Journal of Mathematical Modeling
Vol. -, No. -, 2026, pp. -. Research Article

JMM

A bi-level optimization model for an ambulance routing problem for green, red, and black patients in a post-disaster stage

Raheleh Khanduzi[†*]

[†] *Department of Mathematics and Statistics, Faculty of Basic Sciences and Engineering, Gonbad Kavous University, Gonbad-e Kavus, Iran*

Email(s): khanduzi@gonbad.ac.ir

Abstract. In post-disaster environments, effective allocation and routing of ambulances is crucial to minimize casualties and improve overall emergency response efficiency. This paper develops a novel bi-level programming model to address the ambulance routing problem with triage-based patient categorization, including green, red, and black patients. The upper level focuses on strategic decisions regarding ambulance allocation and dispatching, while the lower level models operational routing decisions performed by responders. The proposed approach integrates triage priorities, limited resources, and road network disruptions, yielding a realistic framework for decision support. A hybrid solution methodology based on Genetic algorithm, tabu search, and teaching learning based optimization is presented. Experimental results on test instances from existing literature demonstrate the model's capability to balance response time efficiency and prioritization of critical patients.

Keywords: Bi-level optimization, ambulance routing, priority of patients, metaheuristics, hybrid solution approach.
AMS Subject Classification 2010: 90B10, 90C59.

1 Introduction

Natural disasters, such as earthquakes, floods, hurricanes, and industrial accidents, create chaotic and resource-constrained environments in which emergency medical response becomes a critical determinant of survival rates. Immediately following a disaster, emergency medical services (EMS) face overwhelming demand, infrastructural damage, uncertain information, and severe time pressure. Among the various tasks in disaster management, the allocation and routing of ambulances are among the most important, as timely medical attention significantly reduces mortality and long-term injury severity. In these

[*]Corresponding author
Received: 16 December 2025/ Revised: 22 February 2026/ Accepted: 25 February 2026
DOI: 10.22124/jmm.2026.32560.2959

high-stakes conditions, decision-makers must determine how best to deploy limited ambulances across multiple affected sites while ensuring that patients with varying levels of seriousness receive appropriate and timely care.

A central component of disaster medical response is triage process of classifying patients based on injury severity and the urgency of required treatment. Classical triage classes typically include red (critical and time-sensitive injuries), green (minor injuries that do not require immediate intervention), and black (deceased or beyond recovery). Incorporating these categories into routing strategies is essential because ambulances must prioritize patients who can benefit most from immediate treatment. However, integrating triage-based priorities introduces complexity, particularly in large-scale disasters where the distribution of patients is heterogeneous and dynamic. Disaster sites may contain a mix of patients from all three groups, and the number and severity distribution of patients may change rapidly as new information becomes available.

In addition to triage complexities, disasters often degrade transportation infrastructure. Roads may be partially or fully blocked, travel times may increase unpredictably, and access routes to affected sites may be uncertain or unstable. These disruptions further complicate routing decisions, requiring robust and adaptive optimization approaches. Traditional vehicle routing problem (VRP) frameworks are insufficient in such scenarios because they generally assume a stable environment, homogeneous service requirements, and fixed travel times. In contrast, post-disaster routing must account for uncertainty, dynamic road conditions, and heterogeneous patient needs.

The hierarchical nature of EMS decision-making also motivates the use of bi-level optimization. In real-world settings, EMS coordination typically involves two layers of decision-making: (1) a strategic layer in which high-level decisions are made determining which sites to serve, how to allocate ambulances, and how to set triage priorities and (2) an operational layer in which routed ambulances perform detailed navigation and patient extraction. These two layers are interdependent: strategic decisions shape operational feasibility, while operational constraints influence optimal strategic planning. Bi-level programming provides a natural mathematical structure for capturing this hierarchy, enabling integrated optimization of strategic allocation and tactical routing.

Despite the importance of these challenges, relatively few studies explicitly modeled ambulance routing within a bi-level framework that incorporates triage categories and post-disaster uncertainties. Existing models tend to treat ambulance allocation and routing as separate problems or overlook triage complexities and infrastructure disruptions. This limits their applicability in real-world disaster scenarios, where combined decision-making and heterogeneous patient needs are essential considerations.

To address these gaps, this paper proposes a comprehensive bi-level programming model for ambulance routing in post-disaster environments with triage-based patient prioritization. The model simultaneously considers strategic allocation of ambulances to disaster sites and operational routing decisions under disrupted transportation networks. The upper level focuses on minimizing weighted unmet demand, emphasizing the rescue of red patients while still accounting for green and black categories. The lower level optimizes routing decisions to minimize total travel time while satisfying capacity limits, triage policies, and infrastructure constraints.

The contributions of this work can be summarized as follows:

- We develop a novel bi-level mathematical formulation that integrates triage classes, ambulance allocation, and routing under disaster conditions.
- We incorporate infrastructure disruptions and heterogeneous patient distributions into an integrated

decision-making framework.

- We propose a practical solution strategy using Genetic algorithm (GA), tabu search (TS), teaching learning based optimization (TLBO), and CPLEX solver to handle the NP-hardness inherent in bi-level programs.

- We evaluate the proposed model on realistic disaster-inspired scenarios, demonstrating improved performance relative to single-level and traditional VRP approaches.

The remainder of the paper is organized as follows. Section 2 reviews relevant literature on EMS routing, triage-based optimization, and bi-level programming. Section 3 presents the mathematical description of the proposed bi-level model. Section 4 describes the hybrid solution methodology, while Section 5 reports computational experiments and results. Section 6 concludes with a discussion of findings and directions for future research.

2 Related work

Research on ambulance routing, EMS optimization, and disaster management has grown significantly over the past decades. This section reviews the major strands of literature relevant to the present work, including classical EMS vehicle routing, triage-based models, post-disaster logistics, and bi-level optimization for emergency response. While many of these issues have been studied separately, few attempts have been made to integrate them into an integrated modeling framework. This motivates the development of a bi-level programming model explicitly designed for triage-based ambulance routing under disaster conditions.

2.1 Ambulance routing and EMS logistics

Ambulance routing has traditionally been studied within the context of the VRP, with early work focusing on minimizing response time or maximizing system coverage. Studies such as Goldberg [16] and Brotcorne et al. [6] analyzed EMS systems using queueing models and VRP variants, emphasizing response time performance. More recent contributions incorporate dynamic travel times and real-time information streams (Schmid [33]; Andersson and Värbrand [2]). However, while these models address operational routing, they generally assume stable road networks and homogeneous patient demands—conditions rarely met in post-disaster environments.

2.2 Triage-based optimization in medical response

Triage prioritization is a critical component of post-disaster medical response. Several studies incorporate triage categories into resource allocation or evacuation planning (Bazyar et al. [3]; Ro et al. [30]). These works highlight the importance of prioritizing red (critical) patients, but most do not integrate routing decisions or consider interactions between ambulance allocation and route planning. A notable exception is the work of Najafi et al. [23], who develop a triage-aware evacuation model. Nevertheless, their model does not adopt a bi-level structure nor explicitly incorporate road network disruptions.

2.3 Post-disaster transportation and medical logistics

Disasters create unique operational conditions characterized by infrastructure damage, uncertain road availability, and unequal spatial distributions of casualties. Several researchers have studied logistics under such constraints. For example, Chang et al. [8] examined transportation network reliability under earthquake scenarios, while Holguín-Veras et al. [18] analyzed humanitarian logistics with infrastructure disruption. In medical logistics, Yi and Özdamar [42] proposed a multi-commodity flow model for delivering medical supplies in disaster zones. Although these studies acknowledge infrastructure disruption, they primarily address supply distribution rather than ambulance routing.

2.4 Multi-objective and bi-level optimization in disaster and emergency response

Recent work in ambulance routing and emergency logistics underscores the complexity of multi-objective decision contexts. For example, recent studies have explored ambulance routing optimization with enhanced service availability and emergency selection criteria, highlighting the growing interest in integrating multiple objectives in EMS routing models. For example, Sarma et al. [32] introduced a multi-objective mathematical model for post-disaster relief logistic network. Gharib et al. [13] proposed a multi-objective stochastic model in a post-disaster environment. Yan et al. [40] developed a multi-objective location-routing problem in the post-disaster stage. Ye et al. [41] presented a multi-objective model for the emergency materials response after petrochemical enterprises.

