



Strategic allocation of medical resources in disaster response: a multi-objective nonlinear dynamic model with risk and cost considerations

Qing Liu^b, Jixiang Dong^b, Na Cui^{a,b,*}, Zimin Liang^c, Jiabin Luo^d

^a School of Information Science and Engineering, University of Jinan, Jinan, China

^b School of Civil Engineering and Architecture, University of Jinan, Jinan, China

^c School of Computer Science, University of Birmingham, Birmingham, United Kingdom

^d Department of Business Analytics and Information Systems, Aston University, Aston Triangle, Birmingham, United Kingdom

ARTICLE INFO

Keywords:

Emergency medical resource allocation
Hospital overcrowding
Multi-objective nonlinear dynamic model
NSGA-II
Simulation

ABSTRACT

Optimizing the allocation of emergency medical resources is very important to ensure the efficiency of the rescue processes, especially after large-scale natural disasters. This study proposes a multi-objective nonlinear dynamic model, aiming to minimize total system rescue costs, considering hospital overcrowding, untreated patients, ambulance operating, and medical personnel related costs, and at the same time maximize the total expected number of survivors. A Non-dominated Sorting Genetic Algorithm II (NSGA-II) is employed as the solution method to solve the proposed model efficiently, which has also been validated through comparative experiments against the Pareto Local Search (PLS) algorithm. Through simulation-based evaluation leveraging real-world data from the 2013 Ya'an earthquake (Sichuan, China), this study further evaluates the effects of critical parameters (e.g., medical personnel availability) on model performance. Experimental results demonstrate that the proposed model and algorithm can provide optimized strategic allocation of medical resources which would further significantly improve rescue efficiency, alleviate hospital overload, and reduce overall operating costs. The identified medical personnel allocation thresholds effectively prevent both service degradation (under-allocation) and resource waste (over-allocation), providing actionable decision-making guidance.

1. Introduction

In recent years, the intensification of global climate change has significantly increased, in terms of both the frequency and severity of natural disasters. Events such as earthquakes, hurricanes, floods, and wildfires have occurred with growing regularity, inflicting severe damage on human society and leading to substantial casualties and economic losses (Zhang et al., 2012, Haddow et al., 2025). For instance, the 2011 Great East Japan Earthquake triggered the Fukushima nuclear disaster, Hurricane Harvey devastated Houston in 2017, the 2020 Australian wildfires caused widespread destruction, and the 2013 Ya'an earthquake resulted in nearly 200 fatalities and over ten thousand injuries within a short period (Norio et al., 2011, Zhang et al., 2016, Emanuel, 2017, Geary et al., 2022). In the aftermath of such disasters, affected regions often descend into chaos, with large numbers of injured individuals in urgent need of medical attention. However, severe shortages of critical medical resources—such as medical personnel, ambulances, hospital beds, essential medications, and life-saving

equipment—pose significant challenges to effective rescue operations, exacerbating the risk of delayed treatment and increased mortality rates.

Recognizing the urgency of disaster response, governments worldwide have placed great emphasis on developing emergency medical rescue systems. As early as 2005, China introduced the National Emergency Response Plan for Public Incidents, which established a structured framework for emergency medical response, including its organizational system, operational mechanisms, and support measures. This was followed by the enactment of laws and regulations such as the Emergency Response Law and the National Emergency Plan for Natural Disaster Relief, aimed at improving the efficiency and effectiveness of medical rescue efforts (Lettieri et al., 2009). Internationally, the World Health Organization (WHO) has launched the Global Health Emergency Plan for 2025–2028, highlighting the crucial role of scientifically allocating medical resources in reducing disaster-related mortality (WHO, 2023). Despite increasing policy support, post-disaster emergency medical response continues to face numerous challenges, particularly in optimizing medical resource allocation and patient treatment strategies,

* Corresponding author.

E-mail addresses: cea_cuin@ujn.edu.cn (N. Cui), zx1525@student.bham.ac.uk (Z. Liang), luoj2@aston.ac.uk (J. Luo).

<https://doi.org/10.1016/j.cie.2025.111541>

Received 4 April 2025; Received in revised form 15 August 2025; Accepted 17 September 2025

Available online 25 September 2025

0360-8352/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

where many pressing issues remain unresolved.

Emergency medical service for post-disaster scenarios is a complex systemic operation that encompasses multiple aspects, including post-disaster medical resource allocation, hospital overload management, medical rescue network optimization, and emergency dispatch decision-making. Effective allocation of medical resources is critical for reducing disaster-related mortality and improving treatment efficiency. In recent years, advancements in operations research, mathematical modeling, and computer science have provided theoretical and practical approaches for addressing the challenges of post-disaster medical resource allocation. Research that integrating multi-objective optimization, multi-period dynamic planning, and intelligent optimization algorithms has becoming a popular field for enhancing medical rescue strategies and optimizing resource allocation.

Existing research on emergency medical services primarily focuses on ambulance dispatching, medical resource allocation, and patient transportation. Many studies have developed mathematical models, such as linear programming, mixed integer programming, and dynamic programming, to optimize the allocation of medical resources, with some studies incorporating heuristic algorithms for solution approaches (Talarico et al., 2015, Liu et al., 2019, Sun et al., 2021). However, most of these studies fail to adequately address the dynamic demand fluctuations after a disaster considering hospital overloaded and resource fairness, which are critical challenges in real-world emergency medical service.

Firstly, the demand for emergency medical services after a disaster is highly uncertain, making traditional static allocation models inadequate for adapting to rapidly changing medical resource needs. The shortage of medical resources in disaster areas, combined with the uncoordinated deployment of medical teams, may lead to traffic congestion, organizational chaos, and resource mismatches, which would further complicate rescue efforts (Kendra and Wachtendorf, 2003, Raj and Chinnadurai, 2021). Secondly, hospitals have limited treatment capacity, and unbalanced patient distribution may result in certain hospitals becoming overwhelmed, delaying critical care and potentially disrupting non-disaster-related medical services (Zoraster et al., 2007). Furthermore, the trade-off between treatment fairness and economic costs could prevent any further exploration. Addressing how to maximize the expected number of survivors while minimizing the total rescue costs is an important real-world problem which hasn't been addressed sufficiently in previous studies.

In recent years, advancements in multi-objective optimization and intelligent algorithms have provided new approaches to post-disaster medical resource allocation. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) has been widely applied in resource allocation and scheduling optimization due to its efficiency in solving multi-objective optimization problems (Deb et al., 2002, Aboud et al., 2022). However, existing studies still lack a systematic framework that comprehensively considers hospital overcrowding, resource fairness, multi-period dynamic optimization, and management and control of operational costs in emergency rescue operations.

In the context of post-disaster emergency rescue, the allocation of medical resources presents multiple complex challenges. First, optimizing casualty distribution is crucial to ensure that severely injured patients receive priority treatment while alleviating hospital overcrowding. Second, balancing resource fairness and treatment efficiency is essential to prevent excessive concentration of medical resources in specific areas, which may lead to inefficiencies. Additionally, leveraging multi-objective optimization techniques and intelligent algorithms to improve solution quality and enhance decision-making reliability is key to increasing rescue efficiency. Lastly, establishing a multi-period dynamic optimization framework to adapt to evolving medical demands and adjust rescue strategies over time is critical to ensuring the sustained effectiveness of emergency response efforts.

The objective of this study is to develop an efficient and systematic framework for post-disaster emergency medical resource allocation. To

address challenges such as inequitable medical resource distribution, hospital congestion, and treatment fairness, we develop a multi-objective optimization model implemented through the NSGA-II algorithm. We subsequently evaluate its performance against the Pareto Local Search (PLS) approach.

Building upon existing research, this study introduces a more comprehensive, dynamic, and efficient optimization framework for emergency medical resource allocation. The key contributions include:

- 1) Developing a dynamic post-disaster medical resource allocation model that integrates hospital congestion, resource fairness, and multi-period variations to enhance rescue efficiency.
- 2) Incorporating fairness and economic cost constraints to simultaneously maximize survival rates and optimize cost control, thereby improving rescue operation sustainability.
- 3) Implementing the NSGA-II optimization algorithm with a Pareto front search mechanism to improve solution quality and provide a set of feasible optimal solutions for emergency response.
- 4) Conducting empirical validation using real-world data from the 2013 Ya'an earthquake (Sichuan, China) to assess model applicability and the algorithm effectiveness, providing actionable insights for policymakers.

The structure of this paper is organized as follows: Section 2 reviews relevant literature and analyzes the strengths and limitations of existing research. Section 3 introduces the notations and mathematical formulations used in the model construction. Section 4 describes the operational mechanisms of post-disaster rescue operations. Section 5 provides a detailed explanation of the NSGA-II algorithm employed in this study. Section 6 validates the effectiveness of the proposed model and algorithm through numerical experiments and case studies. Section 7 concludes the research findings and discusses potential directions for future research.

2. Literature review

In the face of mass casualty incidents (MCIs) and surging demand for medical resources, optimal allocation of limited resources to enhance rescue efficiency is critical. Researchers globally have extensively studied this domain (Jotshi et al., 2009, Wex et al., 2014, Gu et al., 2018, Setiawan et al., 2019, Shetab-Boushehri et al., 2022). This section reviews literature closely aligned with our problem, focusing on multi-objective casualty transportation, dynamic resource scheduling, and hospital capacity management. We then explicitly contrast our work's innovations against existing gaps.

(1) Ambulance positioning and dispatching

To enable rapid response in disasters, studies have optimized ambulance positioning and redeployment. (Wang et al., 2014, Bélanger et al., 2020, Mohri and Haghshenas, 2021, Mahalakshmi et al., 2022, Hemici et al., 2023). Andersson and Värbrand (2007) introduced ambulance repositioning in scheduling problems, optimizing dispatch based on incident urgency and distance. Toro-Díaz et al., (2013) proposed a bi-level genetic algorithm-hypercube queuing framework to minimize response time while maximizing coverage. Recently, Hemici et al., (2023) developed a multi-objective evolutionary algorithm integrating simulated annealing for real-time ambulance dispatching during COVID-19, minimizing response time while maximizing emergency call coverage. A common limitation is the omission of hospital reception capacity, leading to congestion when casualties overwhelm finite resources (Zoraster et al., 2007).

(2) Casualty transportation and triage

Routing optimization for ambulance transportation has been widely explored (Zhang et al., 2017, Yoon and Albert, 2020, Sutherland and Chakraborty, 2023, Shiri et al., 2023). Najafi et al., (2014) designed a dynamic vehicle scheduling model for earthquakes but ignored injury severity, risking suboptimal survival outcomes. Talarico et al., (2015)

Table 1
Chronological comparison of literature (2014–2023) vs. our study.

Authors (Year)	Resource Constraints	Decisions	Survival Probability	Hospital Delay	Multi-period	Objectives	Solution Method	Disaster Scenario
Dean and Nair (2014)	A, H	AD, CT, CA	✓	×	×	Maximize expected survivors	Severity-Adjusted Model	Illustrative example
Liu et al. (2019)	A, H	AD, CT, CA	✓	×	×	Maximize survivors, minimize costs	ϵ -constraint method	Yushu earthquake
Chou et al. (2022)	A, H	AD, CT	×	✓	×	Minimize hospital/treatment delay penalties	Cell Transmission Model	Generic MCIs
Rathore et al. (2022)	A, M, H	AD, CT, CA	✓	×	✓	Maximize expected survivors	Severity-Adjusted Evacuation Model	Generic MCIs
Lin et al. (2023)	M, H	MD	×	✓	×	Improve regional healthcare accessibility	Patient diversion optimization	Post-earthquake
Hemici et al. (2023)	A	AD	×	×	×	Minimize response time, maximize coverage	MOEA/D + Simulated Annealing	COVID-19
Wan et al. (2023)	A	AD	×	×	✓	Minimize unmet demand cost, distribution risk	Fuzzy chance-constrained programming	Brazil floods (2011)
Chang et al. (2023)	A, H	AD, CT	×	×	✓	Minimize expected delivery time	Hybrid simulation optimization	Taiwan earthquake
Yazdani and Haghani (2023)	A, H	AD, CT	✓	×	×	Minimize response time, maximize coverage	MILP + Heuristics	Pandemic outbreaks
This study	A, M, H	AD, MD, CT, CA	✓	✓	✓	Minimize system costs, maximize survivors	NSGA-II + Simulation	Ya'an earthquake (2013)

Resource constraints considered: Ambulances (A), Medical personnel (M), Hospital capacity (H).

Decisions: Ambulance dispatch (AD), Medical staff dispatch (MD), Casualty transport (CT), Casualty allocation (CA).

optimized ambulance routing for triage-based cohorts (red/green codes) yet overlooked dynamic in-hospital treatment constraints. Recent work by Yazdani and Haghani (2023) integrated hospital capacity and multi-period evacuation, but hospital overcrowding penalties remained unaddressed. Crucially, under resource scarcity, triage-driven prioritization is essential (Frykberg, 2005, Raiter et al., 2008, Kim and Oh, 2023). Zhu et al., (2019) minimized deprivation costs in rescue routing using time windows to enforce fairness, while Rathore et al., (2022) embedded severity-adjusted survival probabilities in victim evacuation. Our model unifies these elements, i.e., triage, survival probability decay, and congestion penalties, within a dynamic framework.

(3) Medical personnel scheduling

Scheduling medical staff amid fluctuating demand is another key challenge (Rais and Viana, 2011, Van den Bergh et al., 2013, Defraeye and Van Nieuwenhuysse, 2016). Chen et al., (2016) optimized personnel allocation under uncertainty via a two-stage algorithm but omitted integration with casualty transportation. Lin et al., (2023) proposed regional medical demand reallocation strategies with casualty diversion, while Sun et al., (2023) modeled victim psychology in an evolutionary game-theoretic rescue framework. For nurse scheduling, Maenhout and Vanhoucke (2013) allocated staff across departments using column generation, yet disaster-specific dynamics (e.g., hospital overcrowding) were unaddressed.

(4) Multi-period resource allocation

Multi-period models dynamically coordinate relief distribution and casualty transport (Cao et al., 2018, Rodríguez-Espíndola et al., 2018, Yang et al., 2021, Vahdani et al., 2022, Zhang et al., 2023). Yi and Özdamar (2007) integrated location-distribution decisions using simple routing, while Sheu (2007) predicted time-varying demand via fuzzy clustering. Sheu and Pan (2014) later unified shelter, medical, and distribution networks into a multi-objective mixed-integer program. Recent advances include Wang et al. (2023) three-tier robust optimization for medical supply distribution and Chang et al. (2023) simulation-based casualty collection point allocation. However, none concurrently optimize medical personnel mobility, ambulance routing, and hospital congestion penalties across periods. While Wan et al. (2023) developed a fuzzy chance-constrained model for multi-period disaster planning, their focus was on relief supplies rather than medical human resources. Our model bridges these gaps by synchronizing

ambulance routing, medical personnel dispatch, and hospital resource utilization across time periods.

Table 1 chronologically compares key studies (2014–2023) with our approach, revealing some limitations in current literature: First, most existing models fail to account for hospital overcrowding penalties resulting from resource-patient mismatches (Liu et al., 2019; Chou et al., 2022). Second, while recent work, e.g., Chang et al. (2023) and Wan et al. (2023), has advanced dynamic relief distribution modeling, they overlook the crucial aspect of medical personnel redeployment across facilities. Third, no prior study has successfully integrated the minimization of system-wide costs (combining overcrowding, transportation, and penalty costs) with survivor maximization through Pareto optimization.

Our study addresses these gaps through three key advances: (1) development of a unified framework that incorporates overcrowding penalties, multi-period resource allocation, and NSGA-II-based Pareto optimization; (2) implementation of a discrete-event simulation that models detailed interactions between medical teams, ambulances, and casualties; and (3) empirical validation using 2013 Ya'an earthquake data, which yields practical staffing thresholds for emergency planners. These contributions collectively provide a more comprehensive approach to post-disaster medical resource allocation.

3. Problem description and model formulation

In this section, a mathematical model is proposed to optimize emergency medical resource allocation in mass casualty events. The model specifically take into consideration of hospital overcrowding, casualty transportation and assignment, as well as medical personnel and ambulance scheduling hospital capacity constraints and operational conditions are incorporated to enhance the efficiency and effectiveness of post-disaster medical service.

3.1. Problem statement

We consider a scenario in which a severe disaster occurs, resulting in a large number of casualties within a specific region. Let $J = \{1, 2, \dots, j\}$ represent the set of disaster areas, where casualties are discretely

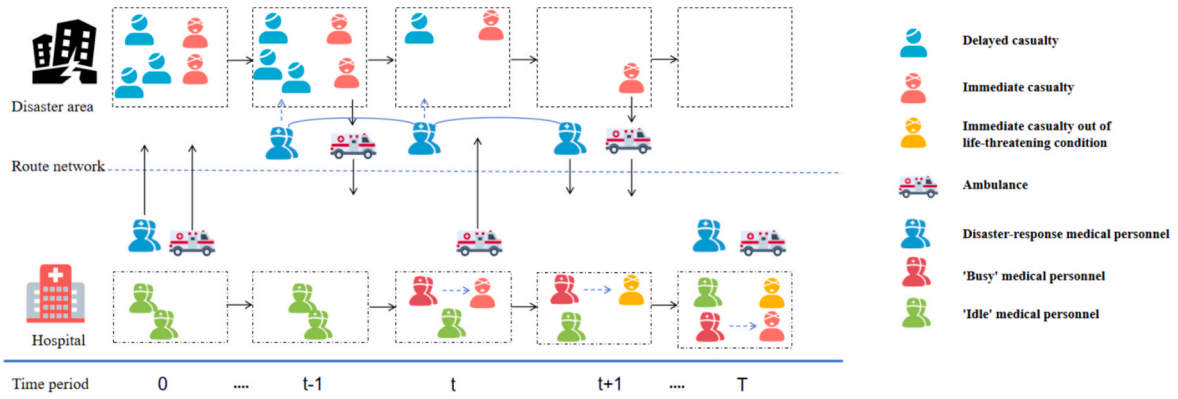


Fig. 1. The schematic diagram of emergency medical service operation.

distributed in groups. Similarly, let $I = \{1, 2, \dots, \hat{i}\}$ denote the set of hospitals located in proximity to these disaster areas. Each hospital i is equipped with a predefined number of medical personnel b_i (e.g., doctors and nurses) and ambulances n_i , which are available for emergency response operations. Rescue actions take place via a road network $G = (V, E)$, where V and E are the sets of nodes and directed arcs, respectively, and $I, J \in V$. The travel time for an ambulance at hospital i to reach disaster area j is π_{ij} . Since rescue activities typically continue for 72 h after a disaster or even for a long time thereafter, to characterize the continuity of medical rescue actions, we divide the entire rescue period into multiple time periods, denoted as $T = \{1, 2, \dots, \hat{t}\}$, and the duration of each time period is Δt .

