exact Optimization: CPLEX guarantees optimal solutions with minimal response times and full critical coverage, but is limited to small instances.
- Metaheuristics: TS and ACO achieve strong compromises in distance and critical coverage, while GA and SA provide more diverse but less efficient solutions.
- Trade-offs: Metaheuristics scale effectively to large instances, offering a practical balance between solution quality, coverage, and operational feasibility.

TABLE IV
PERFORMANCE OF CPLEX AND METAHEURISTICS ON SMALL EMS INSTANCES

Instance	Method	Total Distance (km)	Avg Response (min)	Critical Coverage (calls)	Comp.Time (s)
Small	CPLEX	21.13	11.17	10	0.06
	GA	67.38	26.91	9	0.0024
	TS	24.74	12.61	10	0.0186
	SA	63.16	30.33	8	0.0071
	ACO	51.49	18.96	10	0.0093

Operational Insights and Patient Prioritization

Patient urgency is modeled via critical coverage and response windows. GA maximizes urgent-case coverage, TS and SA minimize response times, and ACO balances efficiency and timeliness, highlighting the need to align algorithm choice with EMS priorities.

B. External Analysis

Literature shows that exact MILP methods provide benchmarks and theoretical guarantees for small-scale or strategic instances [4], [9], [10], while metaheuristics efficiently generate high-quality solutions for complex or large-scale problems [3], [5], [8], [11]. Multi-objective approaches capture trade-offs between response time, travel distance, and critical coverage [6], [15], and learning-based strategies enhance adaptability under uncertainty [12], [14].

We compare exact MILP solutions (CPLEX) with metaheuristics (GA, SA, TS, ACO) on real EMS instances, highlighting trade-offs between solution quality, response time, and critical coverage. Metaheuristics offer flexible, scalable solutions, while MILP provides optimal benchmarks for small instances [15]. The unified MILP also acts as a structural safeguard, ensuring priority constraints are respected.

Future extensions may also investigate Pareto-based optimization strategies such as the augmented ϵ-constraint method to better analyze trade-offs among conflicting EMS objectives. In addition, exploring hybrid frameworks that combnie MILP and metaheuristics to leverage exact optimization for subproblems while scaling efficiently to realistic EMS scenarios. Moreover, automated hyperparameter optimization and configurators may be incorporated to improve robustness and reproducibility of metaheuristic performance across heterogeneous EMS scenarios.

VII. CONCLUSION

This paper addressed the deterministic Ambulance Routing Problem using a unified multi-objective optimization framework that captures key trade-offs between travel efficiency, response time, and critical emergency coverage. Exact MILP solutions obtained with CPLEX were used as optimal benchmarks for small-scale instances, while population-based metaheuristics were applied to realistic EMS scenarios to ensure scalability under operational constraints. The results confirm that no single method optimizes all objectives simultaneously. Metaheuristics exhibit complementary behaviors, with some favoring rapid response and critical coverage and others emphasizing route efficiency, highlighting the importance of multi-objective modeling for EMS decision support. Future work will extend this framework to dynamic and stochastic settings, incorporate real-time traffic and learning-based adaptation, and explore reinforcement learning approaches for adaptive ambulance routing and scheduling.

REFERENCES

[1] J. Tassone and S. Choudhury, "A comprehensive survey on the ambulance routing and location problems," *arXiv preprint arXiv:2001.05288*, 2020.

[2] F. Da Ros, L. Di Gaspero, K. Roitero, D. La Barbera, S. Mizzaro, V. Della Mea, F. Valent, and L. Deroma, "Supporting fair and efficient emergency medical services in a large heterogeneous region," *Journal of Healthcare Informatics Research*, vol. 8, no. 2, pp. 400–437, 2024.

[3] M. E. Schjølberg, N. P. Bekkevold, X. Sánchez-Díaz, and O. J. Mengshoel, "Comparing metaheuristic optimization algorithms for ambulance allocation: An experimental simulation study," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, 2023, pp. 1454–1463.

[4] T. Tlili, S. Abidi, and S. Krichen, "A mathematical model for efficient emergency transportation in a disaster situation," *American Journal of Emergency Medicine*, vol. 36, no. 9, pp. 1585–1590, 2018.

[5] J. V. L. Chew, "Hybrid genetic algorithm with caputo fractional derivative for ambulance routing problems," in *Multi-disciplinary Trends in Artificial Intelligence (MIWAI 2025)*, ser. LNCS, 2025, pp. 317–327.

[6] T. Yan, F. Lu, S. Wang, L. Wang, and H. Bi, "A hybrid metaheuristic algorithm for the multi-objective location-routing problem in the early post-disaster stage," *Journal of Industrial and Management Optimization*, vol. 19, no. 6, pp. 4663–4691, 2023.

[7] J. A. Nasir, Y.-H. Kuo, and R. Cheng, "Clustering-based iterative heuristic framework for a non-emergency patients transportation problem," *Journal of Transport & Health*, vol. 26, p. 101411, 2022.

[8] A. Javidaneh, M. Ataee, and A. A. Alesheikh, "Ambulance routing with ant colony optimization," in *Multi-disciplinary Trends in Artificial Intelligence*, 2025.

[9] T. Zhang and Y. Liu, "A robust optimization model for allocation-routing problems under uncertain conditions," *PLoS ONE*, vol. 20, no. 5, p. e0322483, 2025.

[10] S. M. H. Orazani, A. Heidari, M. Khalilzadeh, and F. Jolai, "A multi-objective mathematical model for emergency and medical routing to provide healthcare services to victims," *Journal of Mathematics & Statistics*, 2025.

[11] Y. Xia, Z. Fu, L. Pan, and F. Duan, "Tabu search algorithm for the distance-constrained vehicle routing problem with split deliveries by order," *PLoS ONE*, vol. 13, no. 5, p. e0195457, 2018.

[12] T. Tlili *et al.*, "Swarm-based approach for solving the ambulance routing problem," *Procedia Computer Science*, vol. 112, pp. 350–357, 2017.

[13] ——, "Tabu search-based hyper-heuristic for solving the heterogeneous ambulance routing problem with time windows," *International Journal of Intelligent Transportation Systems Research*, vol. 22, no. 2, pp. 446–461, 2024.

[14] S. Liu, J. G. C. Costa, Y. Mei, and M. Zhang, "Gpgls: Genetic programming guided local search for large-scale vehicle routing problems," in *Parallel Problem Solving from Nature (PPSN XVIII)*, ser. LNCS, 2024, pp. 36–51.

[15] M. Rabbani *et al.*, "Ambulance routing in disaster response considering variable patient condition: Nsga-ii and mopso algorithms," *Journal of Industrial and Management Optimization*, 2021.

[16] M. Chaieb, J. Jemai, and K. Mellouli, "On the relationships between sub problems in the hierarchical optimization framework," in *Proceedings of the 28th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE)*. Springer, Lecture Notes in Computer Science, 2015. [Online]. Available: https://doi.org/10.1007/978-3-319-19066-2_23

[17] Z. BaniHassan, M. Kazemi, M. Jangi, and H. Tabesh, "An open-access dataset of emergency department admissions at a large teaching hospital in iran," *Data in Brief*, vol. 56, p. 110827, 2024.

[18] Y. A. Kochetov and N. B. Shamray, "Optimization of the ambulance fleet location and relocation," *Journal of Applied and Industrial Mathematics*, vol. 15, pp. 234–252, 2021.