Bi-level optimization has been used to model hierarchical decision structures in emergency management. Typical applications include evacuation planning (Lv et al. [22] and Liu et al. [21]), emergency facility location (Saghehei et al. [31]), and transportation network design (Farahani et al. [11]). These works demonstrate the suitability of bi-level models for capturing strategic and operational decision interactions. However, applications to EMS routing remain limited. Some recent studies, such as Camacho-Vallejo et al. [7], Chen et al. [9], Gao [12] and Wang et al. [39], applied bi-level frameworks to disaster relief logistics, but none consider triage-based ambulance routing.

2.5 Methodological gaps

Despite the development in each of these fields, there is a notable and remarkable gap in integrated models that simultaneously address hierarchical decision-making, triage classes, ambulance allocation, and routing under disaster conditions. Few papers generically capture the interplay between ambulance assignments and operational routing This research addresses this gap by proposing a novel bi-level optimization model that integrates strategic allocation and tactical routing. A metaheuristic-based exact approach is proposed to efficiently solve this complex problem, proposing practical solution for ambulance allocation and route planning.

The concepts of hierarchical decision-making, prioritization of critical patients, allocation and routing of ambulances and capacity and infrastructure limits are novel opinions in EMS systems and have been provoked by experiences in the case of a disaster event. Allocating and routing ambulances are two major operation for the future planning of EMS system and the hospital assignment development planning in countries. Through the development of ambulance routing problem, attention can be paid to the possible impacts of patient priorities, ambulance allocation and operational routing and how best to design the EMS system in front of such hierarchical decision structure. In spite of the fact that classical models can include operational routing decisions and strategic allocation of ambulances, such models are not based

on hierarchical decisions within EMS systems and a bi-level model. The upper-level of the proposed model, constitutes the planner's allocation problem. The problem seeks to allocate ambulances and how to set triage priorities, so as to minimize the service completion time of red code patients, green code patients, and black code patients, considering routing/pickup plans optimized in the lower-level problem. The centralization of the novel model is on a hierarchical decision-making structure and EMS systems. Verily, this study is required to develop the existing models and to propose a new mixed-integer bi-level model for EMS system.

The second noticeable need is to propose an adaptable and suitable solution approach, such as meta-heuristic algorithms for solving this NP-hard bi-level EMS problem. From our viewpoint, this work designs a novel scheme of computational intelligence and generates acceptable computational results in terms of accuracy and CPU time. Most of the existing researches on solving ambulance routing problems developed exact approaches. The exact approaches may not converge at an admissible CPU time for many large-scale instances of NP-hard EMS problems. This prompted us to present a fast solution approach to reduce CPU times.

The principal differences and novelties of this study compared to existing research can be summarized as follows:

Bi-level decision-making structure: Unlike most existing EMS vehicle routing with single-level or multi-objective decision-making structures, this work formulates the ambulance routing problem as a bi-level optimization model. The upper level describes the strategic planner's ambulance allocation decisions, while the lower level explicitly models triage-based ambulance routing decisions, leading to a more practical leader-follower depiction of EMS systems.

Hybrid solution methodology: Methodologically, this study is distinct from available methods by presenting a hybrid bi-level GA-TS-TLBO solution approach, where GA operators, TS local intensification, and TLBO-style learning explore the upper-level solution region and an exact solver warrants feasibility of the lower-level subproblem.

Empirical validation and statistical robustness: Beyond computational experiments and realistic disaster-inspired instances, this study applies Wilcoxon test to statistically validate the robustness of the novel hybrid GA-TS-TLBO approach, which is rarely revealed in similar EMS studies.

Table 1 provides a structured comparison between the most relevant prior studies and the proposed model. As can be seen, existing works address important components of EMS routing or disaster logistics, yet none simultaneously integrate triage-based patient prioritization, hierarchical or bi-level decision-making, and routing on disrupted transportation networks. The proposed model fills this gap by combining strategic ambulance allocation and operational routing in a Stackelberg bi-level configuration while explicitly modeling heterogeneous triage categories. Although prior studies have addressed important aspects of EMS operations such as patient-condition-based routing (Rabbani et al. [26], heterogeneous ambulances (Tikani and Setak, [38], patient-group evacuation (Talarico et al. [37], or bi-level evacuation and logistics (Liu et al. [21]; Wang et al. [39]), none integrate triage-based ambulance allocation, operational routing, and disaster-induced network disruption within a hierarchical decision-making framework. Existing works typically treat routing and allocation separately, ignore the leader-follower structure intrinsic to EMS command protocols, or simplify triage categories without assigning operational priority weights. The proposed study addresses these gaps by introducing the first bi-level mixed-integer

Table 1: Comparison of related studies on EMS routing, triage optimization, and bi-level disaster response

Reference	Problem Domain	Optimization Structure	Triage	Disaster	Allocation	Routing	Solution Strategy
Najafi et al. [23]	Earthquake medical logistics	Multi-objective	Yes	Yes	No	No	Exact
Talarico et al. [37]	Disaster ambulance routing	Single-level	No	Yes	No	Yes	Metaheuristic
Borhanifar and Shadkam [5]	Generic optimization problem	Multi-objective	No	No	No	No	Hybrid metaheuristic
Tikani and Setak [38]	Ambulance routing	Multi-objective	Yes	Yes	No	Yes	Metaheuristic
Rabbani et al. [26]	Ambulance routing	Multi-objective	Yes	Yes	No	Yes	Metaheuristic
Shadkam and Cheraghchi [34]	Earthquake relief prioritization	Hybrid two-phase	Yes	Yes	No	Yes	Hybrid heuristic
Wang et al. [39]	Post-disaster logistics	Bi-level multi-objective	No	Yes	Yes	No	Exact
Rajabi et al. [27]	COVID-19 supply chain network	Multi-objective	No	Yes	No	No	Goal programming
Liu et al. [21]	Emergency evacuation traffic	Bi-level	No	Yes	No	No	Heuristic
Afsharirad [1]	Vehicle routing problem	Single-level	No	No	No	Yes	Exact
This Paper	**Post-disaster ambulance allocation and routing**	**Bi-level**	**Yes**	**Yes**	**Yes**	**Yes**	**Hybrid meta-heuristic and exact**

programming model that simultaneously captures (i) triage categories (red/green/black), (ii) strategic ambulance allocation, (iii) operational routing under disrupted networks, and (iv) Stackelberg-style hierarchical EMS decision flows. The structured comparison in Table 1 highlights these distinctions and clarifies the novelties of the proposed framework. Although several papers evaluate metaheuristics or hybrid metaheuristics (e.g., Borhanifar and Shadkam, [5]; Talarico et al. [37]; Tikani and Setak [38];

Shadkam et al. [35]; Rabbani et al. [26]), none combine GA, TS, TLBO for the upper level, and an exact MILP solver for the lower level. This hybrid architecture improves the convergence and solution quality of this bi-level EMS routing problem.

2.6 Motivation

In real-world EMS disaster response, decision making is inherently hierarchical: an EMS command center (strategic planner) allocates ambulances to affected sites and prioritizes patient groups, while operational routing teams on the ground determine the most efficient routes under the constraints set by these strategic allocations. This reflects a cooperative leader-follower structure, where the leader's strategic decisions condition the feasible set of the follower's routing decisions. Modeling this dependency explicitly motivates the adoption of a bi-level optimization framework, which can capture the sequential nature of command decisions and ground-level routing responses, consistent with Stackelberg hierarchical decision-making in optimization. The key reasons are summarized below:

1. **Fundamental decision hierarchy in real EMS systems.**
   EMS operations inherently follow a leader-follower structure:

   - Upper level (Leader): Central command allocates ambulances to disaster sites and determines triage-based priorities.
   - Lower level (Follower): Operational teams determine feasible routes after receiving upper-level assignments, subject to real road conditions, capacity limits, and disruptions.

   A single-level model collapses this natural sequence and assumes simultaneous decision-making, which is inconsistent with real EMS practice. Our bi-level structure mathematically represents this real-world chain of command.