At the start of recovering stage from a disaster, particularly a large-scale one, there is usually an imbalance between the high demand for medical treatment and the limited availability of medical resources. To enhance the efficiency of first aids and maximize the utilization of medical resources, the Simple Triage and Rapid Treatment (START) method is commonly used to assess the medical conditions of the injured which is able to rapidly classify them into four categories depending on the degree of injury: Minor (walking wounded), Delayed (requiring delayed treatment), Immediate (needing emergent treatment) and Deceased (Liu et al., 2019, Mills et al., 2013, Sung and Lee, 2016). Compared with Minor and Deceased cases, both Delayed and Immediate injuries have a higher need for medical care and treatment and inevitably require greater support from health care resources, especially immediate injuries need to be transported to hospitals for surgical treatment. Therefore, in this paper we focus primarily on these two categories of victims. We denote them by $\lambda = \text{Immediate}$, $\varepsilon = \text{Delayed}$ and define that (i) for Immediate casualties, due to the severity of their injuries, they will be transferred by ambulance to hospitals to receive more specialized and comprehensive treatment (e.g., surgery in the operating room); while (ii) for Delayed casualties, they are transferred to the safe zones within the disaster area to receive on-site medical treatment. Correspondingly, the medical personnel are thus divided into two groups: (i) Group I: staying at the hospital to operate on the Immediate casualties; and (ii) Group II: dispatched to the disaster area to provide on-site first aid for the Delayed casualties. The entire rescue operation is illustrated in Fig. 1.

The assumptions that proposed for the model are summarized as follows (Mills et al., 2013, Liu et al., 2019):

- 1) During each rescue time period, each individual medical worker can only opt to either stay at the hospital or go to the disaster area. Medical personnel working at the disaster area may return to the hospital after completing their rescue tasks.
- 2) All ambulances are homogeneous in nature, and their traveling speed is known. After finishing the rescue tasks, ambulances return to hospitals.

- 3) Medical personnel are dispatched to the disaster areas by regular buses rather than ambulances.
- 4) Ambulance preparation time as well as loading and unloading time are not considered in our model.
- 5) At the outset of rescue actions ($t = 0$), the numbers of Delayed and Immediate casualties have been determined separately, and no additional casualties will be added to the triage and waiting queue at each disaster area after the rescue commences ($t > 0$).

For the second assumption, all ambulances are homogeneous in nature, and their traveling speed is known. Travel time between nodes are modeled as deterministic values. This simplification aligns with strategic disaster operations literature (e.g., Talarico et al., 2015; Liu et al., 2019) to ensure initial model tractability and focus on system-level trade-offs among transportation logistics, hospital capacity, and survival outcomes. Homogeneous ambulances avoid combinatorial complexity in vehicle-specific routing, while deterministic travel time enable a baseline analysis of dynamic resource coordination. In reality, post-disaster road disruptions may cause travel time variability, and ambulances may differ in capabilities. To address these limitations, we will conduct comprehensive sensitivity analyses to validate our model's robustness. (1) Ambulance homogeneity: we will examine ambulance capacity heterogeneity in Section 6.4.2, demonstrating that the optimized strategy remains effective across varying ambulance specifications. (2) Travel time uncertainty: in Section 6.4.3, we will evaluate the impacts of realistic travel time fluctuations ($\pm 30\%$) on rescue outcomes, confirming the solution's stability under actual operational conditions.

The assumption of no new casualties after $t = 0$ (Assumption 5) simplifies the temporal dynamics but aligns with empirical disaster response patterns. In the 2013 Ya'an case, >95 % of identified casualties were registered within the first 12 h (Sichuan Health Commission Report¹), allowing our multi-period model to capture the critical rescue phase. Although casualties may be discovered progressively, their treatment is synchronized with resource deployment – a dynamic implicitly reflected in our time-varying medical personnel allocation and casualty evacuation. This approach is consistent with prior studies addressing large-scale disasters (Liu et al., 2019, Sung and Lee, 2016). We will validate the model's outputs against actual outcomes from the Ya'an case study in Section 6.3 to demonstrate how this deterministic approach produces operationally viable emergency response plans.

The integrated workflow of our methodology is depicted in Fig. 2, encompassing data inputs, modeling, optimization, and decision outputs.

¹ <https://www.sc.gov.cn/>.

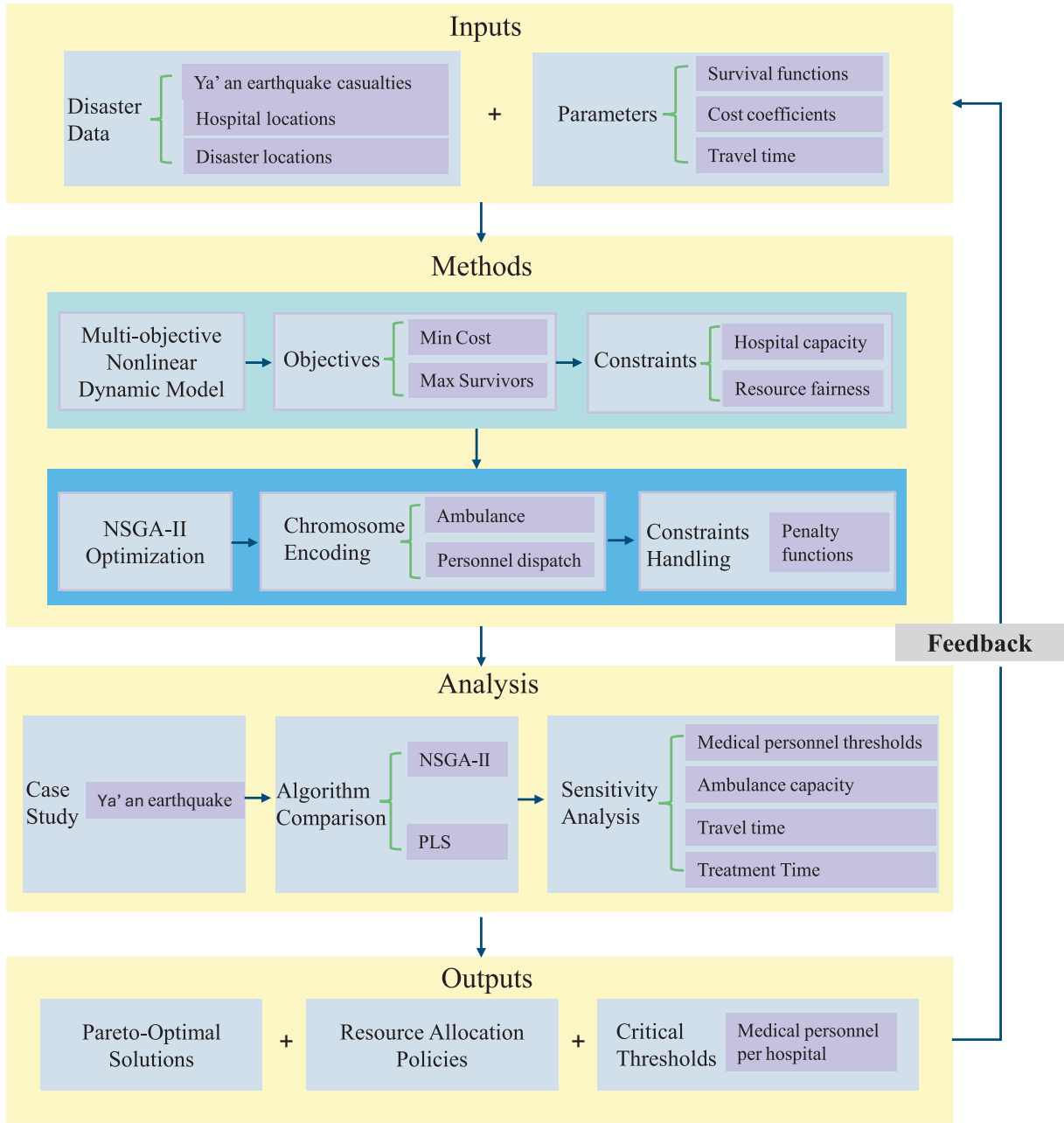


Fig. 2. Integrated Workflow of the Proposed Methodology.

3.2. Model formulation

Our model seeks to balance two conflict objectives: minimizing overall system rescue costs and maximizing the expected number of survivors. By explicitly considering this trade-off, the model aims to enhance the efficiency of emergency medical resource allocation while ensuring the effective treatment of casualties in mass-casualty events. Suppose $f_{\lambda}(t)$ and $f_{\varepsilon}(t)$ denote the survival probabilities of Immediate and Delayed casualties, respectively, as defined by the survival probability function below (Mills et al., 2013).

$$f_{\lambda}(t) = \frac{\beta_{0,\lambda}}{(t/\beta_{1,\lambda})^{\beta_{2,\lambda}} + 1} \quad (1)$$

$$f_{\varepsilon}(t) = \frac{\beta_{0,\varepsilon}}{(t/\beta_{1,\varepsilon})^{\beta_{2,\varepsilon}} + 1} \quad (2)$$

where $\beta_{k,\lambda}, \beta_{k,\varepsilon} > 0$, for $k = 0, 1, 2$. Several studies have demonstrated that timely medical treatment is crucial for casualties and that the survival probability of untreated casualties diminishes over time (Sacco et al., 2005, Sacco et al., 2007, Mizumoto et al., 2011). Although quantified survival probability may be achievable in some specific medical conditions where the survival probability function can be estimated based on the physical condition of the casualties, for other more complex conditions, estimating the survival probability function is difficult. Consequently, we adopted similar assumptions as in previous studies and directly utilized the values of $\beta_{k,\lambda}$ and $\beta_{k,\varepsilon}$ (Jacobson et al., 2012, Mills et al., 2013, Dean and Nair, 2014, Sung and Lee, 2016, Liu et al., 2019).

Following a disaster, the number of casualties in the disaster area rises sharply, necessitating the rapid and efficient allocation of medical resources to address varying triage needs. For Immediate-category casualties, ambulances shuttle cyclically between disaster areas and hos-

pitals to ensure timely medical intervention. Positive integer variables v_{ij}^t are introduced to denote the number of ambulances dispatched from hospital i to disaster area j for Immediate casualties during time period t . Given that the unit travel cost of an ambulance is C , the total operational cost of ambulance deployment can be formulated as follows.

$$S(v) = \sum_{t=1}^T \sum_{i \in I} \sum_{j \in J} C(v_{ij}^t * 2\pi_{ij}) \quad (3)$$

Hospital ambulance and medical staff allocations adjust dynamically based on immediate injury treatment demands. Therefore, there is a set of variables associated with the dynamic rescue operations in hospital i during time period t , including the number of available medical personnel, y_i^t , the number of ambulances, Z_i^t , the number of Immediate casualties requiring treatment, $p_{i\lambda}^t$, and the remaining number of untreated Immediate casualties, $u_{i\lambda}^t$. The relationship between $p_{i\lambda}^t$ and $u_{i\lambda}^t$ is presented as follows, where a_i represents the unit treatment capacity of medical personnel for Immediate casualties.

$$u_{i\lambda}^t = p_{i\lambda}^t - a_i * y_i^t, \forall i \in I, t \in [1, T] \quad (4)$$

The massive casualty influx inevitably causes hospital overcrowding and resource shortages, leading to treatment delays. To mitigate these congestion effects, we develop a treatment impedance function based on medical penalty metrics like the Injury Severity Score (ISS), which quantifies overcrowding penalties as follows:

$$f_i^t(y, p) = F_i + \eta * F_i * \left(\frac{p_{i\lambda}^t}{y_i^t * a_i} \right)^2, \forall i \in I, t \in [1, T] \quad (5)$$

Where F_i represents the treatment cost for hospital i when there is no congestion, and η is a modulation factor between the hospital's free flow treatment and overloading treatment needs. The overcrowding in the hospital can be express as a quadratic relationship between the total number of casualties requiring treatment and the number of available medical personnel in each time period t , i.e., $\frac{p_{i\lambda}^t}{y_i^t * a_i}$.

For delayed casualties (less severe injuries), medical personnel can be deployed to disaster areas for on-site treatment, optimizing resource utilization and reducing hospital burden. After completing field operations, these personnel may either: (1) return to treat immediate casualties in hospitals, or (2) be redeployed to other disaster sites. This creates dynamic fluctuations in personnel distribution between hospitals and disaster areas across time periods. Assuming that positive integer variables X_{ij}^t are denoted as the number of medical personnel dispatched from hospital $i \in I$ to disaster area $j \in J$ in time period t , and the unit travel cost is M_{ij}^t , then the total travel cost of medical personnel is presented as follows.

$$W(X) = \sum_{t=1}^T \sum_{i \in I} \sum_{j \in J} M_{ij}^t X_{ij}^t \quad (6)$$

When medical personnel work in disaster areas, there is a high risk of exposure to disasters. Let variables d_j^t denote the number of medical personnel working at disaster area j in time period t and the unit cost of exposure duration for medical personnel is ω_j^t , the total exposure cost of medical personnel is presented as follows.

$$E(d) = \sum_{t=1}^T \sum_{j \in J} d_j^t * \omega_j^t \Delta t \quad (7)$$

To maximize the number of Immediate and Delayed casualties receiving timely treatment, a penalty cost is introduced to quantify the impact of unsatisfied medical demand resulting from resource shortages. Let $\gamma_{j\lambda}^t$ and γ_{je}^t denote the number of untreated Immediate and Delayed casualties in disaster area j at time period t respectively, and δ_i and δ_e be their corresponding unit penalty costs. Then, the total penalty cost can be written as follows.

Table 2

Nomenclature of the model.

Notation	Description
Sets	
I	Hospitals, $i \in I$
J	Disaster areas, $j \in J$
Parameters	
T	The length of the planning horizon (the amount of time interval), $t \in [1, T]$
Q	The capacity of an ambulance for accommodating Immediate casualties.
π_{ij}	Travel time between hospital i and disaster area j .
n_i	The Initial number of ambulances in hospital i .
b_i	The Initial number of medical personnel in hospital i .
δ_i	Unit penalty cost for untreated Immediate casualties.
δ_e	Unit penalty cost for untreated Delayed casualties.
a_i	The average number of Immediate casualties that can be treated by each medical personnel member.
a_e	The average number of Delayed casualties that can be treated by each medical personnel member.
N_j^i	The total number of Immediate casualties in area j .
N_j^e	The total number of Delayed casualties in area j .
C	Unit time cost of ambulances.
ω_j^t	The risk cost of medical personnel traveling to disaster area j at time period t .
M_{ij}^t	The travel cost of medical personnel traveling to disaster area j at time period t .
Decision variables	
X_{ij}^t	The number of medical personnel dispatched from hospital i to disaster area j at time period t .
v_{ij}^t	The number of ambulances transporting from hospital i to disaster area j at time period t .
y_i^t	The number of remaining medical personnel in hospital i at time period t .
Z_i^t	The number of remaining ambulances in hospital i at time period t .
$\gamma_{j\lambda}^t$	The number of untreated Immediate casualties at disaster area j at time period t .
γ_{je}^t	The number of untreated Delayed casualties at disaster area j at time period t .
$u_{i\lambda}^t$	The number of Immediate casualties at hospital i that have not been treated at time period t .
$p_{i\lambda}^t$	The number of Immediate casualties at hospital i requiring treatment at time period t .
d_j^t	The number of medical personnel working at disaster area j at time period t .
H_i^t	The number of medical personnel occupied in treating the Immediate casualties in hospital i at time period t .

$$P(\gamma) = \sum_{t=1}^T \sum_{j \in J} (\delta_i * \gamma_{j\lambda}^t + \delta_e * \gamma_{je}^t) \quad (8)$$

The sets, parameters and decision variables employed in problem statement and mathematical modeling are listed in Table 2.