2. **Technical dependency between decision layers.**
   The feasibility region of the lower-level routing is endogenously determined by upper-level decisions. In the model, the routing variables and pickup variables are restricted by the upper-level assignment variable through linking constraints. A single-level weighted-objective formulation cannot preserve this asymmetric control structure, because it would require the lower-level routing to select assignments that contradict the hierarchical reality (e.g., an ambulance routing itself to a site without central assignment).

3. **Why weighted multi-objective single-level models fail.**
   We tested a single-level reformulation. The comparison shows that the single-level formulation cannot guarantee triage priority enforcement because the routing objective often overrides the strategic priority structure. It also becomes computationally more expensive, because the routing space becomes unnecessarily large without the hierarchical restriction imposed by the upper level. Thus, the empirical evidence confirms that a single-level model is both less realistic and less efficient.

4. **Bi-level modeling captures EMS-specific objectives more faithfully.**
   In EMS settings, the upper level explicitly minimizes priority-weighted unmet demand (critical for Red patients). The lower level independently minimizes travel time given the upper-level

guidance. These two objectives are not substitutable by simple weights because they occur at different operational layers, have different decision makers, and represent non-comparable goals (strategic vs. tactical). A single-level weighted model forces them into an artificial trade-off, which leads to triage violation (as shown empirically).

5. **Theoretical justification from literature.**
Consistent with prior works in disaster logistics and evacuation modeling (Lv et al. [22]; Liu et al. [21]; Camacho-Vallejo et al. [7]), bi-level programming is appropriate whenever:

- The leader controls high-level assignment decisions.
- The follower executes operational routing conditioned on leader constraints.
- The interaction is sequential, not simultaneous.

Our problem satisfies all three conditions.

We propose a novel bi-level approach, which not only is conceptually aligned with real EMS decision-making but also mathematically necessary to preserve the structure of feasible routing and strategically enforced triage priorities. Both theoretical reasoning and comparative experiments demonstrate the superiority and necessity of the bi-level formulation.

3 Mathematical model description

This section presents a detailed mathematical formulation of the proposed bi-level programming model for the ambulance routing problem in a post-disaster environment. The system consists of a set of disaster sites, each containing green, red, or black patients, and a fleet of ambulances operating on a disrupted transportation network. The adoption of a bi-level programming structure in this study is not only conceptually aligned with EMS disaster operations but also mathematically required to represent the hierarchical decision-making process accurately. In real-world settings, ambulance allocation and triage prioritization are determined centrally by an EMS command unit (upper level), while route planning and patient pickup decisions are executed by operational teams on the ground (lower level). These two decision layers interact sequentially rather than simultaneously: routing decisions are feasible only within the service region authorized at the upper level. Accordingly, the feasibility of lower-level variables is restricted by the assignment variable through linking constraints. Since these constraints establish an asymmetric leader-follower dependency, collapsing the problem into a single-level weighted or multi-objective formulation would destroy the hierarchical structure and generate routing behavior inconsistent with EMS command protocols.
Figure 1 shows comparison of single-level with bi-level optimization in EMS routing. The single-level model jointly determines ambulance allocation, triage prioritization, and routing, using a combined weighted or multi-objective function, which overrides triage and ignores hierarchical command. In contrast, the bi-level model separates upper-level EMS command decisions-allocation, triage setting, and service region-from lower-level routing and pickup decisions, enforcing EMS hierarchy and realistic constraints. These considerations jointly justify the necessity of the bi-level optimization framework in modeling post-disaster EMS routing and allocation.

Figure 2 illustrates the hierarchical structure of the proposed bi-level ambulance routing model. The upper level represents strategic decision-making and is responsible for triage prioritization, allocation

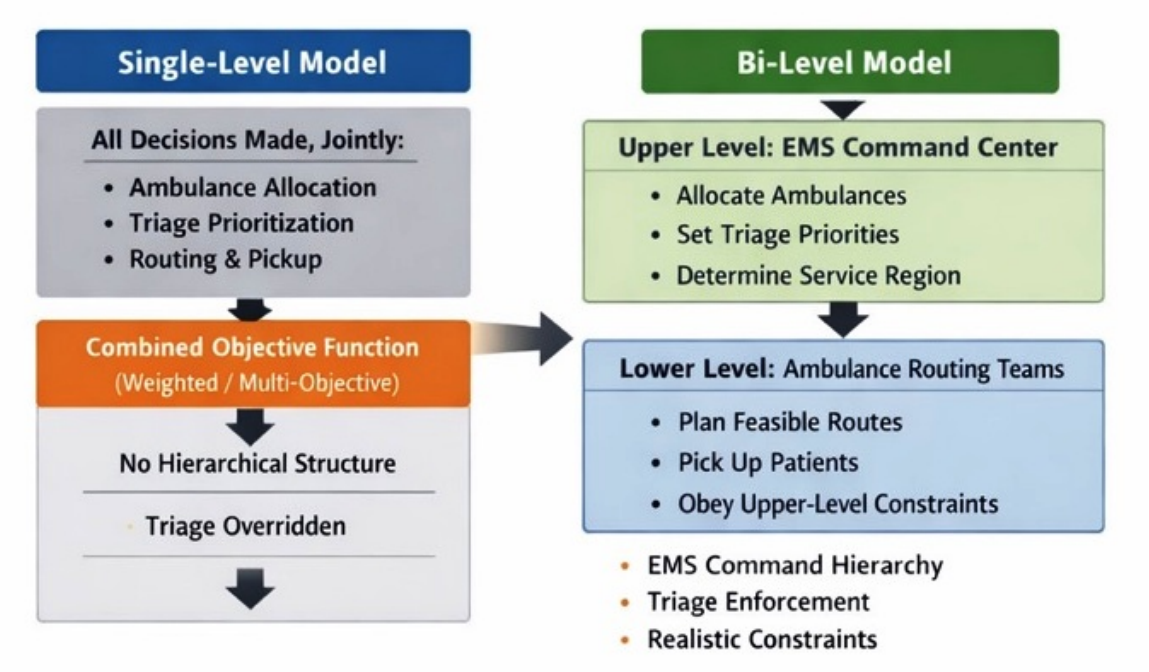

Figure 1: Conceptual difference between single-level and bi-level optimization in EMS routing

of disaster sites, and assignment of patients. The objective at this level is to maximize a weighted service performance measure that reflects the relative urgency of different patient categories. Given the upper-level decisions, the lower level addresses the operational problem of ambulance dispatching and routing. This level determines feasible routes that minimize total travel time while satisfying ambulance availability, capacity, and network constraints. The lower-level problem is formulated as a mixed-integer programming model and is solved exactly for each candidate upper-level solution. The interaction between the two levels is explicitly depicted through directional information flows. Strategic assignments generated at the upper level define the feasible space of routing decisions at the lower level. In contrast, the resulting routing solutions provide performance feedback used to evaluate and update upper-level decisions. This hierarchical structure ensures that operational routing plans are fully aligned with triage-based strategic objectives, which is essential in post-disaster emergency medical service operations.

3.1 Notations:

Sets and indices

I : Set of disaster sites, indexed by i,

A : Set of ambulances, indexed by k,

N : Set of all hospitals in the transportation network, indexed by j,

G : Set of green code patients,

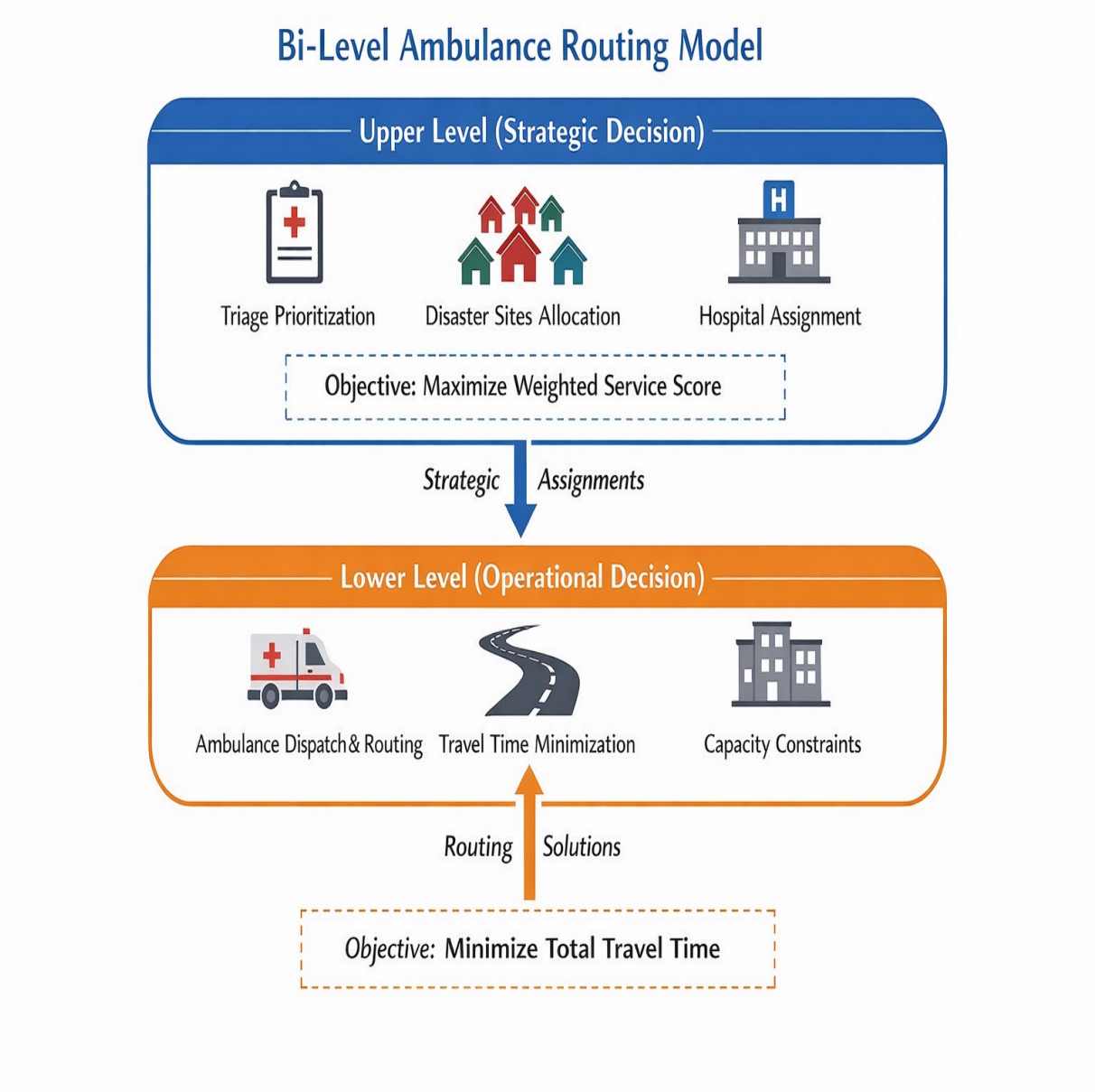

Figure 2: Schematic representation of the proposed bi-level ambulance routing model

R : Set of red code patients,

B : Set of black code patients,

T : Patient triage categories ($T = \{G, R, B\}$), indexed by t.

Parameters

d_{ij} : Travel time between nodes i and j.

p_i^t : Number of patients of type t at site i.

C_k : Capacity of ambulance k.

w_t : Priority weight of triage class t.

Decision variables

x_{ijk} : 1, if ambulance k travels from i to j; 0, otherwise.

y_{ik}^t : Number of patients of type t picked by ambulance k from site i.

z_{ik} : 1, if ambulance k is assigned to site i; 0, otherwise.

3.2 Bi-Level structure

The following bi-level programming model is proposed:

$$\min \quad Z_U(z_{ik}) = \sum_{k\in A}\sum_{i\in I}\sum_{t\in T} w_t p_i^t z_{ik}, \tag{1}$$

$s.t.$

$$\sum_{i\in I} z_{ik} \leq 1, \forall k \in A, \tag{2}$$

$$z_{ik} \in \{0,1\}, \forall i \in I, k \in A, \tag{3}$$

where z solves:

$$\min \quad Z_L(x_{ijk}, y_{ik}^t) = \sum_{k\in A}\sum_{i\in I}\sum_{j\in N} d_{ij} x_{ijk}, \tag{4}$$

$s.t.$

$$\sum_{j\in N} x_{ijk} - \sum_{j\in N} x_{jik} = 0, \forall k \in A, i \in I, \tag{5}$$

$$\sum_{i\in I}\sum_{t\in T} y_{ik}^t \leq C_k, \forall k \in A, \tag{6}$$

$$y_{ik}^t \leq p_i^t z_{ik}, \forall i \in I, t \in T, k \in A, \tag{7}$$

$$x_{ijk} \leq z_{ik}, \forall i \in I, j \in N, k \in A, \tag{8}$$

$$y_{ik}^t \geq 0 \ \text{ integer}, \forall i \in I, k \in A, t \in T, \tag{9}$$

$$x_{ijk} \in \{0,1\}, \forall i \in I, j \in J, k \in A, \tag{10}$$

In this bi-level optimization problem, (1) to (3) form the leader's subproblem, while (4) to (10) are equivalent to the mathematical formulation of the follower's subproblem. $Z_U(z_{ik})$ and $Z_L(x_{ijk}, y_{ik}^t)$ in (1) and (4) show the high-level decision maker's and low-level decision maker's objective functions, respectively. The upper level determines ambulance assignments , and given upper-level assignments, the lower-level

subproblem optimizes routing. The objective function (1) minimizes the service completion time of red code patients, green code patients, and black code patients, while the objective function (4) minimizes travel time between disaster sites. In the proposed objective function, triage classes are differentiated using explicit priority weights satisfying $w_R \gg w_G \gg w_B$. In our implementation, the weight assigned to Black-tag patients is set to a very small value, ensuring that their evacuation never competes with or displaces the rescue of Red-tag patients, who receive the highest priority weight. Black patients represent the 'expectant' subgroup in disaster triage-individuals with very low survival probability but who may be transported only after all viable patients have been addressed and if residual ambulance capacity remains. The model does not force evacuation of Black-tag individuals; rather, they are considered only when strategic resources allow, and their inclusion does not alter routing or service completion times for Red patients. This treatment preserves medical realism while allowing the model to capture eventual evacuation needs without compromising life-saving priorities. Constraints (2) assure that each patient will be visited directly with an ambulance. Constraints (5) ensure that the number of arrivals to disaster site i by ambulance k is equal to the number of departures from disaster site i. Constraints (6) state that the number of patients who will be carried by each ambulance should not exceed the capacity of that ambulance. Constraints (7) prohibit pick-up service of patients to an ambulance that is not assigned to a disaster site. Constraints (8) prevent the follower from traveling ambulances not assigned to a disaster site at all. Finally, constraints (3), (9), and (10) are domain restrictions for decision makers' variables.
The only decision variables passed from the upper level to the lower level are the binary ambulance-site assignment variables z_{ik}. These variables determine whether ambulance k is permitted to serve disaster site i. In the lower-level problem, the routing variables x_{ijk} and pickup variables y^t_{ik} are constrained by z_{ik}. Specifically, constraints $y^t_{ik} \leq p^t_i z_{ik}$ and $x_{ijk} \leq z_{ik}$ enforce that an ambulance may visit or collect patients from a site only if the upper level has authorized that service. Thus, the upper level defines the feasible service region for each ambulance, while the lower level computes the optimal routing and pickup plan within those constraints.
The interaction between upper-level assignments (z_{ik}) and lower-level routing/pickup decisions (x_{ijk}, y^t_{ik}) is expressed using linear linking constraints such as $y^t_{ik} \leq p^t_i z_{ik}, \quad x_{ijk} \leq z_{ik}$, which avoid bilinear terms. The capacity and flow-balance constraints are linear by construction. No product of decision variables appears in the final lower-level formulation; therefore, CPLEX can solve the follower problem directly as a mixed-integer linear program (MILP).