Therefore, a multi-objective nonlinear dynamic model is formulated to minimize system rescue costs while maximizing the expected number of survivors. The formal mathematical definitions of these two objective functions are presented below.

$$\begin{aligned} \min Z_1 = & \sum_{t=1}^T \sum_{i \in I} f_i^t(y, p) \sum_{t=1}^T \sum_{i \in I} \sum_{j \in J} (C * (v_{ij}^t * 2\pi_{ij}) + M_{ij}^t * X_{ij}^t) \\ & + \sum_{t=1}^T \sum_{j \in J} d_j^t * \omega_j^t \Delta t + \sum_{t=1}^T \sum_{j \in J} (\delta_i * \gamma_{j\lambda}^t + \delta_e * \gamma_{je}^t) \end{aligned} \quad (9)$$

$$\max Z_2 = \sum_{t=1}^T \left(\sum_{i \in I} f_{\lambda}(t) * a_i * H_i^t + \sum_{j \in J} f_e(t) * a_e * d_j^t \right) \quad (10)$$

$$s.t. Z_i^0 = n_i - \sum_{j \in J} v_{ij}^0, \forall i \in I \quad (11)$$

$$y_i^0 = b_i - \sum_{j \in J} X_{ij}^0, \forall i \in I \quad (12)$$

$$d_j^0 = 0, \forall j \in J \quad (13)$$

$$u_{ia}^0 = 0, \forall i \in I \quad (14)$$

$$p_{ia}^t = u_{ia}^{t-1} + \sum_{j \in J: t \geq 2\pi_{ij}} v_{ij}^{t-2\pi_{ij}} * Q, \forall i \in I, t \in [1, T] \quad (15)$$

$$u_{ia}^t = p_{ia}^t - a_{ia} * H_i^t, \forall i \in I, t \in [1, T] \quad (16)$$

$$\gamma_{ja}^0 = N_j^i, \forall j \in J \quad (17)$$

$$\gamma_{je}^0 = N_j^e, \forall j \in J \quad (18)$$

$$\gamma_{ja}^t = \gamma_{ja}^{t-1} - v_{ij}^{t-\pi_{ij}} * Q, \forall j \in J, t \in [1, T] \quad (19)$$

$$\gamma_{je}^t = \gamma_{je}^{t-1} - a_e * d_j^{t-\pi_{ij}}, \forall j \in J, t \in [1, T] \quad (20)$$

$$Z_i^t = Z_i^{t-1} - \sum_{j \in J} v_{ij}^t + \sum_{j \in J: t \geq 2\pi_{ij}} v_{ij}^{t-2\pi_{ij}}, \forall i \in I, t \in [1, T] \quad (21)$$

$$y_i^t = y_i^{t-1} - \sum_{j \in J} x_{ij}^t + \sum_{j \in J: t \geq \pi_{ij}} x_{ji}^{t-\pi_{ij}}, \forall i \in I, t \in [1, T] \quad (22)$$

$$d_j^t = d_j^{t-1} + \sum_{i \in I: t \geq \pi_{ij}} x_{ij}^{t-\pi_{ij}} - \sum_{i \in I} x_{ji}^t, \forall j \in J, t \in [1, T] \quad (23)$$

$$H_i^t \leq y_i^t, \forall i \in I, t \in [1, T] \quad (24)$$

$$\sum_{i \in I} \sum_{t=1}^T x_{ij}^t > 0, \forall j \in J \quad (25)$$

$$x_{ij}^t, v_{ij}^t, y_i^t, z_i^t, \gamma_{ja}^t, \gamma_{je}^t, u_{ia}^t, p_{ia}^t, d_j^t \in N^+, \forall i \in I, \forall j \in J, t \in [1, T] \quad (26)$$

The first objective function (9) aims to minimize the total system rescue costs, which consist of five key components: the penalty cost associated with hospital overcrowding, the total operating cost of ambulances $S(v)$, the total travel cost of medical personnel $W(X)$, the total exposure risk cost of medical personnel $E(d)$, and the total penalty cost for untreated casualties $P(\gamma)$. The second objective function (10) is designed to maximize the expected number of survivors after the disaster. Here, $\sum_{i \in I} f_{ia}(t) * a_{ia} * H_i^t$ and $\sum_{j \in J} f_{je}(t) * a_e * d_j^t$ denote respectively the number of expected Immediate and Delayed casualties treated at hospitals and in disaster areas across all the time periods.

The constraints primarily govern the dynamic status of ambulances, medical personnel, and casualties, as well as the capacity limitations of hospitals. Constraints (11) and (12) ensure the initial conservation of ambulances and medical personnel at each hospital immediately after dispatch at time period 0. Constraints (13) and (14) specify that, at the onset of emergency medical services, no medical personnel are initially stationed at disaster areas, and no Immediate casualties remain untreated in hospitals. Constraints (15) define the number of Immediate casualties awaiting treatment at each time period t , while constraints (16) calculate the number of untreated Immediate casualties at the end of time period t . Constraints (17) and (18) show the initial number of Immediate and Delayed casualties at each disaster area. Meanwhile, Constraints (19) and (20) describe the dynamic evolution of Immediate and Delayed casualty numbers over time at disaster areas. Constraints (21) maintain ambulance conservation at hospitals across consecutive time periods. Constraint (22) characterizes the availability of medical personnel at hospitals in each time period, which is determined by the sum of the remaining personnel from the previous period and those returning from disaster areas, minus the personnel dispatched in the current period. Similarly, Constraints (23) describe the availability of medical personnel at disaster areas in each time period. Constraints (24) reveal the relationship between the occupied and total available medical

personnel in hospitals. Constraints (25) ensure that Delayed casualties at disaster areas receive treatment throughout all time periods. Finally, Constraints (26) enforce the non-negativity of decision variables.

The proposed post-disaster emergency medical resource allocation problem is formulated as a bi-objective, non-convex, nonlinear mixed-integer program with inherent NP-hard complexity. The model's incorporation of medical personnel scheduling and dynamic, multi-period resource allocation substantially increases its computational challenges. The hospital congestion function introduces additional nonlinearity, further complicating the optimization. To address these challenges, we implement the NSGA-II algorithm and validate the model's effectiveness through post-disaster rescue simulations, with framework and algorithmic details presented in subsequent sections.

4. Simulation framework for evaluating rescue strategies

In this study, simulation techniques are employed to analyze the actual rescue process in scenarios involving mass casualties caused by large-scale natural disasters. Based on this rescue process, we further describe the strategic medical resource allocation model developed in this research, with a particular focus on key decision-making aspects, including the allocation of medical personnel, casualty triage, ambulance dispatching, and the distribution of casualties among hospitals—all of which are critical components of emergency medical services. The discrete-event simulation (DES) model was developed in Python 3.9 using native coding to ensure seamless integration with the optimization formulation, which was implemented with an Intel Core i5-11320H CPU (3.20 GHz) and 16 GB of RAM. The simulation tracked rescue operations over a horizon of $T = 10$ h, with optimization decisions updated hourly ($\Delta t = 1$) and state transitions evaluated per minute ($\Delta t_s = 1$). This granularity captured dynamic interactions among ambulances, personnel, and casualties, as governed by the logic in Code 1.

Following large-scale disasters, medical personnel first gather at designated hospitals to await deployment. Concurrently, field teams triage casualties into four severity groups. Those classified as Immediate or Delayed are stabilized at secure locations pending hospital transfer.

Step 1. Based on casualty distributions across disaster areas, each hospital must determine ambulance and medical staff allocations for dispatch at time period t , while retaining sufficient personnel for incoming emergencies. Ambulances transport Immediate casualties to hospitals, while dispatched staff remain on-site to treat Delayed casualties.

Step 2. Once medical allocations are determined, ambulances and staff must travel from hospitals to assigned disaster areas. Given potential road damage and traffic congestion that may delay rescues, optimal route planning is essential. We preemptively identify the shortest hospital-to-disaster area paths to minimize time loss during critical rescue operations.

Step 3. Following the scheduled transit period, ambulances and medical teams arrive at their assigned disaster areas. Ambulances transport Immediate casualties back to hospitals via pre-determined routes, with each vehicle constrained by its patient capacity (Mills et al., 2018). Meanwhile, medical personnel treat Delayed casualties on-site according to their individual capabilities before returning to their base hospitals.

Step 4. Upon ambulance arrival at the destination hospital, Immediate casualties are admitted for treatment and the ambulance status changes from 'busy' to 'idle'. The system then verifies whether any disaster sites still have untreated Immediate casualties. If present, the process continues to Step 6; otherwise, it proceeds to Step 8.

Step 5. When Immediate casualties arrive at the hospital, available ('idle') medical staff provide emergency treatment, with their status changing to 'busy' during care. Upon treatment completion, they revert to 'idle' and return to Step 5. If no staff are available, patients wait until the next provider is free.

Step 6. Ambulance deployment decisions must consider current

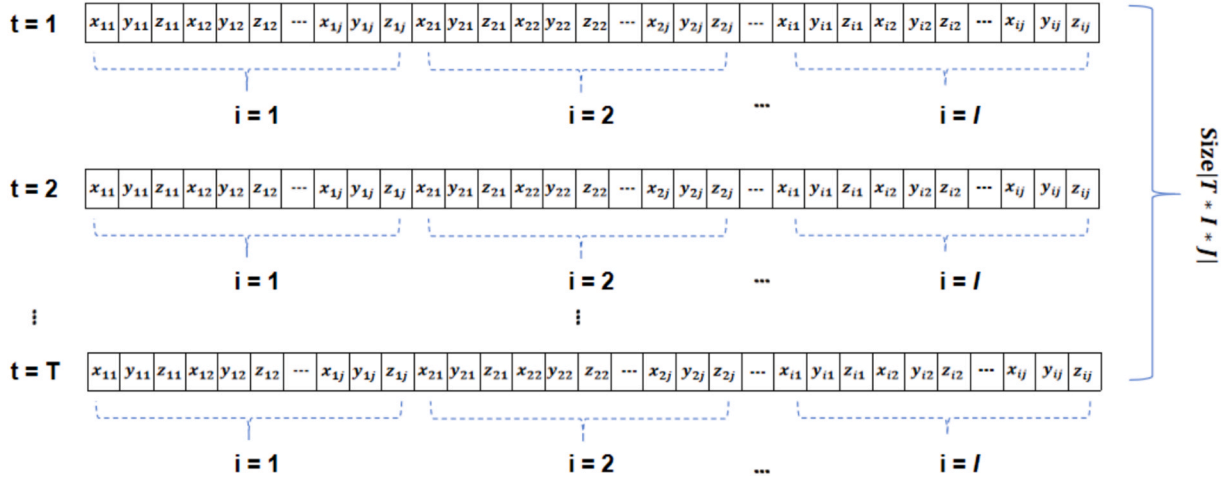


Fig. 3. The example of chromosome structure.

hospital capacity, including the number of Immediate casualties awaiting treatment and available medical staff. If capacity permits, 'idle' ambulances are redeployed to disaster areas for Immediate casualty transport (returning to Step 1). Otherwise, the process advances to Step 7.

Step 7. When hospital capacity for emergency admissions is exhausted, ambulances remain on standby ('idle') at the hospital until capacity becomes available, at which point they resume operations (returning to Step 4).

Step 8. The ambulance finishes the transportation mission.

The simulation framework serves solely to evaluate the objective values (total rescue cost and expected survivors) of resource allocation plans generated by the optimization model. It emulates real-world rescue dynamics (e.g., ambulance routing, hospital congestion) but does not replace the optimization process.

Code 1 Simulation Pseudo-code for Emergency Rescue Process.

Input: P (Population), T (Time periods), I (Number of hospitals), J (Number of disaster areas) $light[j]$ (Number of delayed casualties at disaster area j)
 $\pi[i][j]$ (Travel time from hospital i to disaster area j)
 $rb[j][t]$ (Number of medical personnel at disaster area j at time period t)
 $ambu[i]$ (Number of ambulances in hospital i)
 $doc[i][t]$ (Number of medical personnel in hospital i at time period t)
 $sum[i][t]$ (Immediate casualties transported to hospital i during period t)
 $severe[j]$ (Immediate casualties at disaster area j)
 $uncured[i][t]$ (Immediate casualties not treated in hospital i)
 a (doctors' treatment capabilities), Q (ambulance capacity)
Decision variables: X^* (Ambulance from hospital i to disaster area j at time period t)
 Y^* (medical personnel traveling from hospital i to disaster area j at time period t)
 Z^* (medical personnel returning to hospital i from disaster area j at time period t)
1: **function** $evaluate_p$
2: $result \leftarrow []$
3: **for** individual $\leftarrow 1$ to Population **do** –Randomly generate initial solution $\{X^*, Y^*, Z^*\}$.
4: **for** $t \leftarrow 1$ to T **do**
5: **for** $i \leftarrow 1$ to I **do**
6: **for** $j \leftarrow 1$ to J **do**
7: **if** $light[j] > 0$ and $t - 1 \geq \pi[i][j]$ **then** –Check if there are delayed casualties needing treatment at the disaster area and if medical personnel have arrived.
8: $Update\ rb[j][t] \leftarrow Y^*$ –Assign medical personnel to disaster area
9: **if** $Z^* \leq rb[j][t]$ **then** –Conditions met for medical personnel to return to the hospital.
10: $Update\ Z^*$
11: **else then**
12: **Trigger penalty** –Trigger penalty if conditions are not met.
13: **elif** $light[j] \leq 0$ **then**
14: **if** $\pi[i][j] \leq T - t$ **then**
15: $Update\ Z^*$ –Rescue operation is complete, medical personnel can return.
16: $Update\ X^*$ and Y^*
17: $Update\ doc[i][t] \leftarrow Z^*$ –Update the medical personnel in the hospital.

(continued on next column)

(continued)

Code 1 Simulation Pseudo-code for Emergency Rescue Process.

18: **if** $X^* > ambu[i]$ or $Y^* > doc[i][t]$ **then** –Check if resources exceed available ambulances or personnel.
19: **Trigger penalty** –Exceeds available resources.
20: **if** $X^* > ambu[i]$ and $Y^* \leq doc[i][t]$ **then** –Valid resource allocation.
21: $doc[i][t] \leftarrow doc[i][t] - Y^*$
22: $uncured[i][t] \leftarrow sum[i][t] - doc[i][t] * a$ –Update uncured casualties.
23: **elif** $Y^* > doc[i][t]$ and $X^* \leq ambu[i]$ **then**
24: $ambu[i] \leftarrow ambu[i] - X^*$
25: $severe[j][t] \leftarrow severe[j][t] - X^* * Q$ –Calculate immediate casualties at disaster area.
26: $Update\ ambu[i], sum[i][t], severe[j], doc[i][t], uncured[i][t]$
27: **else then**
28: $doc[i][t] \leftarrow doc[i][t] - Y^*$
29: $ambu[i] \leftarrow ambu[i] - X^*$
30: $severe[j] \leftarrow severe[j] - X^* * Q$
31: $uncured[i][t] \leftarrow sum[i][t] - doc[i][t] * a$
32: $Update\ ambu[i], sum[i][t], severe[j], doc[i][t], uncured[i][t]$
33: **for** $j \leftarrow 1$ to J **do**
34: $light[j] \leftarrow rb[j][t] * a$ –Calculate delayed casualties at disaster area.
35: $Update\ light[j]$
36: **if** $sum(rb[j]) < 0$ **then** –If no medical personnel available at disaster area, trigger penalty.
37: **Trigger penalty**
38: **end if**
39: **end for**
40: $Calculate\ the\ objective\ function\ F1, F2\ by\ Eq.\ (9)\ (10).$
41: $Objective\ function\ plus\ penalty\ value$
42: $result \leftarrow result \cup [F1, F2]$
43: **return result**
44: **end function**

5. Implementation of the NSGA-II algorithm for multi-objective optimization

The post-disaster emergency medical resource allocation model proposed in this study involves NP-hard problems related to casualty transportation and allocation, as well as the dispatch of medical personnel and ambulance, all of which are embedded within a dynamic multi-period decision-making framework. Additionally, the hospital overcrowding penalty function exhibits strong nonlinearity, further complicating the model's solvability. To efficiently address the challenges posed by this multi-objective nonlinear dynamic optimization model, the NSGA-II algorithm was employed as the solution approach. The NSGA-II-generated solutions were validated through a Python-executed DES that emulated real-world rescue dynamics.

The NSGA-II algorithm is a classic multi-objective evolutionary algorithm. The general process for obtaining the optimal Pareto front is as follows: Initially, a population of solutions is randomly generated and

Table 3

Initialization of a randomly generated chromosome.

	<i>i</i> = 1									<i>i</i> = 2								
	<i>j</i> = 1			<i>j</i> = 2			<i>j</i> = 3			<i>j</i> = 1			<i>j</i> = 2			<i>j</i> = 3		
<i>t</i> = 1	5	8	0	1	2	0	4	10	0	2	8	0	2	7	0	0	4	0
<i>t</i> = 2	3	6	0	6	2	0	0	2	0	8	2	0	1	0	0	1	3	0
<i>t</i> = 3	5	4	2	2	2	0	3	2	0	2	0	0	6	3	2	2	6	4

initialized. Subsequently, mutation and crossover operations are applied to the initial solution set to produce multiple new candidate solutions. Secondly, the candidate solutions undergo non-dominated sorting, which classifies them into different ranks based on their dominance relationships; solutions that are non-dominated are considered to be of higher quality and are thus retained. For solutions within the same rank, crowding distance is used as a metric to eliminate inferior solutions. Finally, based on the rank and crowding distance of the solution set, the best solutions are selected from the candidates to form the next generation of solution sets. Through the iteration of these steps, NSGA-II is capable of progressively generating a series of solutions that approximate the Pareto front as the result of multi-objective optimization (Deb et al., 2002, Zheng and Doerr, 2023).

5.1. Chromosome representation and population initialization

Let the set $\{x, y, z\}$ denote the key decision variables in the model, where x and y respectively represent the number of ambulances and medical personnel from hospital i to disaster area j during time period t , and z represents the number of medical personnel returning from disaster area j to hospital i during time period t . In accordance with the definition of decision variables, the chromosomes are encoded using a real-number representation scheme, and a population consists of a set of chromosomes. At the commencement of the evolutionary process, each chromosome is randomly initialized to ensure diversity within the population. Each chromosome represents a medical rescue scheduling scheme, encompassing key decision variables such as x , y , and z . The structure of the chromosome is shown in Fig. 3.

Assuming the model comprises 2 hospitals, 3 disaster areas, and 3 rescue periods, each hospital is equipped with a fleet of 10 ambulances and 20 medical personnel. Table 3 presents an initialization of a randomly generated chromosome, where each row corresponds to a rescue period. Taking the third row as an example, [5, 4, 2] indicates that in the third time period, 5 ambulances and 4 medical personnel are dispatched from hospital 1 to disaster area 1, while 2 medical personnel return to hospital 1 from disaster area 1; [2, 2, 0] signifies that during the third time period, 2 ambulances and 2 medical personnel depart from hospital 1 to disaster area 2, with no medical personnel returning to hospital 1 from disaster area 2 during the same time period; [2, 6, 4] denotes that 2 ambulances and 6 medical personnel are dispatched from hospital 2 to disaster area 3, and 4 medical personnel return within the same period. During evolution, each chromosome (resource allocation plan) is evaluated via DES (Section 4) to compute F1 and F2.

5.2. Evolutionary operators

In the NSGA-II algorithm, two primary evolutionary operators are employed: the Simulated Binary Crossover (SBX) operator and the Polynomial Mutation (PM) operator.