4 Hybrid solution approach

Solving the proposed bi-level programming model for ambulance routing in post-disaster conditions requires advanced computational strategies that can handle the multi-layered hierarchical structure and combinatorial complexity. Traditional exact algorithms become computationally intractable for medium and large-scale instances due to NP-hardness. Therefore, this study employs a hybrid solution framework integrating GA, TS, and TLBO and provides mathematical sub-formulations, algorithmic pseudocode, complexity remarks, and illustrative flowcharts. Each method addresses different optimization needs of the bi-level model, combining global exploration, local intensification, and adaptive learning.

The hybrid structure is designed such that GA performs the global search and generates diverse candidate solutions. The TS refines promising solutions locally to escape suboptimal neighborhoods. The TLBO then enhances solution quality through learning mechanisms in a parameter-free environment,

enabling additional exploitation and diversity management. Collectively, this tri-layer framework ensures computational efficiency and high-quality solutions.

4.1 The GA

The GA is responsible for global exploration and generating highly diverse initial solutions. Its population-based nature makes it well-suited to complex search spaces with multiple local minima. The GA is employed to efficiently explore the complex solution space of the ambulance routing problem, where disrupted transportation networks, heterogeneous patient categories, and strict time-critical evacuation requirements make exact optimization computationally challenging. Inspired by natural evolution, GAs represent candidate routing plans as chromosomes and iteratively improve them through selection, crossover, and mutation operators (Holland [17]; Goldberg [15]). Each chromosome encodes the sequence of sites visited and the assignment of green, red, and black patients to ambulances. At the same time, the fitness function evaluates travel times, priority-weighted patient evacuations, and resource constraints. Similar to other applications of GAs in vehicle routing (Prins [25]; Ombuki-Berman and Hanshar [24]), the algorithm balances intensification and diversification to identify high-quality, near-optimal plans within practical computational time. This makes GAs particularly suitable for post-disaster emergency response, where rapid and robust decision-making is essential.

A triage-based seeding strategy is employed, whereby red patients are assigned first to minimize delays. After, green patients are distributed based on nearest-neighbor criteria. Black patients are assigned last if evacuation is required. For the leader's subproblem, the selection operator employs tournament selection to balance exploration and convergence. Crossover uses route-based operators, i.e., order crossover and partially mapped crossover. Mutation includes node swap mutation, route segment reversal, and ambulance reassignment mutation. Then, the CPLEX solver is utilized to ensure feasibility by obtaining exact solutions for the follower's subproblem. Below is the pseudocode describing the GA-CPLEX algorithm, signified as Algorithm 1.

Algorithm 1: GA-CPLEX to solve the bi-level VRP

Input: Leader's subproblem data, GA parameters
Output: Best solution found
1: Generate an initial population P of feasible solutions
2: Call *cplexmilp* solver to obtain an exact solution for the follower's subproblem
3: Evaluate the fitness of each chromosome in P
4: **while** the stopping criterion is not met **do**
5: Select parent chromosomes using tournament selection
6: Apply the crossover operator to produce offspring
7: Apply the mutation operator to maintain diversity
8: Call *cplexmilp* solver to solve the follower's subproblem
9: Evaluate offspring fitness
10: Create a new population using the elitism strategy
11: **end while**
12: Return the best chromosome found

4.2 The TS

The TS is a powerful metaheuristic applied to the ambulance routing problem to efficiently navigate its highly combinatorial search space, particularly under post-disaster conditions where road disruptions, heterogeneous patient priorities, and tight evacuation times create complex routing constraints. TS iteratively explores neighboring routing solutions while using short-term and long-term memory structures—known as tabu lists—to prevent cycling and encourage diversification (Glover [14]; Laguna [19]). In this context, each move may involve reordering site visitation sequences, reallocating patients among ambulances, or adjusting travel paths to avoid blocked roads. The tabu mechanism helps escape local optima and guides the search toward promising regions of the solution space, consistent with the success of TS applications in vehicle routing and emergency logistics (Taillard et al. [36]; Cordeau et al. [10]). Due to its balance of intensification and diversification, tabu search provides high-quality routing strategies within reasonable computational time, making it suitable for time-sensitive ambulance dispatch decisions in disaster environments.

Here, TS performs local improvement on elite GA solutions. It is designed to intensify the search within promising regions identified by GA. TS applies the following neighborhood operations:

- **Swap**: exchange two patient visits,
- **Relocate**: move a patient from one ambulance route to another,
- **2-opt**: reverse route subsections to reduce travel time.

A tabu list stores recently applied moves to avoid cycling. Aspiration criteria allow tabu moves if they yield a globally superior solution. Below is the pseudocode describing the TS-CPLEX algorithm, signified as Algorithm 2.

Algorithm 2: TS-CPLEX to solve the bi-level VRP

Input: Leader's subproblem data, TS parameters, Initial solution S
Output: Best solution found
1: best = S
2: tabu list = empty
3: **while** the stopping condition is not met **do**
4: Generate neighborhood $N(S)$
5: For each neighbor S' in $N(S)$:
6: Call *cplexmilp* solver to solve the follower's subproblem
7: Evaluate fitness $Z_U(S')$
8: Select the best non-tabu S' unless aspiration applies
9: Apply selected move and update tabu list
10: **if** $Z_U(S') < Z_U$(best) **then**
11: best = S'
12: **end if**
13: $S = S'$
14: **end while**
15: return best

4.3 The TLBO

The TLBO is a population-based metaheuristic inspired by the pedagogical interaction between teachers and learners, and it has been successfully applied to complex routing and emergency-response problems due to its simplicity, parameter-free structure, and strong convergence characteristics. In the ambulance routing problem, each learner represents a candidate routing plan encoding visit sequences and patient assignment, while the teacher phase drives the population toward better performance by improving global solution quality, and the learner phase enhances diversity through peer-to-peer interactions (Rao et al. [28]). TLBO has shown competitive performance in various transport and routing applications, outperforming several classical metaheuristics in solution quality and robustness (Rao and Patel [29]; Liu and Liu [20]; Bhatia et al. [4]). These strengths make TLBO suitable for time-critical ambulance routing, where rapid generation of high-quality solutions is essential for saving lives in post-disaster conditions. In TLBO, the best solution acts as the "teacher," pulling weaker solutions toward better performance:

$$L_{\text{new}} = L_{\text{old}} + r(L_{\text{teacher}} - T_F \cdot L_{\text{mean}}),$$

where $r \in [0,1]$ is a vector of independent uniform random number and T_F is the teaching factor randomly as 1 or 2. Learners update their knowledge through random peer interactions. Below is the pseudocode describing the TLBO-CPLEX algorithm, signified as Algorithm 3.

Algorithm 3: TLBO-CPLEX to solve the bi-level VRP

Input: Leader's subproblem data, TLBO parameters, Population of solutions
Output: Best solution found
1: Identify Teacher (best solution)
2: **while** not converged **do**
3: // Teacher Phase
4: For each learner *Li*:
5: Update *Li* towards Teacher
6: Call *cplexmilp* solver to solve the follower's subproblem
7: Evaluate fitness *Li*
8: // Learner Phase
9: For each pair (Li, Lj):
10: **If** $Z_U(Li) < Z_U(Lj)$:
11: $Lj = Lj + r(Li - Lj)$
12: **Else**:
13: $Li = Li + r(Lj - Li)$
14: Call *cplexmilp* solver to solve the follower's subproblem
15: Evaluate fitness of updated solutions
16: **end while**
17: return best solution

The TLBO was selected due to several characteristics that make it particularly well-suited for bi-level ambulance routing:

1. Parameter-free nature: Unlike GA (requiring crossover/mutation rates) or PSO (requiring inertia and acceleration coefficients), TLBO has no algorithm-specific parameters. This substantially simplifies calibration when the lower-level must be solved repeatedly with CPLEX.

2. Fast convergence: Preliminary tests showed that TLBO converged faster than GA and PSO for our problem structure, reducing the number of expensive lower-level CPLEX calls.

3. Strong exploitation capability: Since TLBO relies on instructor–learner and learner–learner phases, it provides strong local refinement—useful in Stackelberg-type bi-level problems where each evaluation is computationally costly.

4. Demonstrated performance in routing and combinatorial problems: Recent studies (e.g., Rao et al. [28]; Liu and Liu [20]; Bhatia et al. [4]) report competitive or superior performance of TLBO compared to classical GA/PSO variants in VRP settings.