The SBX operator effectively explores new solutions within a continuous search space $[x_l^i, x_u^i]$ and maintains population diversity. Initially, two parent individuals are selected from the current population, denoted as $P^1 = (x_1^1, x_2^1, \dots, x_n^1)$ and $P^2 = (x_1^2, x_2^2, \dots, x_n^2)$, which are n -dimensional. The disparity between the parents is assessed using Eq. (27). This step is crucial for preventing premature convergence of the solution set due to the genetic operators, thus ensuring that the search is

Table 4Values of x_l and x_u for decision variables in this experiment.

	x	y	z
x_l	1	7	0
x_u	19	40	10

not confined to a narrow region.

$$\delta = \begin{cases} x_i^2 - x_i^1, x_i^2 - x_i^1 \geq 10^{-10} \\ 10^{-10}, \text{otherwise} \end{cases} \quad (27)$$

The SBX crossover operator is applied to the two selected parent individuals and the crossover parameters $\beta_{q_i}^k$ are computed to adjust the degree of crossover according to Eq. (28).

$$\beta_{q_i}^k = \begin{cases} (rand \times \alpha_i^k)^{1/(1+\eta)}, rand \leq \frac{1}{\alpha_i^k} \\ (1/2 - rand \times \alpha_i^k)^{1/(1+\eta)}, \text{otherwise} \end{cases} \quad (28)$$

where the distribution index of the crossover operator η is an arbitrary non-negative real number; the larger the value of η , the higher the probability of generating a 'near-parent' solution. Here, $\eta=90$. $rand$ is a random real number in the range $[0,1]$ and $rand \times \alpha_i^k$ is used to ensure that $\beta_{q_i}^k$ is not greater than 1. α_i^k is a random parameter for executing crossover of gene i for individual k , $\alpha_i^k = 2 - \frac{1}{\beta_i^{k+1+\eta}}$, $k \in \{1, 2\}$, where β_i^k is calculated by Eq. (29).

$$\begin{cases} \beta_i^1 = 1 + \left(2 \times \frac{x_i^1 - x_i^l}{\delta}\right) \\ \beta_i^2 = 1 + \left(2 \times \frac{x_i^2 - x_i^l}{\delta}\right) \end{cases} \quad (29)$$

The offspring individuals C_i^1 and C_i^2 are obtained from Eq. (30).

$$SBX(P^1, P^2) = \begin{cases} C_i^1 = 0.5[(x_i^1 + x_i^2) - \beta_{q_i}^1 \cdot \delta] \\ C_i^2 = 0.5[(x_i^1 + x_i^2) + \beta_{q_i}^2 \cdot \delta] \end{cases} \quad (30)$$

In the Polynomial Mutation (PM) operator, the individual gene values are adjusted based on random numbers using polynomial probability distributions (Eqs. (31) and (32)), with small perturbations applied within the specified variable range $[x_l, x_u]$.

$$p_{k+1} = p_k + \delta_k(x_u - x_l) \quad (31)$$

$$\delta_k = \begin{cases} (2 \times rand_k)^{\frac{1}{(1+\eta_m)}} - 1, rand_k \leq 0.5 \\ 1 - \left[2(1 - rand_k)^{\frac{1}{(1+\eta_m)}}\right], \text{otherwise} \end{cases} \quad (32)$$

Here, $rand_k$ is a value randomly generated from a uniform distribution within the range $[0, 1]$; η_m is the mutation index; x_u and x_l are the lower and upper bounds of the gene, representing the valid range of gene values; p_k is the selected individual from the k -th generation; p_{k+1} is the individual obtained after the mutation operation in the $(k+1)$ -th generation.

The values of x_l and x_u set in this experiment are shown in Table 4.

Table 5

Constraints	Violations	Penalty terms
(23)	The number of medical personnel returning from disaster area j to hospital i exceeds the number of medical personnel from hospital i present at disaster area j .	$\begin{cases} \text{vio}_{f1} = \text{vio}_{f1} + \text{punish}_{f1} \\ \text{vio}_{f2} = \text{vio}_{f2} + \text{punish}_{f2} \end{cases}$
(21)	The number of ambulances dispatched from hospital i to all disaster areas j exceeds the number of ambulances available within hospital i .	$\begin{cases} \text{vio}_{f1} = \text{vio}_{f1} + \text{punish}_{f1} \\ \text{vio}_{f2} = \text{vio}_{f2} + \text{punish}_{f2} \end{cases}$
(22)	The number of medical personnel dispatched from hospital i to all disaster areas j exceeds the number of medical personnel within hospital i .	$\begin{cases} \text{vio}_{f1} = \text{vio}_{f1} + \text{punish}_{f1} \\ \text{vio}_{f2} = \text{vio}_{f2} + \text{punish}_{f2} \end{cases}$
(25)	At time period t , the number of medical personnel at all disaster areas j is negative.	$\begin{cases} \text{vio}_{f1} = \text{vio}_{f1} + \text{punish}_{f1} \\ \text{vio}_{f2} = \text{vio}_{f2} + \text{punish}_{f2} \end{cases}$

The SBX and PM operators were tuned through systematic parameter sensitivity analyses (Section 6.2.1). This ensured optimal exploration–exploitation trade-offs, with P_c controlling solution diversity and P_m maintaining population convergence properties.

5.3. Constraint handling strategy

During the solution process, to ensure that the algorithm generates solutions that are both feasible and continuously optimized, this paper adopts a penalty strategy to handle individuals that violate constraint conditions. When an individual breaches these constraints, the penalty term introduced into the objective function diminishes its competitive ability during the selection and evolution phases. Specifically, by assigning a substantial penalty coefficient, the two objective functions are combined with the penalty coefficient through a weighted aggregation, leading to a significant deterioration in the performance of individuals who violate constraints on both objective functions, thereby naturally eliminating them from the optimization process. This penalty strategy not only effectively enhances the convergence speed and solution quality of the algorithm but also aids in identifying an optimal solution set that meets the constraint conditions within the framework of multi-objective optimization. The specific idea is delineated as follows.

The two objective functions constructed in this paper are ($\min f_1$, $\max f_2$). Given the distinct optimization structures of the two objectives,

to facilitate the solution process, the maximization problem of the second objective function is transformed into a minimization problem, denoted as $\min -f_2$.

A. Establish penalty terms,

$\begin{cases} \text{vio}_{f1} = 0 \\ \text{vio}_{f2} = 0 \end{cases}$, punish_{f1} , punish_{f2} two exceptionally large penalty values.

B. Impose penalties on individuals that breach the constraint conditions. For each instance of violation of constraints a, b, c, or d as shown in Table 5, the penalty terms vio_{f1} and vio_{f2} are cumulatively incremented.

C. Introduce penalty terms in the objective functions, as shown in Eq. (33). The accumulated penalty terms are denoted by vio_{f1} and vio_{f2} .

$$\begin{cases} \min F_1 = \min f_1 + \text{vio}_{f1} \\ \min F_2 = \min -f_2 - \text{vio}_{f2} \end{cases} \quad (33)$$

5.4. Performance evaluation indicator

This paper employs the HyperVolume (HV) metric to assess the quality of the generated Pareto front. The HV measure evaluates the diversity and coverage of the Pareto front by calculating the volume between a reference point set in the objective space and the area covered by the Pareto front. In a two-dimensional objective space, the reference point is defined as $\text{ref_point} = [\text{ref1}, \text{ref2}]$, which corresponds to the maximum vertex of the objective space. In this study, both ref1 and ref2 are assigned a value of 10. There are also two critical boundary points on the Pareto front: the Ideal Point and the Nadir Point. The Ideal Point is obtained by optimizing each objective function individually to find the minimum solution values, while the Nadir Point represents the set of maximum optimization solutions that satisfy the constraints. Typically, the Ideal Point is denoted as $\text{ideal} = [\text{ideal1}, \text{ideal2}]$, and the Nadir Point is denoted as $\text{nadir} = [\text{nadir1}, \text{nadir2}]$.

The HyperVolume (HV) calculation is formulated as:

$$HV = \text{Volume}(PF, \text{ref_point}, \text{ideal}, \text{nadir})$$

Where Volume denotes the function for calculating the volume, and PF represents the set of solutions on the Pareto front. A larger HV value indicates a higher quality of the generated Pareto front.

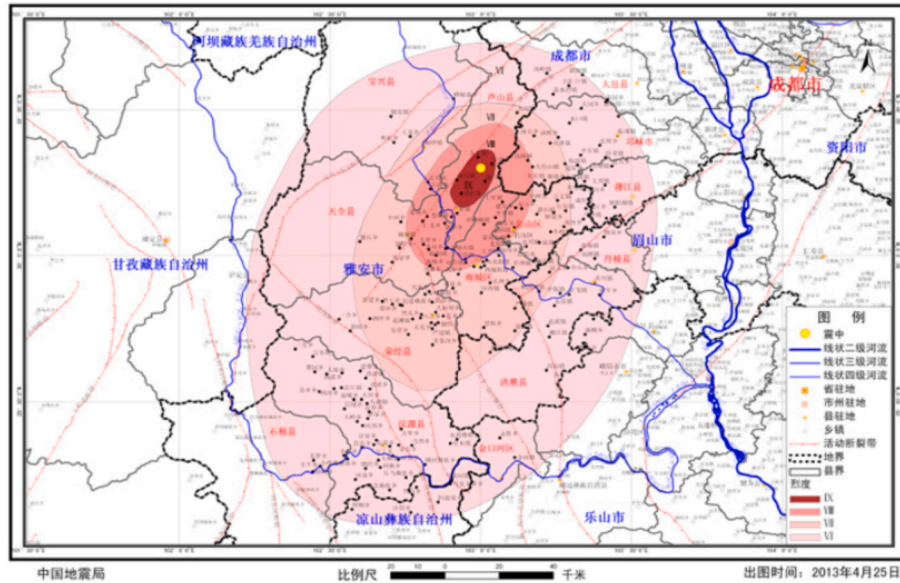


Fig. 4. Seismic intensity map of Ya'an Earthquake in Sichuan Province, China, 2013.

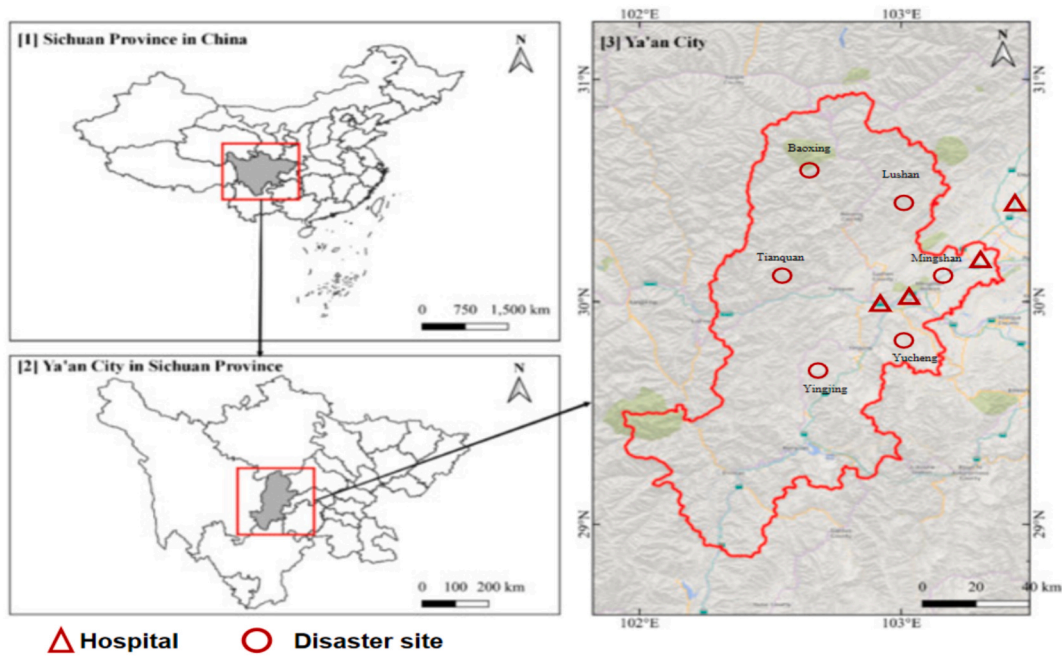


Fig. 5. Geographical locations of selected disaster areas and hospitals in the case study.

6. Case study

To validate the model's applicability in large-scale disaster scenarios, we analyze floods, hurricanes, and tsunamis as representative events, which universally exhibit three critical characteristics: (1) medical demand surges exceeding local capacity due to mass casualties (Lettieri et al., 2009); (2) infrastructure disruption triggering cascading economic losses and impeding resource deployment (Kendra and Wachtendorf, 2003); and (3) spatially heterogeneous casualty distribution necessitating dynamic resource reallocation across fragmented zones (Sheu, 2007). Among these, earthquakes typically induce the most severe human impacts, characterized by extreme casualty volumes and geographical spread. The 2013 Ya'an earthquake (M7.1) epitomizes these challenges: affecting > 15,000 km² within 72 h (Fig. 4, China Earthquake Administration), causing 196 fatalities and > 10,000 injuries, and generating heterogeneous medical demands across fragmented regions. Consequently, we adopt the Ya'an event as a validation benchmark — its extreme scale and complexity demonstrating the model's extensibility to cross-disaster mass-casualty incidents (MCIs). Through this case study, we test the feasibility of the proposed emergency medical resource allocation model, along with the performance and solution efficiency of the NSGA-II algorithm. Numerical experiments were conducted in Python on a laptop (Intel i5-11320H CPU, 3.20 GHz, 16 GB RAM).

6.1. Dataset and model parameter configuration

This study constructs an experimental network for case analysis by selecting six severely affected areas and four hospitals that participated in the rescue operations during the 2013 Ya'an earthquake. Specifically, the six disaster areas are Lushan County, Baoxing County, Yucheng District, Tianquan County, Mingshan District, and Yingjing County; the four hospitals are Ya'an People's Hospital, The 4th People's Hospital of Ya'an City, Mingshan People's Hospital, and West China Hospital of Sichuan University. The geographical locations of these disaster areas and hospitals are detailed in Fig. 5 (Zhang and Zheng, 2024). Furthermore, the model assumes medical personnel apply a "one-to-one" treatment protocol for Immediate casualties and a "one-to-four" protocol for Delayed casualties (Frykberg, 2005). This corresponds to unit

service rates of 1 and 4, respectively. This setting aligns with the international START triage protocol (Sacco et al., 2007) and WHO disaster medicine guidelines (Organization, 2021). Empirical data from 12,578 medical records during the 2013 Sichuan Ya'an earthquake¹ show actual ratios averaged 1:1.03 for Immediate casualties (3.2 % above the model) and 1:3.84 for Delayed casualties (4.1 % below the model). These values closely match the model's "one-to-one" and "one-to-four" assumptions, validating the ratio's feasibility in real-world mass casualty events. Additionally, Mills et al., (2013) identified this configuration through Pareto analysis as the key strategy for optimizing overall casualty survival rates and resource utilization efficiency in disaster scenarios. The ambulance travel cost per hour is ¥10; the risk exposure cost for medical personnel at the disaster area is estimated at ¥15 per person; given that medical personnel are transported to the disaster area via bus, the unit travel cost is designated as ¥5 per person; and for casualties that remain untreated, the unit penalty cost is assessed at ¥150. The parameter data referenced in this case study are obtained from the People's Government of Sichuan Province and the Sichuan Provincial Health Commission.²

This study employs the framework proposed by Mills et al. (2013), which defines five emergency rescue scenarios. These scenarios primarily differ in their level of 'pessimism' regarding the scale of mass-casualty incidents and the urgency of casualty treatment. Each scenario corresponds to a distinct underlying distribution of casualty severity, resulting in unique sets of survival probability functions. Specifically, in Scenario 1, the survival probabilities are low, and most of the casualties are immediate; in Scenario 5, survival probabilities are much larger, and there are not many immediate casualties; and Scenarios 2–4 are in the middle in terms of the severity of the event. Liu et al. (2019) simplified this framework into three representative scenarios: Pessimistic, Moderate, and Optimistic. To enable finer modeling of severity-dependent survival functions while preserving original scenario distinctions, this study reconstructs survival function parameters $\beta_{e,\lambda}$ and $\beta_{e,e}$ for all five scenarios. This reconstruction leverages linear interpolation based on Mills et al., (2013) original framework and Liu et al., (2019) simplified framework. Both parameters $\beta_{e,\lambda}$ and $\beta_{e,e}$ exhibit

² <https://www.sc.gov.cn/>.

Table 6

Values of survival probability parameters for different disaster scenarios (Mills et al., 2013, Liu et al., 2019).

Scenarios	$\beta_{0,\lambda}$	$\beta_{1,\lambda}$	$\beta_{2,\lambda}$	$\beta_{0,\varepsilon}$	$\beta_{1,\varepsilon}$	$\beta_{2,\varepsilon}$
Scenario 1	0.09	17	1.01	0.57	61	2.03
Scenario 2	0.17	32	1.16	0.67	99.5	2.10
Scenario 3	0.24	47	1.30	0.76	138	2.17
Scenario 4	0.40	69	1.44	0.79	149	2.29
Scenario 5	0.56	91	1.58	0.81	160	2.41

increasing values as scenario severity decreases (i.e., from Scenario 1 to Scenario 5), as shown in Table 6.

6.2. Performance evaluation of NSGA-II algorithm

In this section, we first optimize the key parameter settings of the NSGA-II algorithm, such as P_c and P_m , through a series of experiments to ensure its best performance. Then, a comparative analysis between the NSGA-II and PLS algorithms is conducted to further validate the efficiency of the proposed approach.