4.4 Hybrid GA-TS-TLBO framework

This subsection describes a practical hybrid metaheuristic approach that solves a bi-level ambulance routing problem where the upper-level subproblem are handled by a hybrid population-based/metaheuristic method and the lower-level subproblem is solved exactly with CPLEX solver. The upper-level metaheuristic approach mixes GA operators, TS local intensification, and TLBO-style learning to balance exploration and exploitation for ambulance allocation and dispatching. The lower-level is treated as a parameterized operational routing subproblem. For each candidate upper-level decision variables, CPLEX solves the induced lower-level problem, and the resulting follower decision variables are used to evaluate the leader's fitness. This hybrid approach is appropriate because the lower-level subproblem is modeled as a mixed-integer program solvable by a commercial solver and the upper-level space is large and combinatorial so that an exact bi-level algorithm is impractical.
The hybrid GA-TS-TLBO design is not intended as a redundant combination of metaheuristics but as a structured integration of methods with complementary roles. The GA provides global exploration of the highly combinatorial upper-level space, TS intensifies the search around promising assignments to escape local minima, and TLBO accelerates convergence through population-level learning without additional parameters. Preliminary experiments conducted during algorithm design indicated that GA alone converges slowly, TS alone is highly dependent on initial conditions, and TLBO alone lacks the exploratory capability required for a bi-level assignment-routing problem. By combining these components, the hybrid framework achieves both exploration and exploitation, yielding faster convergence and more stable solutions.

4.4.1 Hybrid algorithm design

Population initialization (GA-style): Create a population of P candidate upper-level solutions z.

Upper-level evaluation: For each z, call CPLEX to solve the lower-level problem $\min Z_L(x,y)$. Use the optimal lower-level solution x,y to compute the leader fitness $\min Z_U(z)$.

Selection and reproduction (GA): Apply selection, crossover, and mutation to produce offspring.

Learning phase (TLBO-inspired): Apply teacher and learner phases to the population to accelerate convergence using population statistics rather than extra parameters.

Intensification (TS): Apply a TS on promising individuals (elite subset) to refine them locally and escape local optima.

Replacement: Integrate offspring and improved individuals to form the next generation.

Stopping: Stop on max generations or no-improvement threshold.

At every evaluation that proposes a new z, the algorithm invokes CPLEX to produce an exact lower-level response and enforce bilevel feasibility and optimality. The GA provides robust global search and natural diversity the use of crossover and mutation. TLBO brings parameter-light guided learning: the teacher phase moves the population toward the best-known solution, and the learner phase allows pairwise exchange/learning—helpful for accelerating convergence without many hyperparameters. TS provides focused, memory-based local search to intensify around reasonable solutions and avoid cycling. CPLEX ensures the lower-level is solved optimally, preserving the correctness of the bilevel evaluation. Combining them yields a balance: GA and TLBO drive exploration and guided learning; TS intensifies and corrects; CPLEX enforces exact follower reaction.

4.4.2 Genetic operators

The upper-level decision variable z is encoded as a chromosome/vector. We use the problem-appropriate encoding include the binary string for yes/no ambulance assignment decisions. Generate P random and apply:

Selection: Roulette-wheel based on leader fitness.

Crossover: Problem-specific crossover (i.e., swap segments in assignment problems).

Mutation: Bit-flip mutation.

4.4.3 TLBO-style learning phases

Teacher phase:

Identify teacher z_t (best leader fitness). For each individual z_i, update:

$$z_i' = z_i + r \cdot (z_t - T_f \cdot \bar{z}), \tag{11}$$

where $\bar{z}$ is the population mean, r is random in [0,1], and T_f is either 1 or 2.

Learner phase:

- Pair individuals randomly.
- For a pair (i, j), if z_i better than z_j, update $z_j \leftarrow z_j + r \cdot (z_i - z_j)$ else the reverse.
- Evaluate modified individuals.

The TLBO phases require fewer parameters than other local search heuristics and complement GA diversity.

4.4.4 The TS intensification

- Select an elite subset of each generation.
- Run a TS (fixed iterations, small tabu tenure) exploring local moves that modify a few components of z for each elite individual.
- Maintain a tabu list to avoid reversal of recent moves.
- Evaluate each neighbor by solving the lower-level with CPLEX.
- Accept improvements and update the elite individual.
- Return improved elites into population replacement pool.

4.4.5 Pseudocode of hybrid GA-TS-TLBO

Below is the pseudocode describing the hybrid GA-TS-TLBO algorithm, signified as Algorithm 4.
To control the computational cost, we include the following mechanisms in the solution approach:

Feasibility screening: Before invoking CPLEX, each candidate upper-level solution is checked for basic feasibility (capacity bounds, assignment structure). Infeasible candidates are discarded without lower-level evaluation.

Solution caching / memoization: We store previously evaluated upper-level solutions and their corresponding lower-level optimal values. If TLBO generates a duplicate or nearly identical assignment vector, the stored value is reused without calling CPLEX.

Limited TLBO population size: The TLBO population and iteration count were calibrated experimentally to balance solution quality and computational efficiency.

Warm-starting CPLEX: The lower-level solver is warm-started using the previous iteration's feasible solution, which significantly reduces solution times, especially for medium-sized instances.

5 Experimental results

This section presents a focused computational evaluation of the proposed bi-level programming model and hybrid GA-TS-TLBO solution framework. The experiments are designed to assess the practical performance of the approach on realistic disaster-response scenarios, with particular emphasis on solution quality, computational efficiency, and compliance with triage-based priorities. To ensure a fair and meaningful comparison, all experiments are conducted on benchmark instances commonly used in the ambulance routing literature. Special attention is given to medium-sized instances, which are representative of operationally relevant post-disaster situations and pose significant computational challenges for exact solution methods.

Algorithm 4: hybrid GA-TS-TLBO to solve the bi-level VRP

Input: population size P, $max_{generations}$ G, elite fraction E, TS_{iters}, TLBO and GA operators
1: Initialize population Pop of size P
2: Compute $Z_{U(z)}$ for each z in Pop by solving lower-level $Z_{L(x,y)}$ with CPLEX
3: **for** $gen = 1$ **to** G:
4: GA selection + reproduction
5: Parents = select(Pop)
6: Offspring = crossover and mutate(Parents)
7: Evaluate Offspring via CPLEX
8: TLBO phases
9: **if** $TLBO_{enabled}$:
10: teacher = argmin Fitness(Pop U Offspring)
11: evaluate new candidates via CPLEX
12: PopOff = Pop U Offspring
13: PopOff = teacher phase(PopOff, teacher)
14: PopOff = learner phase(PopOff)
15: **else**:
16: PopOff = Pop U Offspring
17: TS intensification on elites
18: Elites = select top(PopOff, fraction=E)
19: for each e in Elites:
20: $e_{improved} = tabu_{search}(e, TS_{iters})$
21: evaluate neighbors via CPLEX
22: replace e with $e_{improved}$ in PopOff
23: Replacement: keep the best P individuals
24: $Pop = select_{best}(PopOff, P)$
25: **if** stopping criteria met: **break**
27: Return the best z and associated x, y from evaluations

5.1 Test instances and data description

The computational study is based on realistic disaster-inspired instances adapted from the benchmark data of Talarico et al. [37]. The cases differ in the number of disaster sites, ambulances, and patients, capacity constraints, and are grouped into small, medium, and large categories. Patients are classified into red and green triage categories, and priority weights are assigned accordingly. The upper-level problem is solved using a hybrid algorithm combined with the GA-TS-TLBO algorithm, while the lower-level routing problem is solved using the CPLEX solver. Each instance is solved 30 times, and average results are reported to reduce stochastic effects.

5.2 Overall performance of the proposed approach

To ensure reproducibility, all algorithmic parameters used in the GA-TS-TLBO framework are explicitly reported in Table 2. These include GA parameters (population size, maximum iteration, probability for crossover, and probability for mutation), TS settings (tabu list size and number of reiterations

without any improvement in the best solution), and TLBO parameters (population size and maximum iteration). Parameter values were selected through preliminary sensitivity testing on medium-sized instances to balance exploration capacity, computational cost, and convergence speed, especially given that each upper-level evaluation requires solving a lower-level subproblem with CPLEX. The stopping criteria-maximum number of generations and a no-improvement threshold were also calibrated based on observed convergence behavior.