6.2.1. Parameter setting

To establish an optimal NSGA-II configuration, we first systematically analyzed the impact of population size (N) under a fixed computational budget of 200,000 function evaluations. Five population sizes (N = 50, 100, 200, 400, 500) were tested through 10 independent replications per configuration, with HV quantifying Pareto front convergence and diversity. Boxplot analysis (Fig. 6) revealed that small populations (N = 50) yielded the lowest median HV due to insufficient diversity limiting search space exploration. While medium-sized populations (N = 100) achieved median HV values comparable to N = 200, their larger interquartile range (IQR) indicated poorer solution stability. Large populations (N = 400, 500) demonstrated significantly reduced median HV values and maximal IQR, reflecting increased performance volatility and convergence limitations. The N = 200 configuration delivered optimal balance, achieving the highest median HV (99.78) and smallest IQR – signifying superior solution quality and stability. This population size additionally permitted 1,000 generations within the evaluation budget, enabling effective exploration–exploitation balance. Consequently, subsequent experiments adopted N = 200 with 1,000 generations based on empirical validation of solution quality, stability, and computational efficiency.

Subsequently, we conducted extensive experiments to determine the

optimal combination of crossover probability (P_c) and mutation probability (P_m). The ranges $P_c \in [0.1, 1.0]$ (step 0.1) and $P_m \in [0.001, 0.01]$ (step 0.001) were selected based on standard practices in evolutionary algorithms (Deb et al., 2002). The population size was set to 200, and the number of iterations was fixed at 1000. For each combination of P_c and P_m , the algorithm was executed independently 10 times, and the corresponding HyperVolume (HV) values were recorded.

The results, illustrated in Fig. 7, indicate that the HV value was closest to 100 when $P_c = 0.8$ and $P_m = 0.005$, suggesting that this parameter configuration yields the best performance for the NSGA-II

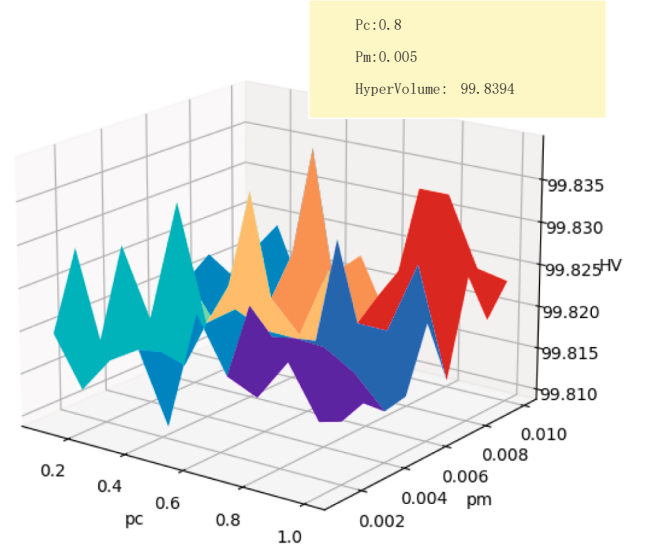


Fig. 7. The HV values of NSGA-II with different combinations of P_c and P_m .

Table 7

Experimental results.

Scenarios	Total system rescue cost (RMB)	Total expected number of survivors
Scenario 1	4078459.87	5344.47
Scenario 2	4113956.62	6321.47
Scenario 3	4132987.37	7088.34
Scenario 4	4257787.61	7680.50
Scenario 5	4341882.78	8086.60

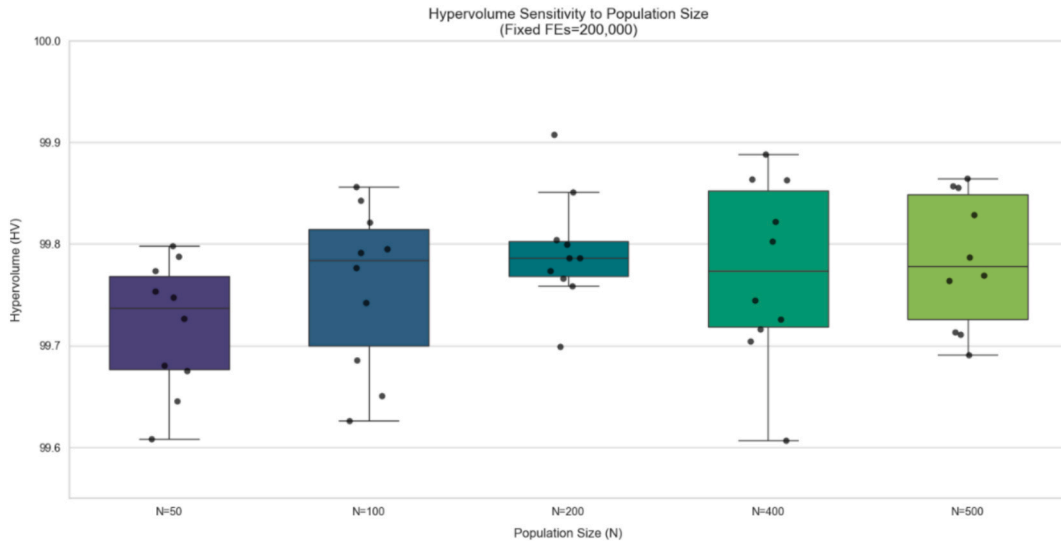


Fig. 6. HV distribution under different population sizes (Fixed FEs = 200,000).

Table 8
NSGA-II parameter settings.

Parameter	NSGA-II
Population size	200
Crossover rate	0.8
Mutation rate	0.005
Maximum generation	1000
Chromosome length	792

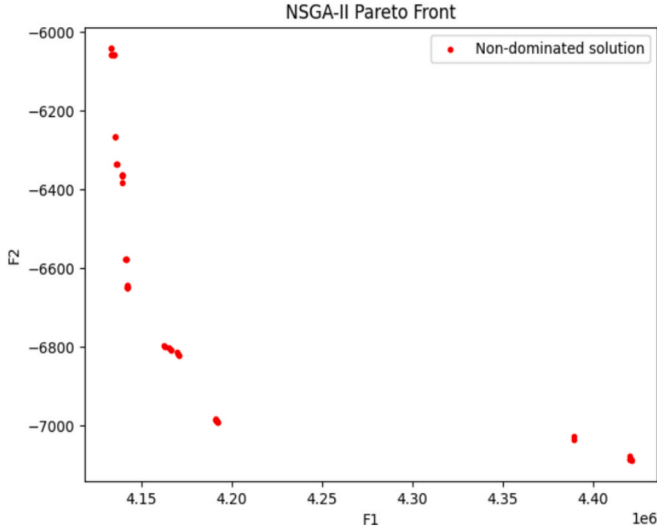


Fig. 8. NSGA-II Pareto Front.

algorithm. Consequently, these optimal settings ($P_c = 0.8$, $P_m = 0.005$) will be adopted in subsequent experiments.

The NSGA-II algorithm was employed to solve the model and analyze the impact of parameters $\beta_{e,\lambda}$ and $\beta_{e,\varepsilon}$ in the survival probability function on the model's objective function values across different emergency rescue scenarios. Table 7 presents the total system rescue cost and the expected number of survivors under five distinct rescue scenarios. Analysis of the data in Table 7 reveals that as the rescue scenario shifts from Scenario 1 (pessimistic) to Scenario 5 (optimistic), the expected number of survivors increases significantly, while the total system rescue cost also gradually rises. This result suggests that emergency managers can evaluate the 'degree of pessimism' in rescue scenarios based on their preferences for the two decision objectives, enabling flexible adjustments to rescue strategies. By aligning rescue plans with actual needs, optimal rescue outcomes can be achieved. Given that Scenario 3 represents a moderate case, we focus our subsequent analysis on the results for this scenario.

6.2.2. Comparison of algorithm performance

The proposed model is a multi-period mixed-integer nonlinear program (MINLP) with NP-hard complexity, evidenced by

- Objective (9): Non-convex hospital congestion penalty function, which introduces quadratic nonlinearity.
- Constraints (15)–(23): Dynamic resource flow and casualty allocation form a combinatorial decision tree.

Time complexity of one solution evaluation (simulation is $O(T \cdot I \cdot J)$, where $T = 10$ (time periods), $I = 4$ (hospitals), $J = 6$ (disaster areas)). NSGA-II addresses this complexity via polynomial-time operations per generation: $O(MN^2)$ (Deb et al., 2002), where $M = 2$ objectives and $N = 200$ population size.

To assess the solution efficiency of the NSGA-II algorithm, a comparative analysis was conducted with the Pareto Local Search (PLS)

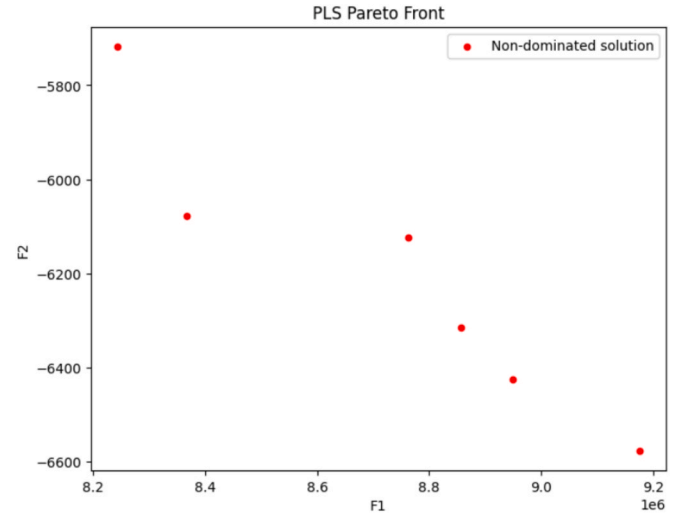


Fig. 9. PLS Pareto Front.

Table 9
Comparison results.

Algorithm	NSGA-II (This study)	PLS
CPU (s)	313.71	409.45
HyperVolume value	99.85	99.17
min f_1 (RMB)	4132987.37	8243262.57
max f_2 (Expected number of survivors)	7088.34	6576.21

algorithm. The PLS algorithm, a multi-objective local search algorithm, employs local search strategies to find high-quality solutions in the vicinity of the Pareto front. To ensure a fair and effective comparison, the same model parameters, chromosome representation, and population initialization were used in both algorithm settings. The parameter settings for the NSGA-II are presented in Table 8, while in the PLS algorithm, maxEvaluations is set to 200,000 and samples to 200.

Figs. 8 and 9 depict the Pareto fronts obtained using the NSGA-II and PLS algorithms, respectively, where the X-axis represents the total system rescue cost, and the Y-axis indicates the expected number of survivors after the disaster. Table 9 provides a detailed comparison of the results obtained from both algorithms.

The comparative analysis indicates that the NSGA-II algorithm achieves not only a lower total system rescue cost but also a higher expected number of survivors than the PLS algorithm. The Pareto front generated by NSGA-II is significantly superior, demonstrating that the solutions produced by the PLS algorithm are dominated by those of NSGA-II. This confirms the NSGA-II algorithm's capability to identify more effective multi-objective optimization solutions. Furthermore, the computational efficiency and solution quality of both algorithms are assessed. The CPU time for NSGA-II and PLS is 313.71 s and 409.45 s, respectively, while the solution quality, measured by HyperVolume (HV) values, is 99.85 for NSGA-II and 99.17 for PLS. These results further validate the superior performance of the NSGA-II algorithm over the PLS algorithm.

6.2.3. NSGA-II algorithm performance

To comprehensively validate the convergence behavior of the NSGA-II algorithm, we monitored the progression of both objective functions and the HyperVolume (HV) metric across generations using the optimal parameters identified in Section 6.2.1 ($N = 200$, $P_c = 0.8$, $P_m = 0.005$) under Scenario 3. Fig. 10 present NSGA-II HyperVolume Trajectory with two integrated components:

- Objective Function Progression (Upper Subplot):

NSGA-II HyperVolume Trajectory

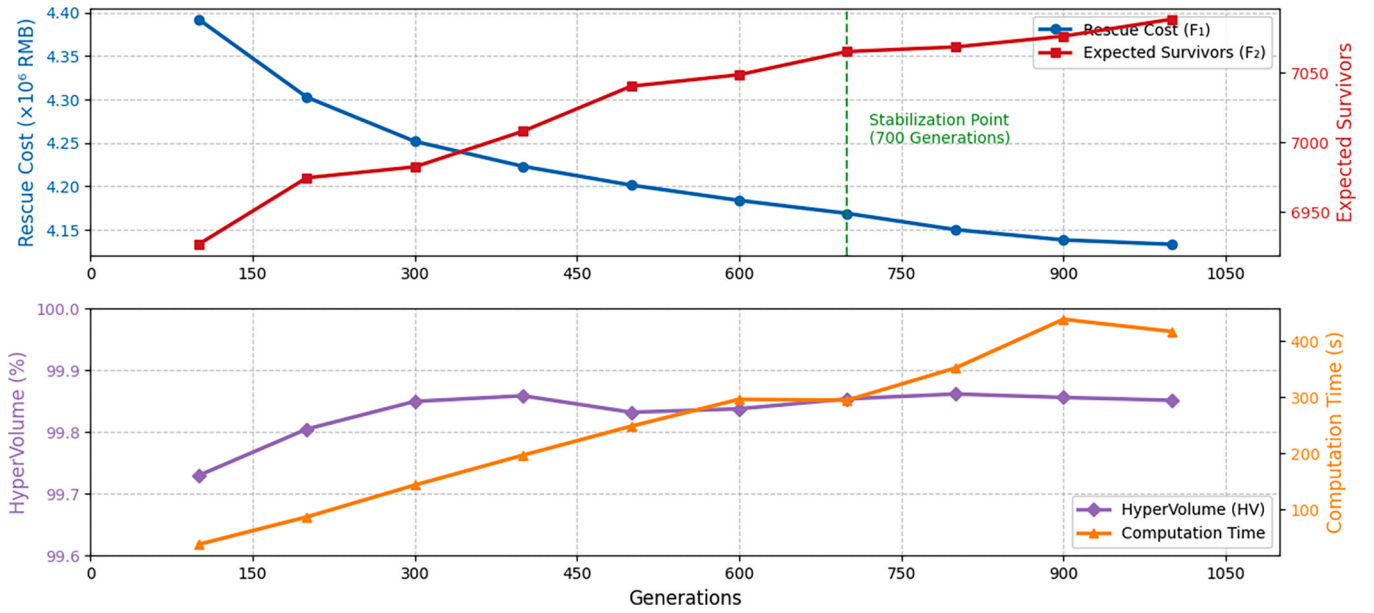


Fig. 10. NSGA-II HyperVolume Trajectory.

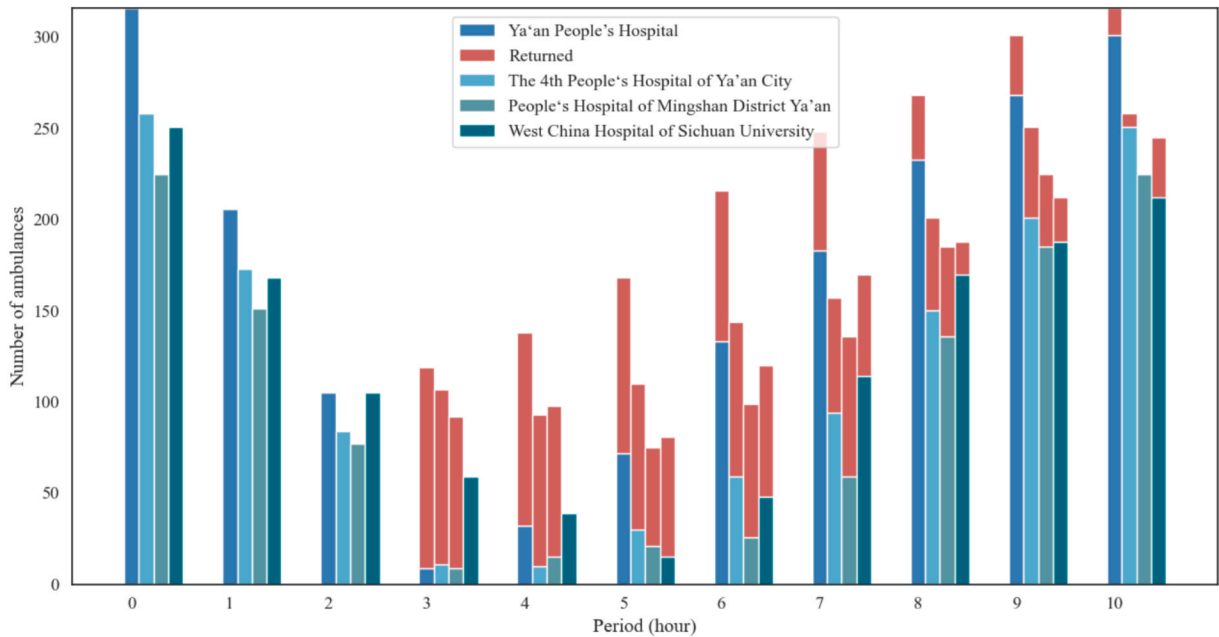


Fig. 11. Dynamic variations in the number of ambulances at each hospital over time.

- 1) Minimum rescue cost (F1) decreased sharply by 5.9 % within the first 300 generations, achieving 90 % of its total reduction during this phase.
 - 2) Maximum expected survivors (F2) increased by 2.3 % in early stages, reaching 98.5 % of optimal value by generation 400.
 - 3) Both objectives stabilized after 700 generations (vertical dashed line), with marginal fluctuations ($<0.5\%$) in subsequent iterations.
- HyperVolume and Computation Time (Lower Subplot):
- 1) HV exhibited exponential growth during generations 0–300 (1.2 % increase), reflecting rapid Pareto front refinement.

- 2) Convergence threshold (HV improvement $< 0.01\%$ /generation) was achieved at generation 750 (vertical dashed line), attaining 99.85 % maximal HV.

This analysis confirms NSGA-II's capability to efficiently navigate the complex solution space of our nonlinear dynamic model, delivering stable Pareto-optimal solutions within 750 generations. Notably, NSGA-II converged 40 % faster than PLS (750 vs. 1,200 generations) while achieving superior solution quality (HV 99.85 vs. 99.17), demonstrating its efficacy for time-sensitive emergency decision-making.

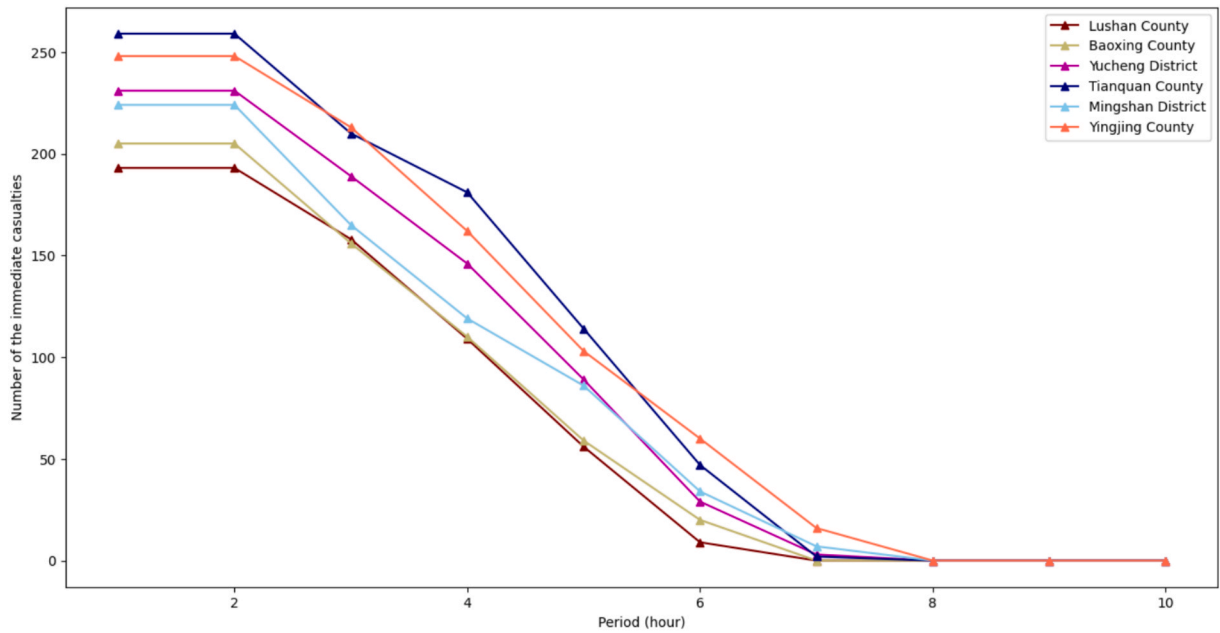


Fig. 12. Dynamic changes in the number of Immediate casualties at each disaster area over time.

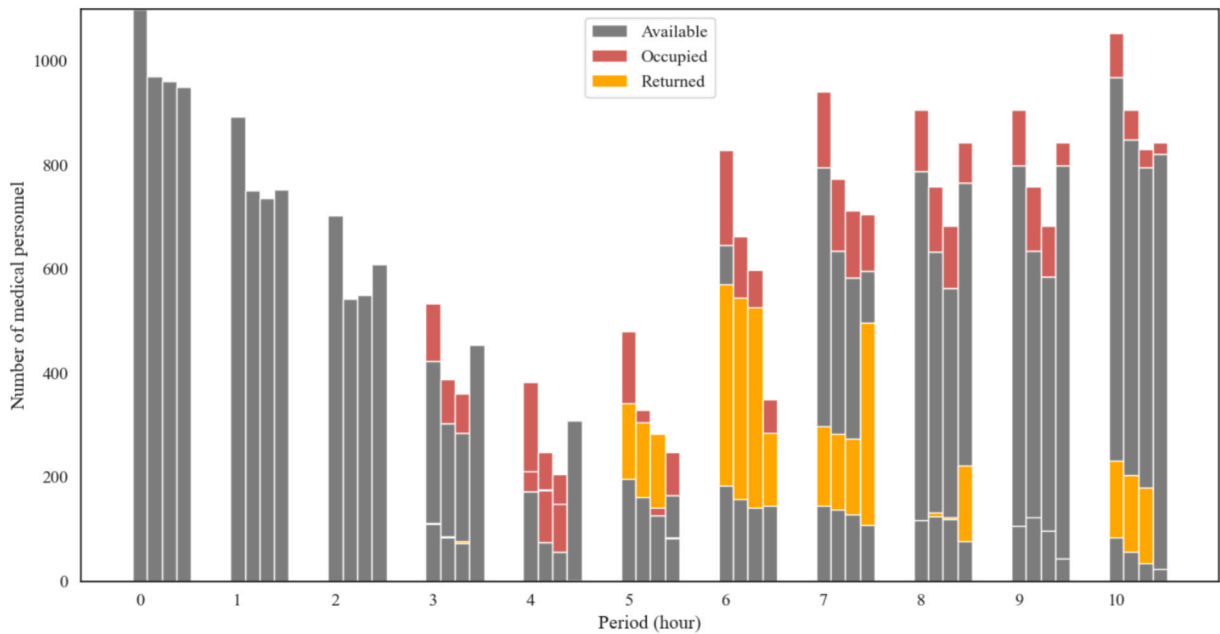


Fig. 13. Dynamic variations in the number of medical personnel at each hospital over time.

6.3. Result analysis

In the context of Immediate casualty rescue operations, Fig. 11 illustrates the dynamic variations in the number of ambulances across four hospitals throughout the rescue period. At time period $t = 0$, the initial allocation of ambulances to each hospital is depicted, representing the starting point of the rescue mission. As operations progress, hospitals dispatch ambulances in successive waves to transport Immediate casualties from disaster areas. Starting from time period $t = 3$, ambulances (marked in red) begin returning to their respective hospitals with casualties for treatment. These returning ambulances, together with those that remain idle at hospitals, constitute a new wave of rescue resources. They are subsequently reassigned to additional rescue missions, ensuring continuous operational efficiency until the entire rescue

effort is successfully completed.

Fig. 12 depicts the dynamic changes in the number of Immediate casualties across six disaster areas throughout the rescue period. Before time period $t = 2$, the number of Immediate casualties in each disaster area remains unchanged, primarily due to the travel time required for rescue vehicles to reach the affected disaster areas from hospitals. After time period $t = 2$, the number of casualties at the disaster areas gradually declines as emergency medical treatment is administered. By time period $t = 8$, all disaster areas have been effectively cleared of Immediate casualties, resulting in the cessation of ambulance dispatches and the subsequent return of all ambulances to their respective hospitals.

In this process, the scheduling of medical personnel is closely linked to the number of Immediate casualties awaiting treatment in hospitals and the number of Delayed casualties in various disaster areas, as

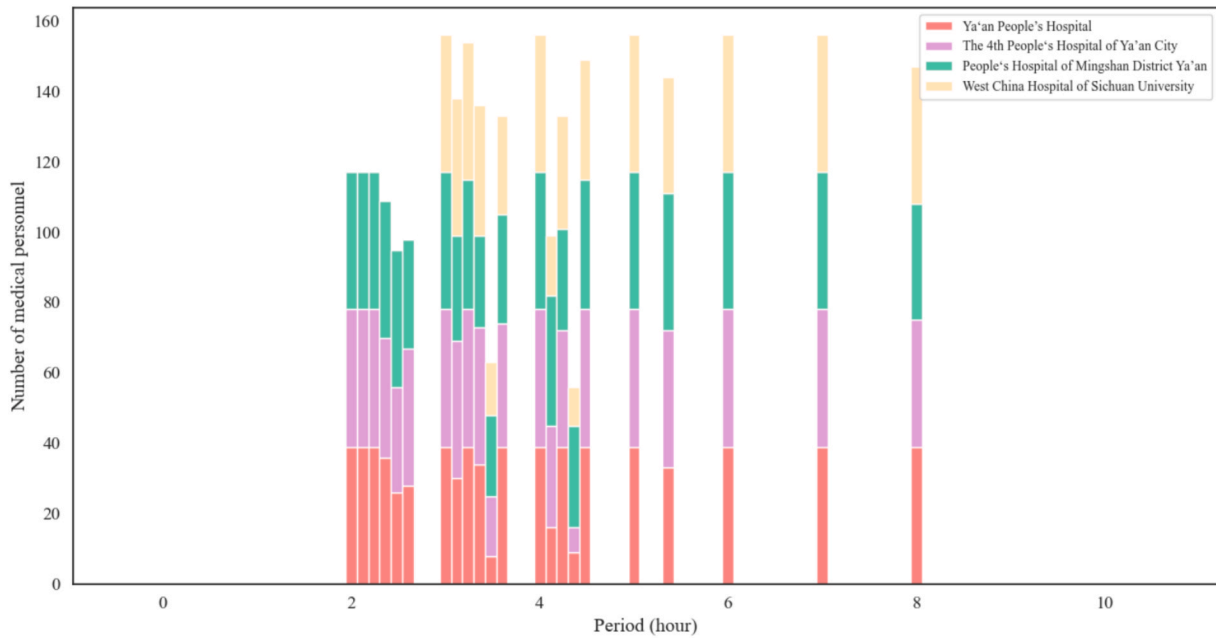


Fig. 14. Dynamic allocation of medical personnel from hospitals to disaster areas over time.

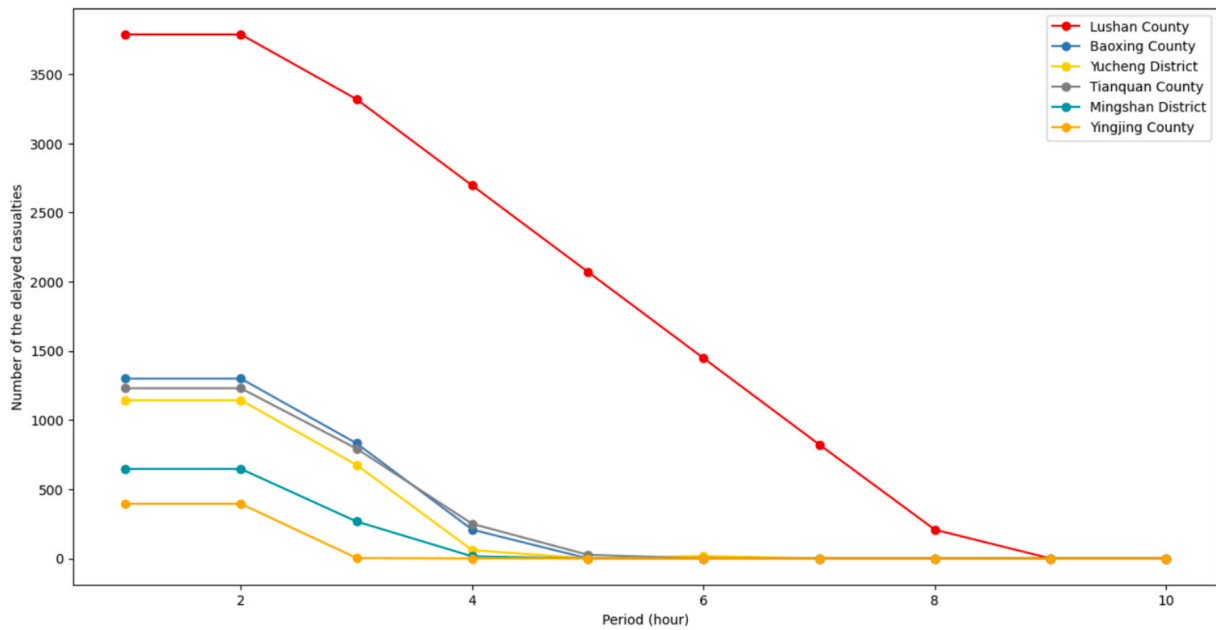


Fig. 15. Dynamic changes in the number of Delayed casualties at each disaster area.

illustrated in Fig. 13. Taking time period $t = 3$ as an example, Ya'an People's Hospital, the 4th People's Hospital of Ya'an City, and People's Hospital of Mingshan District Ya'an received 110, 85, and 74 Immediate casualties for treatment, respectively, while West China Hospital of Sichuan University had not yet admitted any Immediate casualties. Each hospital allocated medical personnel based on its actual patient load, ensuring an efficient distribution of resources. The number of occupied medical personnel is represented by the red bar in Fig. 13. It is important to note that these medical personnel remain occupied until the treatment of the current Immediate casualties is completed, after which they are reassigned to subsequent rescue missions to maintain continuous emergency response efforts.

During the rescue process for Delayed casualties, the dynamic dispatch of medical personnel and the changes in the number of Delayed

casualties at each disaster area are illustrated in Figs. 14 and 15. Similar to the ambulance dispatch process, before time period $t = 2$, medical personnel were still en route to their assigned disaster areas, and no Delayed casualties had yet received treatment. Once medical personnel arrived at the disaster areas, the number of Delayed casualties gradually decreased. For instance, Fig. 15 shows that all Delayed casualties in Yingjing County were treated by time period $t = 3$. Consequently, at the beginning of time period $t = 4$, the medical personnel stationed in Yingjing County began returning to the hospital, as reflected in the yellow bar of Fig. 13 after period $t = 5$. Furthermore, Fig. 14 indicates that by time period $t = 4$, no additional medical personnel were dispatched to Yingjing County from any hospital. Observations from Figs. 14 and 15 further reveal that after time period $t = 6$, all disaster areas—except for Lushan County—had been cleared of Delayed

Table 10
Model-DES-Actual Response Comparison.

Metric	Optimization Model	DES Results	Actual Ya'an Response
Ambulances Deployed	1,050	1,050	>4,000
Medical Personnel Utilized	3,983	3,983	9,889
Total System Cost (¥M)	4.13	4.15	N/A
Survivors Treated	7,088	7,012	~7,100 (estimated)

Table 11
Critical threshold values of medical personnel allocation in hospital settings.

Hospital	Number of medical personnel
Ya'an People's Hospital	778
The 4th People's Hospital of Ya'an City	897
People's Hospital of Mingshan District Ya'an	906
West China Hospital of Sichuan University	785

casualties, eliminating the need for further rescue efforts in those disaster areas.

Furthermore, comparisons with real-world cases indicate that, for the treatment of the same number of casualties, actual rescue operations required over 4,000 ambulance deployments and the involvement of 9,889 medical personnel. In contrast, the medical rescue strategy proposed in this study achieved the same rescue objectives with the optimized dispatch of only 1,050 ambulances and 3,983 medical personnel. The results presented in Figs. 11–15 demonstrate that all casualties in the disaster areas could be treated within 10 h. This highlights the high efficiency of the proposed emergency medical resource allocation strategy in terms of response time, resource utilization, and task coordination. The findings underscore its potential advantages in real-world disaster response scenarios and provide valuable insights for emergency management decision-making.

Validation via DES confirmed the operational feasibility of the NSGA-II-derived strategy. Resource utilization was reduced by 60–74 % compared to real-world responses (Table 10), while treatment completion times aligned with model predictions (max deviation: 4.2 %). Minor cost variations (DES: ¥4.15 M vs. Model: ¥4.13 M) stemmed from travel time uncertainties quantified in Section 6.4.3.

6.4. Sensitivity analysis

6.4.1. Impact of medical personnel allocation thresholds on hospital overcrowding penalties

In emergency medical service operations, the availability of medical personnel is a critical factor that directly influences hospital capacity and the number of casualties that can be effectively treated. Sufficient medical personnel can significantly enhance the efficiency of emergency medical response. However, for a given number of casualties, determining the minimum required medical personnel to prevent hospital overload and congestion is of substantial practical importance. This section conducts sensitivity analysis experiments to examine the critical threshold of medical personnel and its impact on the effectiveness of emergency medical rescue services. All experiments are based on the earthquake scenario in Ya'an, Sichuan, as described in Section 6.1.

A series of numerical experiments were conducted to determine the critical number of medical personnel required for the four hospitals in the case study, as presented in Table 11. These values represent the minimum staffing levels necessary to maintain the effective operation of the emergency medical service network. If the number of medical personnel falls below these thresholds, the dynamic medical rescue network fails to function efficiently. When casualty demand reaches its peak, hospitals operating at their critical staffing levels face a high risk of overcrowding, as illustrated in Fig. 16.

Fig. 17 depicts the trend of hospital overcrowding penalty values for the four hospitals as the number of medical personnel increases. Given the similar trends observed across the hospitals, Ya'an People's Hospital is selected for detailed analysis. When the number of medical personnel is set at the critical level of 778, the hospital overcrowding penalty at Ya'an People's Hospital reaches 76,515. However, with an increase in medical personnel to 779, this penalty drops sharply to one-fourth of its previous value. This finding highlights the significant impact of medical personnel availability on hospital overcrowding. However, as staffing levels continue to rise, the marginal effect of reducing overload risk gradually diminishes.