Table 2: The input parameters of GA, TS and TLBO algorithms

Algorithm	Parameter	Value
GA	Population size	50
	Maximum iteration	100
	Probability for crossover	0.75
	Probability for mutation	0.25
TS	Tabu list size	$\|I\|$
	Number of reiterations without any improvement in the best solution	15
TLBO	Population size	50
	Maximum iteration	100

Table 3 summarizes the overall computational performance of the proposed bi-level approach for different instance sizes. The table reports the average upper-level objective value, the corresponding lower-level travel time, and the average CPU time.

Table 3: Overall computational performance of the proposed bi-level approach

Instance size	objective value	travel time	CPU time (s)
Small (10 sites)	892.4	318.7	42.6
Medium (25 sites)	1041.8	746.2	91.3
Large (50 sites)	1127.6	1324.9	176.8

The results indicate that the proposed solution method is able to consistently generate feasible solutions for all tested instances. As expected, both the objective value and computational time increase with instance size; however, the observed CPU times remain within ranges suitable for decision support in disaster response settings.

5.3 Impact of ambulance fleet size

To analyze the effect of ambulance availability on system performance, a sensitivity analysis with respect to the number of ambulances was conducted. Instances were categorized into low, medium, and high fleet-size scenarios.

Table 4: Effect of ambulance fleet size on service completion times

Fleet size	e_R (red patients)	e_G (green patients)	CPU time (s)
Low	128.6	94.3	64.1
Medium	97.4	72.8	88.6
High	95.1	71.9	121.4

As shown in Table 4, increasing the number of ambulances substantially reduces the latest service completion times for both red and green patients when moving from a low to a medium fleet size. However, the marginal benefit of additional ambulances diminishes for high fleet sizes, indicating that excessive resources do not necessarily translate into proportional service improvements.

5.4 Problem structure and solution quality

This subsection examines how key structural parameters influence solution quality and validates the ability of the proposed bi-level model to reflect realistic triage policies. Table 5 reports the impact of varying the priority weight assigned to red patients. The results clearly show that increasing the weight w_R leads to earlier service completion for red patients at the expense of longer completion times for green patients.

Table 5: Sensitivity analysis with respect to the triage priority weight w_R

w_R	e_R (red patients)	e_G (green patients)	objective value
1	121.3	79.6	984.7
2	109.8	88.1	1076.2
5	94.2	102.4	1289.5
10	82.7	118.9	1546.8

This behavior is consistent with real-world triage principles, where critical patients are prioritized even if this increases waiting times for less severe cases. The monotonic reduction in e_R confirms that the upper-level decisions effectively guide lower-level routing outcomes.

Finally, Table 6 compares the proposed bi-level approach with a classical single-level routing formulation. The comparison highlights the advantage of explicitly modeling hierarchical decision-making and triage priorities.

Table 6: Comparison between the proposed bi-level model and a single-level routing approach

Method	objective value	CPU time (s)	Red-patient priority satisfied
Single-level routing	1186.4	214.7	No
Proposed bi-level model	1041.8	91.3	Yes

Overall, the computational results demonstrate that the proposed bi-level model not only improves solution quality and computational efficiency but also ensures compliance with triage-based priorities, which are essential in post-disaster emergency medical operations.

5.5 Statistical validation of results

To ensure that the observed performance improvements of bi-level over the single-level approach are not due to random variability inherent to the metaheuristic, we conducted formal statistical significance tests. Specifically, each algorithm was executed independently for 30 runs. Performance metrics were then compared between the proposed bi-level approach and the single-level baseline. Since normality could not be guaranteed for all metrics, both a parametric two-sample t-test and a non-parametric Wilcoxon rank-sum test were applied. The results show that the proposed bi-level approach achieves statistically significant improvements across all primary performance measures at the 95% confidence level. These findings confirm that the reported gains are robust and not attributable to random chance. Computational time differences were statistically significant, indicating that performance gains of single-level were achieved with a significant increase in runtime. The results are summarized in Table 7.

Table 7: Statistical significance tests comparing bi-level and single-level approaches

Performance metric	Single-level	Bi-level	T-test(p-value)	Wilcoxon(p-value)	Significant
Objective Value	1186.4	1041.8	3.018	2.024	Yes
CPU time	214.7	91.3	4.214	3.198	Yes

5.6 Sensitivity analysis on ambulance capacity and triage composition

We evaluated three capacity scenarios-low, medium, and high of Talarico et al [37] by proportionally scaling C_k across all ambulances. For each scenario, 30 runs were performed. The analysis shows that increasing ambulance capacity significantly reduces the number of trips and the lower-level travel time. The reduction in unmet demand for red patients is most pronounced when increasing capacity from low to medium. Beyond this point, improvements become marginal, indicating diminishing returns. The results are summarized in Tables 8.

Table 8: Sensitivity analysis on ambulance capacity C_k

C_k	Average service completion time	% Red Patients Served
Low	1041.8	82.1%
Medium	1108.8	90.5%
High	1214.8	95.4%

We generated three disaster-severity scenarios by altering patient composition:

- Red-dominant: High proportion of critical patients,
- Balanced: Baseline triage distribution,
- Green-dominant: Few critical patients, many low-severity patients.

The findings indicate that in red-dominant scenarios, the upper-level prioritization becomes more influential, and the model shifts resources aggressively toward high-severity sites. The bi-level structure is particularly effective here, reducing red-patient unmet demand. In green-dominant scenarios, the model

exhibits more balanced routing behavior and minimizes travel time more efficiently.
To address it, we have conducted a sensitivity analysis examining (i) variations in ambulance capacity C_k and (ii) changes in the ratio of Red/Green/Black patients to evaluate model robustness under different stress levels. The results demonstrate that the model responds in a stable and interpretable manner: increasing ambulance capacity improves patient throughput and reduces delays, while shifts toward higher proportions of critical (Red) patients increase system congestion and response times, as expected. These analyses confirm that the model captures realistic system behavior under both capacity stress and demand-severity stress. As the proportion of Red patients increases, system performance degrades smoothly, indicating that the model realistically captures overload conditions and prioritization effects. The results are summarized in Table 9.

Table 9: Sensitivity analysis on patient severity mix

Red/Green/Black Ratio	Average service completion time	% Red Patients Served
20%/60%/20%	1041.8	96.8%
35%/50%/15%	1117.2	90.5%
50%/40%/10%	1196.5	81.3%
65%/30%/5%	1287.4	68.9%

Overall, the sensitivity analysis confirms that the proposed model behaves consistently and predictably under varying operational capacities and demand severities, supporting its applicability for emergency planning and stress-testing scenarios.

6 Conclusion

This paper investigated a triage-based ambulance routing problem arising in post-disaster environments and proposed a novel bi-level programming framework to capture the hierarchical nature of emergency medical decision-making explicitly. The upper level of the model represents strategic allocation and prioritization decisions among heterogeneous patient groups, while the lower level determines operational ambulance routes under capacity and infrastructure constraints. By integrating these two decision layers, the proposed formulation provides a unified and realistic representation of post-disaster emergency medical service operations.

To address the computational challenges associated with bi-level optimization, a hybrid solution strategy combining genetic algorithm, tabu search, teaching-learning-based optimization, and an exact solver for the lower-level problem was developed. Computational experiments conducted on realistic benchmark instances demonstrate that the proposed approach consistently produces high-quality feasible solutions within short computation times. In particular, the results show that the model effectively enforces triage-based priorities, leading to earlier service completion for critical patients without compromising overall operational feasibility.

Several directions for future research can be identified. Extensions of the model could incorporate dynamic information updates, stochastic travel times, or time-dependent patient deterioration. Furthermore, integrating real-time dispatching and relocation decisions within the bi-level framework represents a promising avenue for enhancing operational realism. Finally, applying the proposed approach to large-scale real-world case studies would further support its practical applicability and managerial relevance.

References

[1] M. Afsharirad, *A mixed integer linear programming model for vehicle routing problem for non-complete graphs: Behshahr (Iran) case study*, J. Math. Model. **13(4)** (2025) 787-801.