Insufficient staffing is unable to meet the treatment demands of a large number of casualties, potentially leading to disorganized emergency medical services and hospital congestion. Conversely, excessive staffing may result in resource inefficiency and reduced individual service effectiveness. Considering these factors, the optimal number of

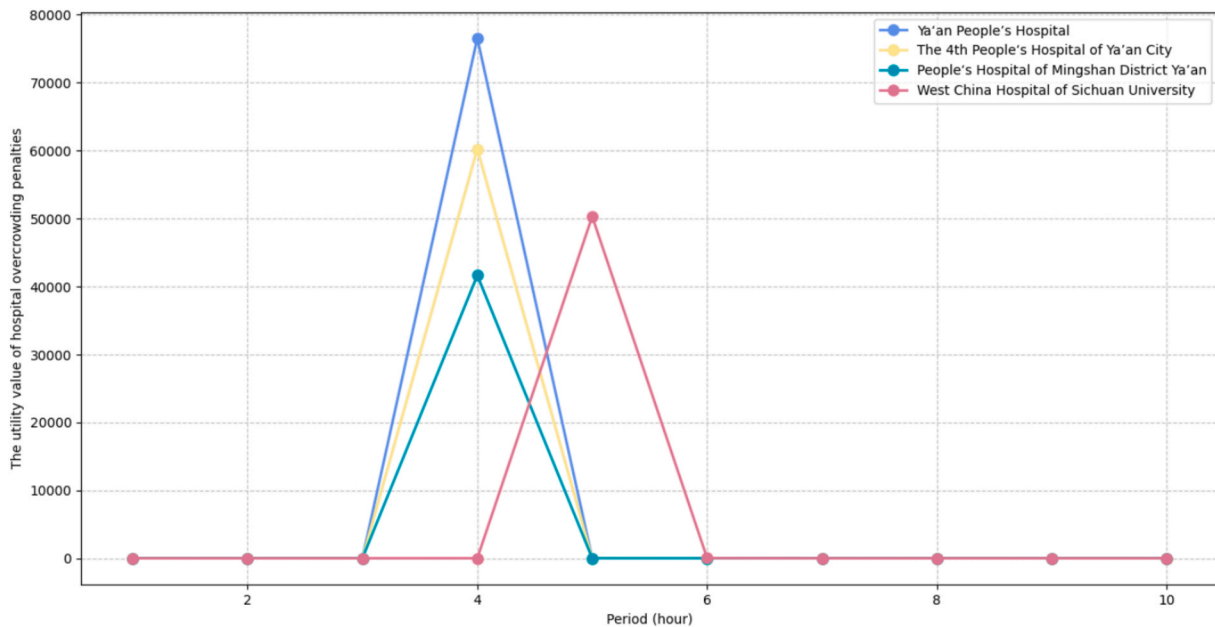


Fig. 16. Impact of medical personnel allocation thresholds on hospital overcrowding penalties.

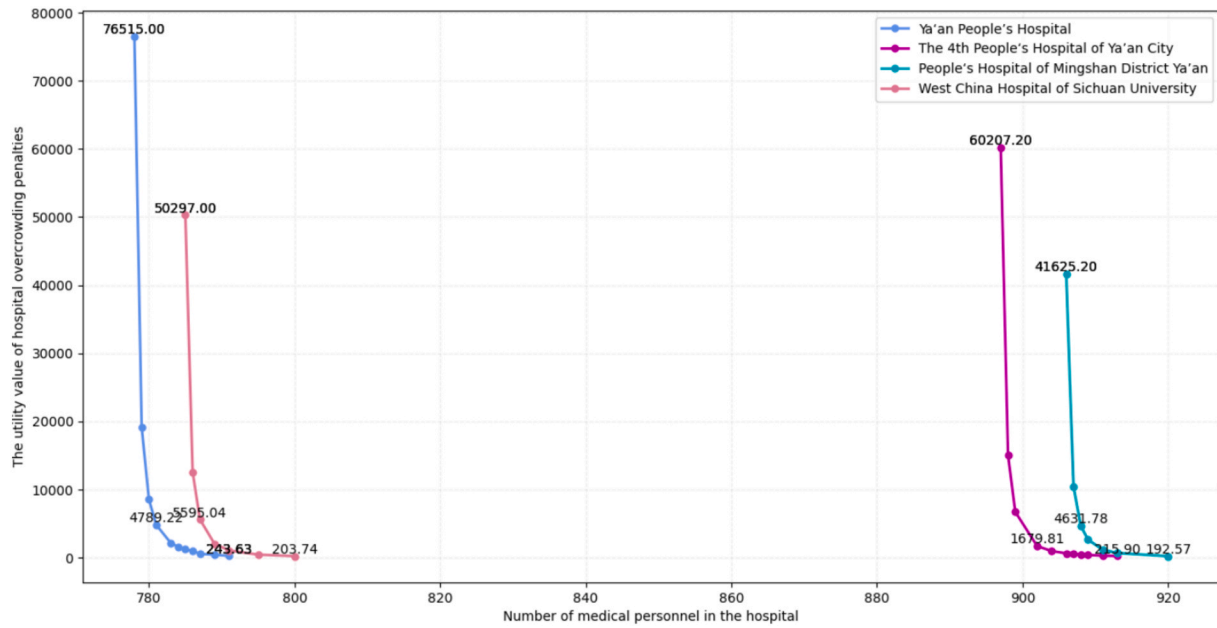


Fig. 17. Variations in overcrowding penalty utility metrics across four hospitals.

Table 12

Impact of medical personnel shortages on system performance.

Shortage level	Hospital congestion penalty (RMB)	%Change	Expected survivors	Survivor reduction
0 % (Baseline)	1,202.93	—	7088	—
10 %	3,580.44	+197.6 %	6792	4.2 %
20 %	6,210.75	+416.5 %	6455	8.9 %
30 %	11,836.20	+884.2 %	6208	12.4 %

medical personnel for each hospital in this study is estimated as follows:

- **Ya'an People's Hospital:** Approximately 787 personnel, reducing the overcrowding penalty to 1,202.93.
- **The 4th People's Hospital of Ya'an City:** Approximately 903 personnel, reducing the overcrowding penalty to 1,236.16.
- **People's Hospital of Mingshan District Ya'an:** Approximately 911 personnel, reducing the overcrowding penalty to 1,163.64.
- **West China Hospital of Sichuan University:** Approximately 791 personnel, reducing the overcrowding penalty to 1,033.62.

These staffing configurations effectively meet medical rescue demands while maximizing the utility of available medical personnel.

To systematically evaluate the impact of hospital overcrowding penalties on emergency system rescue costs, this study conducted a comparative analysis of cost structures across varying medical personnel allocation scenarios. The findings revealed a substantial disparity in cost distribution: at the critical staffing levels {778, 897, 906, 785}, overcrowding penalties accounted for 5.1 % of total system operational costs, whereas a more appropriate staffing configuration {787, 903, 911, 791} reduced this proportion to 0.1 %. This pronounced difference can be attributed to the inefficient resource distribution under constrained medical staffing conditions, which compromises the timely delivery of critical care at both hospitals and disaster areas. The resultant service bottlenecks significantly inflate overcrowding-related costs within the overall system rescue cost framework. Conversely, strategic optimization of medical personnel allocation not only mitigates overcrowding penalties but also enables more efficient resource deployment for core rescue operations, thereby establishing an optimized cost-control mechanism for emergency response systems.

We further tested stepwise personnel shortages (10 %, 20 %, 30 %

Table 13

Impact of homogeneous ambulance capacity (Q) on key performance metrics (Scenario 3).

Capacity (Q)	Ambulance operating cost (RMB)	Total expected number of survivors	Feasibility
1	42,260	7,088.34	Feasible
2	30,660	7,286.47	Feasible
3	—	—	Infeasible

below optimal thresholds) in Table 12. Hospital congestion penalties increased exponentially. Concomitantly, survivor counts decreased by 12.4 % under 30 % shortages due to delayed treatments. This confirms that personnel shortages below critical thresholds create compounding system failures.

6.4.2. Impact of homogeneous ambulance capacity on key performance metrics

To validate the assumption of ambulance homogeneity and investigate its impact on the model, we conducted a sensitivity analysis by adjusting the key parameter Q (ambulance capacity). While the model assumes a fixed capacity for all ambulances (homogeneous fleet), we evaluated system performance under different resource configurations by varying Q. Specifically, we tested three scenarios: Q = 1, Q = 2, and Q = 3. Table 13 summarizes the computational results.

From Table 13, we can find that increasing capacity from Q = 1 to Q = 2 yielded significant improvements:

- (1) Total operating costs decreased substantially by approximately 27.4 % (from ¥42,260 to ¥30,660). This stems directly from the reduced number of trips required to transport the same volume of Immediate casualties.

Table 14

Sensitivity analysis of travel time variability impacts.

Scenario	Ambulance Cost (RMB)	Total expected number of survivors	Ambulance Cost Change	Survivor Change	Personnel threshold stability
Baseline	42,260	7088	--	--	
High Delayed (+30 %)	54,938	6862	+30 %	-3.2 %	$\pm 2\%$
Low Delayed (-30 %)	29,582	7212	-30 %	+1.7 %	$\pm 2\%$

(2) The expected number of survivors increased significantly by approximately 2.8 % (from 7088.34 to 7286.47). This improvement occurs because faster batch transportation ($Q = 2$) shortens the overall treatment time for casualties while ensuring hospitals operate within capacity, allowing for fuller utilization of available medical personnel within hospitals.

Notably, when setting $Q = 3$, the optimization model became infeasible under the given constraints on medical personnel. The solver failed to find a feasible solution, explicitly triggering the hospital overcrowding penalty mechanism. This result stems directly from the model's integrated dynamic resource coordination mechanism: while larger ambulances transport more casualties per trip, they also cause surges in hospital admission demand. At $Q = 3$, the influx of casualties per ambulance trip exceeded the current available medical personnel at hospitals (y_i^j), violating constraints related to medical personnel allocation and hospital congestion management (e.g., Equations (5), (12), (22), (24)). These constraints prevent treatment delays and hospital overload.

This infeasibility highlights a crucial managerial insight: effective emergency medical response requires coordinated scaling of both transportation resources (ambulance capacity/fleet size) and treatment resources (medical personnel/hospital capacity). Increasing ambulance capacity alone, without proportionally enhancing hospital treatment capacity, can create system bottlenecks or even cause failure. Our model effectively captures this interdependence of critical resources.

These results demonstrate the model's sensitivity to Q . Specifically, $Q = 2$ balances transportation efficiency (cost reduction) with manageable hospital load under the medical personnel constraints, achieving a significantly higher expected survival rate compared to $Q = 1$. These quantifiable trade-offs provide clear evidence for emergency planners to coordinate ambulance procurement/sizing decisions with hospital treatment capacity expansions.

6.4.3. Sensitivity analysis on travel time variability

To evaluate model robustness under travel time variability, we introduced $\pm 30\%$ stochastic perturbations to π_{ij} following Zhang et al. (2017). Three scenarios were simulated:

- (1) Baseline: Deterministic travel time derived from Baidu Maps API;
- (2) High Delay: $\pi_{ij} \times 1.3$ (simulating severe road damage);
- (3) Low Delay: $\pi_{ij} \times 0.7$ (optimistic conditions).

Key outcomes are summarized in Table 14:

- (1) Ambulance operating costs increased by $\leq 30\%$ under high-delay conditions;
- (2) Expected survivors decreased by 3.2 % in worst-case delays due to prolonged transportation reducing survival probabilities;
- (3) Resource allocation stability: Optimal medical personnel thresholds (Section 6.4.1) maintained $\pm 2\%$ variation tolerance.

These results demonstrate acceptable resilience to travel time fluctuations. While delays moderately increase operational costs and marginally reduce survivor counts, the core resource allocation strategy maintains robustness. Explicit stochastic routing formulations are reserved for future work.

Table 15

Impact of treatment time extensions on survival rates.

Treatment time	Total expected number of survivors	Change
1 h (Baseline)	7088	--
2 h	6502	-8.3 %
3 h	5879	-17.1 %

6.4.4. Sensitivity analysis on treatment time delays

To evaluate the impact of medical resource shortages, we extended the fixed treatment time per casualty from baseline (1 h) to 2 and 3 h. This simulates scenarios where resource constraints prolong treatment durations. As shown in Table 15, extending to 2 h reduces expected survivors by 8.3 %, while a 3-hour treatment time reduces survivors by 17.1 %. This underscores that medical resource adequacy—not just allocation efficiency—is critical for survival outcomes.

7. Conclusions

Facing practical challenges such as post-disaster medical resource shortages, hospital overcrowding, and fairness in treatment, this study constructs a multi-objective nonlinear dynamic optimization model aimed at minimizing overall system rescue costs and maximizing the expected number of survivors. The model is solved using the NSGA-II algorithm. Numerical experiments based on the real case of the Ya'an earthquake in Sichuan Province validate the feasibility of the model and the efficiency of the NSGA-II algorithm, with further comparisons to the PLS algorithm demonstrating the superior performance of NSGA-II.

The results indicate that the rational planning of dynamic medical resource allocation is crucial for improving emergency medical service efficiency. The Pareto frontier obtained through NSGA-II provides a set of optimal decision-making solutions under different trade-offs between economic cost and rescue effectiveness, enabling decision-makers to select appropriate allocation strategies of medical personnel based on specific rescue needs. Furthermore, experimental analysis reveals a critical threshold for medical personnel allocation. Falling below this threshold may reduce hospital operational capacity and compromise treatment efficiency, while excessive allocation may lead to resource waste and negatively impact medical services in non-affected areas. These findings offer valuable insights for formulating more targeted emergency medical resource allocation strategies.

Future research should address several critical dimensions to enhance real-world applicability:

- (1) Dynamic Casualty Evolution: Integrate predictive models for time-varying casualty influx (e.g., AI-based injury trajectory forecasting) to replace static casualty assumptions, enabling adaptive resource reallocation during prolonged disasters.
- (2) Multi-modal Transportation Coordination & Vehicle Heterogeneity: Optimize joint scheduling of ambulances, helicopters, and drones for rapid victim extraction in terrain-fragmented zones, accounting for route damage and fuel constraints. Specifically, incorporate vehicle heterogeneity (e.g., differences in capacity, speed, medical equipment levels, and cost-effectiveness among ambulance types) into the dispatch and routing decisions.
- (3) Mobile Medical Facility Deployment: Model temporary field hospitals and mobile clinics as flexible resource buffers to

Table 16
Key Findings and Actionable Policy Recommendations.

Key Findings	Actionable Recommendations	Quantitative Evidence Source
Critical medical personnel thresholds exist	Pre-define and enforce minimum medical staff thresholds for participating hospitals. Allocate resources to ensure these thresholds are met before disasters occur. Implement real-time monitoring to prevent staffing levels from falling below critical values during response.	Section 6.4.1, Table 11, Table 12, Fig. 17
Ambulance capacity must align with hospital capability	Coordinate ambulance fleet design (capacity Q) with hospital surge capacity. Avoid deploying high-capacity ambulances ($Q \geq 3$) without proportional increases in medical staff. Establish real-time dispatch protocols linking ambulance routing to hospital availability.	Section 6.4.2, Table 13
Static resource plans are insufficient	Replace static allocation plans with dynamic, multi-period scheduling systems. Invest in decision-support tools (e.g., NSGA-II-integrated platforms) enabling continuous re-optimization based on evolving casualty distributions and resource status.	Section 6.3 (DES validation vs Static), Section 6.2.2 (Dynamic model performance)
Trade-offs between cost and survival are quantifiable	Use Pareto-optimal solution sets (e.g., Fig. 7) to explicitly evaluate cost-survival trade-offs during budget planning. Prioritize funding to minimize hospital congestion penalties and untreated casualty costs, which significantly impact system efficiency.	Section 6.2.2, Fig. 8, Table 9
Triage-driven resource matching improves outcomes	Mandate standardized field triage protocols (e.g., START) to direct Immediate/Delayed casualties to hospital/on-site care. Pre-define resource types (e.g., Group I/II personnel) and dispatch rules aligned with injury severity.	Section 6.3 (Overall efficiency gains), Section 3.1/Fig. 1 (Conceptual validation)
Model-informed strategies enhance operational efficiency	Integrate optimization models (like ours) into national/regional disaster response platforms. Build databases of historical MCIs (casualty patterns, resource constraints) for model calibration and scenario training.	Section 6.3, Table 10

alleviate hospital overload, especially in remote areas with limited healthcare infrastructure.

- (4) Human Behavior Integration: Incorporate crowd psychology (e.g., panic-driven patient self-evacuation) and volunteer workforce management into triage protocols to mitigate operational chaos observed in real incidents (e.g., traffic bottlenecks near hospitals during the 2013 Ya'an response).
- (5) We will integrate Weibull-distributed treatment times (parameterized via trauma datasets from Zhang et al., 2022) and develop a decomposition-based heuristic (e.g., scenario reduction + rolling-horizon NSGA-II) to manage computational complexity.

These extensions will transform our model into a decision-support toolkit that addresses spatial, temporal, and behavioral complexities inherent in disaster ecosystems.

8. Implications for policy and practice

The findings of this study provide concrete, actionable insights for policymakers and emergency management agencies seeking to optimize medical resource allocation during mass-casualty disasters. Based on our multi-objective optimization model, NSGA-II solution framework, and empirical validation using the 2013 Ya'an earthquake case, we summarize the following key recommendations in Table 16.

The study's framework yields concrete operational guidelines across disaster management phases. Pre-disaster preparations should incorporate the derived hospital-specific personnel thresholds (e.g., Table 11) into emergency readiness protocols, complemented by strategically deployed mobile command units equipped with real-time optimization capabilities. During disaster response, the model demonstrates that hourly recalibration of ambulance routes and medical staff allocation, guided by continuously updated Pareto frontiers (e.g., Fig. 8), achieves superior outcomes in balancing survival rates against resource costs. Post-event analyses must systematically compare actual resource deployment with model recommendations to iteratively improve threshold parameters, while the validated optimization framework should be integrated into WHO emergency response standards. This phased implementation approach directly addresses the critical triad of challenges identified in the study: dynamic resource allocation under uncertainty, preventable mortality minimization, and health system overload prevention during mass casualty incidents.

CRedit authorship contribution statement

Qing Liu: Writing – original draft, Software, Formal analysis, Data curation, Conceptualization. **Jixiang Dong:** Validation, Software, Data curation. **Na Cui:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Zimin Liang:** Validation, Software. **Jiabin Luo:** Writing – review & editing, Conceptualization.

Funding

This research was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 71874069) and General Program of the Natural Science Foundation of Shandong Province (NSFSP) (Grant No. ZR2024MG033).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Prof. Miqing Li of University of Birmingham, United Kingdom, for technical support. The authors also extend their sincere gratitude to the anonymous reviewers for their insightful comments and constructive suggestions, which have greatly helped in improving the quality of this paper.

Appendix A

Inputs for the case study of 2013 Ya'an earthquake (Sichuan, China)

This section shows how the inputs are obtained for the case study when considering the 2013 Ya'an earthquake (Sichuan, China).

Casualty estimation in disaster zones

Accurate casualty assessment in disaster-affected regions presents significant methodological challenges due to the dynamic nature of emergency situations and limitations in real-time data collection. To establish a reliable baseline for our analysis, this study utilizes official casualty statistics published by the People's Government of Sichuan Province (<https://www.sc.gov.cn/>). These validated epidemiological data, which reflect the most authoritative assessment available, are comprehensively presented in Table A.1.

Table A.1
Number of casualties in disaster areas.

Disaster area	Immediate casualties	Delayed casualties
Lushan County	193	3788
Baoxing County	205	1300
Yucheng District	231	1144
Tianquan County	259	1230
Mingshan District	224	647
Yingjing County	248	395

Geographic Distribution of Disaster Areas and Healthcare Facilities

The geographic coordinates of affected areas and medical facilities were determined through the following methodology: First, we identified key response hospitals from official rescue reports published by the National Health Commission of the People's Republic of China (<https://www.nhc.gov.cn/>) following the Ya'an earthquake. From these participating institutions, we selected four representative hospitals for our analysis. Additionally, we included six severely affected disaster zones with high casualty rates.

All geographic coordinates were precisely obtained using Baidu Maps API, with the complete dataset presented in Table A.2. This spatial data provides the foundation for our subsequent analysis of emergency response logistics and medical resource allocation.

Table A.2
Geographical coordinates of disaster areas and hospitals.

Hospital	Longitude	Latitude
Ya'an People's Hospital	103.061785	30.008764
The 4th People's Hospital of Ya'an City	103.02472	29.996436
People's Hospital of Mingshan District Ya'an	103.110175	30.092292
West China Hospital of Sichuan University	104.067284	30.64723
Disaster area	Longitude	Latitude
Lushan County	103.025094	30.260316
Baoxing County	102.832206	30.25992
Yucheng District	103.034418	30.010896
Tianquan County	102.891121	30.113984
Mingshan District	103.290617	30.155611
Yingjing County	102.847325	29.844524

Number of casualty classes treated by medical personnel

In this study, we assume the following treatment ratios:

- (1) Immediate casualties: One medical staff member per patient (1:1 ratio)
- (2) Delayed casualties: One medical staff member can treat four patients (1:4 ratio)

Due to the dynamic nature of disaster response, accurately estimating the exact number of medical personnel in this case study is challenging. This uncertainty arises because the medical teams included both local hospital staff and external support personnel dispatched from other healthcare facilities.

To address this limitation, we derived a reasonable approximation of available medical staff based on official statistics from the National Health Commission of the People's Republic of China (<https://www.nhc.gov.cn/>). The estimated distribution of medical personnel across hospitals is presented in Table A.3.

Table A.3
Number of medical personnel in hospitals.

Hospital	Number of medical personnel
Ya'an People's Hospital	1100
The 4th People's Hospital of Ya'an City	971
People's Hospital of Mingshan District Ya'an	962
West China Hospital of Sichuan University	950

Ambulance deployment across hospitals

The number of ambulances available at each hospital was obtained from official records published by the People's Government of Sichuan Province (<https://www.sc.gov.cn/>). These data, which reflect the emergency medical transport capacity of healthcare facilities in the region, are presented in Table A.4.

Table A.4
Number of ambulances in each hospital.

Hospital	Number of ambulances
Ya'an People's Hospital	316
The 4th People's Hospital of Ya'an City	258
People's Hospital of Mingshan District Ya'an	225
West China Hospital of Sichuan University	251

Data availability

The datasets related to the 2013 Ya'an earthquake are available from the following sources: National Health Commission of China: <http://www.nhc.gov.cn/>; People's Government of Sichuan Province: <https://www.sc.gov.cn/>; Additional datasets concerning the emergency medical response during the 2013 Lushan County earthquake in Ya'an City can be accessed at: Sichuan Provincial Health Commission: https://wsjkw.sc.gov.cn/scwsjkw/wstylwsjjyqk/ztwzlmgl_3.shtml.

References

- Aboud, A., Rokbani, N., Fdhila, R., Qahtani, A. M., Almutiri, O., Dhahri, H., Hussain, A., & Alimi, A. M. (2022). DPb-MOPSO: A dynamic pareto bi-level multi-objective particle swarm optimization algorithm, *Applied soft computing*, Vol. 129109622.
- Andersson, T., & Värbrand, P. (2007). Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society*, 58(2), 195–201.
- Bélanger, V., Lanzarone, E., Nicoletta, V., Ruiz, A., & Soriano, P. (2020). A recursive simulation-optimization framework for the ambulance location and dispatching problem. *European Journal of Operational Research*, 286(2), 713–725.
- Cao, C., Li, C., Yang, Q., Liu, Y., & Qu, T. (2018). A novel multi-objective programming model of relief distribution for sustainable disaster supply chain in large-scale natural disasters, *Journal of Cleaner Production*, Vol. 1741422-1435.
- Chang, K., Chen, T., Yang, F., & Chang, T. (2023). Simulation optimization for stochastic casualty collection point location and resource allocation problem in a mass casualty incident. *European Journal of Operational Research*, 309(3), 1237–1262.
- Chen, P., Lin, Y., & Peng, N. (2016). A two-stage method to determine the allocation and scheduling of medical staff in uncertain environments, *Computers & Industrial Engineering*, Vol. 99174-188.
- Chou, C., Chiang, W., & Chen, A. Y. (2022). Emergency medical response in mass casualty incidents considering the traffic congestions in proximity on-site and hospital delays, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 158102591.
- Dean, M. D., & Nair, S. K. (2014). Mass-casualty triage: Distribution of victims to multiple hospitals using the SAVE model. *European Journal of Operational Research*, 238(1), 363–373.
- Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197.
- Defraeye, M., & Van Nieuwenhuysse, I. (2016). Staffing and scheduling under nonstationary demand for service: A literature review, *Omega*, Vol. 584-25.
- Emanuel, K. (2017). Will Global Warming Make Hurricane forecasting more Difficult? *Bulletin of the American Meteorological Society*, 98(3), 495–501.
- Frykberg, E. R. (2005). Triage: Principles and practice. *Scandinavian journal of surgery*, 94(4), 272–278.
- Geary, W. L., Buchan, A., Allen, T., Attard, D., Bruce, M. J., Collins, L., Ecker, T. E., Fairman, T. A., Hollings, T., Loeffler, E., Muscatello, A., Parkes, D., Thomson, J., White, M., & Kelly, E. (2022). Responding to the biodiversity impacts of a megafire: A case study from south-eastern Australia's Black Summer. *Diversity and Distributions*, 28(3), 463–478.
- Gu, J., Zhou, Y., Das, A., Moon, I., & Lee, G. M. (2018). Medical relief shelter location problem with patient severity under a limited relief budget, *Computers & Industrial Engineering*, Vol. 125720-728.
- Haddow, G. D., Bullock, J. A., & Coppola, D. P. (2025). Chapter 9 - Emergency Management and the Terrorist Threat, in HADDOW, G. D., BULLOCK, J. A. & COPPOLA, D. P. (Ed.), *Introduction to Emergency Management (Eighth Edition)*. Butterworth-Heinemann, pp. 555-626.
- Hemici, M., Zouache, D., Brahmi, B., Got, A., & Drias, H. (2023). "A decomposition-based multiobjective evolutionary algorithm using Simulated Annealing for the ambulance dispatching and relocation problem during COVID-19", *Applied Soft Computing*, Vol. 141110282.
- Jacobson, E. U., Argon, N. T., & Ziya, S. (2012). Priority assignment in emergency response. *Operations Research*, 60(4), 813–832.
- Jotshi, A., Gong, Q., & Batta, R. (2009). Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. *Socio-Economic Planning Sciences*, 43(1), 1–24.
- Kendra, J. M., & Wachtendorf, T. (2003). Elements of resilience after the world trade center disaster: Reconstituting New York City's Emergency Operations Centre. *Disasters*, 27(1), 37–53.
- KIM, K. & OH, B. (2023). Prehospital triage in emergency medical services system: A scoping review, *International Emergency Nursing*, Vol. 69101293.
- Lettieri, E., Masella, C., & Radaelli, G. (2009). Disaster management: Findings from a systematic review, *Disaster Prevention and Management*, Vol. 18117-136.
- Lin, Y., Lin, C., & Lin, C. (2023). "Post-earthquake Patient diversion strategies for improving regional emergency healthcare accessibility, *International Journal of Disaster Risk Reduction*, Vol. 96103985.
- Liu, Y., Cui, N., & Zhang, J. (2019). Integrated temporary facility location and casualty allocation planning for post-disaster humanitarian medical service. *Transportation Research Part E Logistics and Transportation Review*, 1281.
- Maenhout, B., & Vanhoucke, M. (2013). An integrated nurse staffing and scheduling analysis for longer-term nursing staff allocation problems. *Omega*, 41(2), 485–499.
- Mahalakshmi, S., Ragunthar, T., Veena, N., Sumukha, S., & Deshulkarni, P. R. (2022). Adaptive ambulance monitoring system using IOT. *Sensors Measurement*, Article 24100555.
- Mills, A. F., Argon, N. T., & Ziya, S. (2013). Resource-based patient prioritization in mass-casualty incidents. *Manufacturing & Service Operations Management*, 15(3), 361–377.
- Mills, A. F., Argon, N. T., & Ziya, S. (2018). Dynamic distribution of patients to medical facilities in the aftermath of a disaster. *Operations Research*, 66(3), 716–732.
- Mizumoto, T., Sun, W., Yasumoto, K., & Ito, M. (2011). Transportation scheduling method for patients in MCI using Electronic Triage Tag. In *International Conference on eHealth, Telemedicine, and Social Medicine* (p. pp.).
- Mohri, S. S. & Haghshenas, H. (2021). An ambulance location problem for covering inherently rare and random road crashes, *Computers & Industrial Engineering*, Vol. 151106937.
- Najafi, M., Eshghi, K., & de Leeuw, S. (2014). A dynamic dispatching and routing model to plan/ re-plan logistics activities in response to an earthquake. *OR Spectrum*, 36(2), 323–356.
- Norio, O., Ye, T., Kajitani, Y., Shi, P., & Tatano, H. (2011). The 2011 eastern Japan great earthquake disaster: Overview and comments. *International Journal of Disaster Risk Science*, 2(1), 34–42.
- Organization, W. H. (2021). *Mass casualty management systems : Strategies and guidelines for building health sector capacity*. Geneva: WHO Press.
- Rais, A., & Viana, A. (2011). Operations research in healthcare: A survey. *International Transactions on Operational Research*, 18(1), 1–31.
- Raiter, Y., Farfel, A., Lehavi, O., Goren, O. B., Shamiss, A., Priel, Z., Koren, I., Davidson, B., Schwartz, D., Goldberg, A., & Bar-Dayan, Y. (2008). Mass casualty incident management, triage, injury distribution of casualties and rate of arrival of casualties at the hospitals: Lessons from a suicide bomber attack in downtown Tel Aviv. *Emergency Medicine Journal*, 25(4), 225–229.
- Raj, A. S., & Chinnadurai, M. (2021). An queueing model with improved delay sensitive medical packet transmission scheduling system in e-health networks. *Journal of Ambient Intelligence and Humanized Computing*, 12(3), 3493–3504.
- Rathore, N., Jain, P. K. & Parida, M. (2022). A sustainable model for emergency medical services in developing countries: a novel approach using partial outsourcing and machine learning, *Risk Manag Health Policy*, Vol. 15193-218.
- Rodríguez-Espíndola, O., Albores, P., & Brewster, C. (2018). Dynamic formulation for humanitarian response operations incorporating multiple organisations, *International Journal of Production Economics*, Vol. 20483-98.
- Sacco, W. J., Navin, D. M., Fiedler, K. E., Waddell, R. N., Long, W. B., & Buckman, R. J. (2005). Precise formulation and evidence-based application of resource-constrained triage. *Academic Emergency Medicine*, 12(8), 759–770.
- Sacco, W. J., Navin, D. M., Waddell, R. N., Fiedler, K. E., Long, W. B., & Buckman, R. J. (2007). A new resource-constrained triage method applied to victims of penetrating injury. *The Journal of Trauma*, 63(2), 316–325.
- Setiawan, E., Liu, J., & French, A. (2019). Resource location for relief distribution and victim evacuation after a sudden-onset disaster. *IISE Transactions*, 51(8), 830–846.
- Shetab-Boushehri, S., Rajabi, P., & Mahmoudi, R. (2022). "Modeling location-allocation of emergency medical service stations and ambulance routing problems considering the variability of events and recurrent traffic congestion: A real case study", *Healthcare Analytics*, Vol. 2100048.

- Sheu, J. (2007). An emergency logistics distribution approach for quick response to urgent relief demand in disasters. *Transportation Research Part E: Logistics and Transportation Review*, 43(6), 687–709.
- Sheu, J. & Pan, C. (2014). “A method for designing centralized emergency supply network to respond to large-scale natural disasters”, *Transportation Research Part B: Methodological*, Vol. 67284–305.
- Shiri, D., Akbari, V. & Tozan, H. (2023), “Online optimisation for ambulance routing in disaster response with partial or no information on victim conditions”, *Computers & Operations Research*, Vol. 159106314.
- Sun, H., Wang, Y., Zhang, J. & Cao, W. (2021), A robust optimization model for location-transportation problem of disaster casualties with triage and uncertainty, *Expert Systems with Applications*, Vol. 175114867.
- Sun, W., Zhu, C. & Li, H. (2023), Evolutionary game analysis of emergency rescuer dispatching under bounded rationality, *International Journal of Disaster Risk Reduction*, Vol. 96103927.
- Sung, I., & Lee, T. (2016). Optimal allocation of emergency medical resources in a mass casualty incident: Patient prioritization by column generation. *European Journal of Operational Research*, 252(2), 623–634.
- Sutherland, M. & Chakraborty, R. K. (2023), An optimal ambulance routing model using simulation based on patient medical severity, *Healthcare Analytics*, Vol. 4100256.
- Talarico, L., Meisel, F. & Sörensen, K. (2015), Ambulance routing for disaster response with patient groups, *Computers & Operations Research*, Vol. 56120–133.
- Toro-Díaz, H., Mayorga, M. E., Chanta, S., & McLay, L. A. (2013). Joint location and dispatching decisions for emergency medical services. *Computers & Industrial Engineering*, 64(4), 917–928.
- Vahdani, B., Veysmoradi, D., Mousavi, S. M. & Amiri, M. (2022), Planning for relief distribution, victim evacuation, redistricting and service sharing under uncertainty, *Socio-Economic Planning Sciences*, Vol. 80101158.
- van den Bergh, J., Belien, J., de Bruecker, P., Demeulemeester, E., & de Boeck, L. (2013). Personnel scheduling: A literature review. *European Journal of Operational Research*, 226(3), 367–385.
- Wan, M., Ye, C. & Peng, D. (2023), Multi-period dynamic multi-objective emergency material distribution model under uncertain demand, *Engineering Applications of Artificial Intelligence*, Vol. 117105530.
- Wang, W., Yang, K., Yang, L. & Gao, Z. (2023), Distributionally robust chance-constrained programming for multi-period emergency resource allocation and vehicle routing in disaster response operations, *Omega*, Vol. 120102915.
- Wang, Y., Colledanchise, M., Marzinotto, A., & Ögren, P. (2014). A distributed convergent solution to the ambulance positioning problem on a streetmap graph. *IFAC Proceedings Volumes*, 47(3), 9190–9196.
- Wex, F., Schryen, G., Feuerriegel, S., & Neumann, D. (2014). Emergency response in natural disaster management: Allocation and scheduling of rescue units. *European Journal of Operational Research*, 235(3), 697–708.
- WHO (2023) A Global Health Strategy for 2025–2028 - advancing equity and resilience in a turbulent world: fourteenth General Programme of Work.
- Yang, M., Liu, Y. & Yang, G. (2021), “Multi-period dynamic distributionally robust pre-positioning of emergency supplies under demand uncertainty”, *Applied Mathematical Modelling*, Vol. 891433–1458.
- Yazdani, M. & Haghani, M. (2023), Logistics of patient evacuation in response to disease Outbreaks: Critical considerations for transportation planning, *Transportation Research Interdisciplinary Perspectives*, Vol. 22100975.
- Yi, W., & Özdamar, L. (2007). A dynamic logistics coordination model for evacuation and support in disaster response activities. *European Journal of Operational Research*, 179 (3), 1177–1193.
- Yoon, S. & Albert, L. A. (2020), A dynamic ambulance routing model with multiple response, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 133101807.
- Zhang, H., Zhang, X., Comfort, L. & Chen, M. (2016), “The emergence of an adaptive response network: The April 20, 2013 Lushan, China Earthquake”, *Safety Science*, Vol. 9014–23.
- Zhang, L., Liu, X., Li, Y., Liu, Y., Liu, Z., Lin, J., Shen, J., Tang, X., Zhang, Y., & Liang, W. (2012). Emergency medical rescue efforts after a major earthquake: Lessons from the 2008 Wenchuan earthquake. *The Lancet*, 379(9818), 853–861.
- Zhang, M. & Zheng, W. (2024), “Building disaster resilience through emergency Plan updates: A case study of Ya’an, China”, *International Journal of Disaster Risk Reduction*, Vol. 100104173.
- Zhang, M., Zhang, Y., Jiao, Z. & Wang, J. (2023), Improving relief operations via optimizing shelter location with uncertain covariates, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 176103181.
- Zhang, Z., Qin, H., Wang, K., He, H. & Liu, T. (2017), Manpower allocation and vehicle routing problem in non-emergency ambulance transfer service, *Transportation Research Part E: Logistics and Transportation Review*, Vol. 10645–59.
- Zheng, W. & Doerr, B. (2023), Mathematical runtime analysis for the non-dominated sorting genetic algorithm II (NSGA-II), *Artificial Intelligence*, Vol. 325104016.
- Zhu, L., Gong, Y., Xu, Y., & Gu, J. (2019). Emergency relief routing models for injured victims considering equity and priority. *Ann. Oper. Res.*, 283(1), 1573–1606.
- Zoraster, R. M., Chidester, C., & Koenig, W. (2007). Field triage and patient maldistribution in a mass-casualty incident. *Prehospital and Disaster Medicine*, 22(3), 224–229.