[2] T. Andersson, P. Varbrand, *Decision support tools for ambulance dispatch and relocation*, J. Oper. Res. Soc. **58(2)** (2007) 195–201.

[3] J. Bazyar, M. Farrokhi, A. Salari, H.R. Khankeh, *The principles of triage in emergencies and disasters: a systematic review*, Prehosp Disaster Med. **35(3)** (2020) 305–313.

[4] S. Bhatia, N. Sharma, H. Sharma, *A hybridized teaching–learning-based optimization algorithm to solve capacitated vehicle routing problem*, in Computer Vision and Robotics: Proceedings of CVR 2022 (pp. 527-539), Singapore, Springer Nature Singapore.

[5] Z. Borhanifar, E. Shadkam, *The new hybrid COAW method for solving multi-objective problems*, Int. J. Found. Comput. Sci. Tech. **5** (2015).

[6] L. Brotcorne, G. Laporte, F. Semet, *Ambulance location and relocation models*, Eur. J. Oper. Res. **147(3)** (2003) 451–463.

[7] J. F. Camacho-Vallejo, E. González-Rodríguez, F.J. Almaguer, R. G. González-Ramírez, *A bi-level optimization model for aid distribution after the occurrence of a disaster*, J. Clean. Prod. **105** (2015) 134–145.

[8] H.S. Chang, C.H. Liao, *Planning emergency shelter locations based on evacuation behavior*, Nat. Hazards Obs. **76(3)** (2015) 1551–1571.

[9] Y. X. Chen, P.R. Tadikamalla, J. Shang, Y. Song, *Supply allocation: bi-level programming and differential evolution algorithm for Natural Disaster Relief*, Clust. Comput. **23(1)** (2020) 203–217.

[10] J.F. Cordeau, M. Gendreau, G. Laporte, J.Y. Potvin, F. Semet, *A guide to vehicle routing heuristics*, J. Oper. Res. Soc. **53(5)** (2002) 512–522.

[11] R.Z. Farahani, E. Miandoabchi, W.Y. Szeto, H. Rashidi, *A review of urban transportation network design problems*, Eur. J. Oper. Res. **229(2)** (2013) 281–302.

[12] X. Gao, *A bi-level stochastic optimization model for multi-commodity rebalancing under uncertainty in disaster response*, Ann. Oper. Res. **319(1)** (2022) 115–148.

[13] M. Gharib, S.M.T. Fatemi Ghomi, F. Jolai, *A multi-objective stochastic programming model for post-disaster management*, Transportmetr A Transp Sci. **18(3)** (2022) 1103–1126.

[14] F. Glover, *Tabu search—part I*, ORSA Journal on computing **1(3)** (1989) 190–206.

[15] D.E. Golberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addion Wesley, 1989.

[16] J. Goldberg, *Operations research models for the deployment of emergency services vehicles* EMS Mgmnt. J. **1(1)** (2004) 20–39.

[17] J.H. Holand, *Adaptation in Natural and Artificial Systems*, Ann Arbor: The University of Michigan Press, **32**(1975).

[18] J. Holguín-Veras, M. Jaller, L.N. Van Wassenhove, N. Pérez, T. Wachtendorf, *On the unique features of post-disaster humanitarian logistics*. J. Oper. Manag. **30(7-8)** (2012) 494–506.

[19] M. Laguna, *Tabu search*. In Handbook of Heuristics (pp. 941-958). Cham: Springer Nature Switzerland, 2025.

[20] X.C. Liu, Q. Liu, *A discrete teaching-learning-based optimization algorithm for the capacitated vehicle routing problem*, Open J. Transp. Technol. **3** (2014) 16-21.

[21] Y. Liu, Z. Zhang, L. Mo, B. Yu, Z. Li, *A bi-level emergency evacuation traffic optimization model for urban evacuation problem*, Comput.-Aided Civ. Infrastruct. Eng. **40(3)** (2025) 369–391.

[22] N. Lv, X. Yan, K. Xu, C. Wu, *Bi-level programming based contra flow optimization for evacuation events*, Kybernetes **39(8)** (2010) 1227–1234.

[23] M. Najafi, K. Eshghi, W. Dullaert, *A multi-objective robust optimization model for logistics planning in the earthquake response phase*, Transp. Res. E: Logist. Transp. Rev. **49(1)** (2013) 217–249.

[24] B. Ombuki-Berman, F.T. Hanshar, *Using genetic algorithms for multi-depot vehicle routing*. In Bio-inspired algorithms for the vehicle routing problem (pp. 77-99). Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[25] C. Prins, *A simple and effective evolutionary algorithm for the vehicle routing problem*, Comput. Oper. Res. **31(12)** (2004) 1985-2002.

[26] M. Rabbani, N. Oladzad-Abbasabady, N. Akbarian-Saravi, *Ambulance routing in disaster response considering variable patient condition: NSGA-II and MOPSO algorithms*, J. Ind. Manag. Optim. **18(2)** (2022) 1035-1062.

[27] R. Rajabi, E. Shadkam, S.M. Khalili, *Design and optimization of a pharmaceutical supply chain network under COVID-19 pandemic disruption*, Sustainable Operations and Computers **5** (2024) 102–111.

[28] R. V. Rao, V.J. Savsani, D.P. Vakharia, *Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems*, Computer-Aided Design **43(3)** (2011), 303–315.

[29] R.V. Rao, V. Patel, *An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems*, Sci. Iran. **20(3)** (2013) 710-720.

[30] Y.S. Ro, S.D. Shin, K.J. Song, W.C. Cha, J.S. Cho, *Triage-based resource allocation and clinical treatment protocol on outcome and length of stay in the emergency department*, Emerg. Med. Australas **27(4)** (2015) 328–335.

[31] E. Saghehei, A. Memariani, A. Bozorgi-Amiri, *Implementing solution algorithms for a bi-level optimization to the emergency warehouse location-allocation problem*, Int. J. Supply Oper. Manag. **10(2)** (2023) 151–173.

[32] D. Sarma, U.K. Bera, A. Singh, M. Maiti, *A multi-objective post-disaster relief logistic model*. In 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC) (pp. 205-208). IEEE.

[33] V. Schmid, *Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming*, Eur. J. Oper. Res. **219(3)** (2012) 611–621.

[34] E. Shadkam, M. Cheraghchi, *Prioritization of earthquake relief using a hybrid two-phase approach*, J. Appl. Res. Ind. Eng. **9(4)** (2022) 493–506.

[35] E. Shadkam, S. Safari, S. S. Abdollahzadeh, *Finally, which meta-heuristic algorithm is the best one?*, Int. J. Deci. Sci. Risk Manag. **10(1-2)** (2021) 32–50.

[36] E. Taillard, P. Badeau, M. Gendreau, F. Guertin, J. Y. Potvin, *A tabu search heuristic for the vehicle routing problem with soft time windows*, Transp. Sci. **31(2)** (1997) 170–186.

[37] L. Talarico, F. Meisel, K. Sorensen, A*mbulance routing for disaster response with patient groups*, Comput. Oper. Res. **56**(2015) 120–133.

[38] H. Tikani, M. Setak, *Ambulance routing in disaster response scenario considering different types of ambulances and semi soft time windows*, J. Ind. Syst. Eng. **12(1)** (2019) 95–128.

[39] Q. Wang, Y. Liu, H. Pei, *Modelling a bi-level multi-objective post-disaster humanitarian relief logistics network design problem under uncertainty*, Eng. Optim. **56(8)** (2024) 1220–1254.

[40] T. Yan, F. Lu, S. Wang, L. Wang, H. Bi, *A hybrid metaheuristic algorithm for the multi-objective location-routing problem in the early post-disaster stage*, J. Ind. Manag. Optim. **19(6)** (2023) 4663-4691.

[41] J. Ye, W. Jiang, X. Yang, B. Hong, *Emergency materials response framework for petrochemical enterprises based on multi-objective optimization*, Energy **269** (2023) 126670.

[42] W. Yi, L. Ozdamar, *A dynamic logistics coordination model for evacuation and support in disaster response*, Eur. J. Oper. Res. **179(3)** (2007) 1177–1193.