

A Grid-Based Reinforcement Learning Approach for Ambulance Dispatching

Carlos Cubillas¹ , Juan M. Alberola¹ , Victor Sanchez-Anguix² , Jaume Jordán¹  *

¹ Valencian Research Institute for Artificial Intelligence, Valencia (Spain)

² Instituto Tecnológico de Informática, Grupo de Sistemas de Optimización Aplicada Universitat Politècnica de València, Camí de Vera s/n, 46022, Valencia (Spain)

* Corresponding author: chcub@doctor.upv.es (C. Cubillas), jalberola@dsic.upv.es (J. M. Alberola), vsanchez@eio.upv.es (V. Sanchez-Anguix), jjordan@dsic.upv.es (J. Jordán).

Received 11 May 2025 | Accepted 16 May 2026 | Published 28 August 2026



ABSTRACT

Ambulance dispatching for critical life-threatening emergencies (type-1) poses significant challenges in rural and resource-constrained settings, where ensuring both rapid response and long-term coverage is critical. Existing methods often overlook ambulance type constraints and focus narrowly on minimizing immediate response time. In this paper, we propose a dispatching strategy called GARD (Grid-aware Reinforcement Dispatching), which partitions the service area into grids categorized by historical emergency intensity. A reinforcement learning agent uses this structure to select ambulances, balancing proximity with anticipated system-wide impact. Experiments based on real-world Emergency Medical Services (EMS) data demonstrate that GARD improves performance metrics for high-priority emergencies while preserving better future coverage.

KEYWORDS

Ambulance Dispatching, Critical Life Risk Emergencies, Emergency Medical Services, Grid-Based Strategy, Reinforcement Learning

DOI: 10.9781/ijimai.2026.2247
eLocator: e2247

I. INTRODUCTION

EMERGENCY Medical Services (EMS) are dedicated to delivering urgent medical assistance to individuals facing life-threatening situations. Their responsibilities include assessing emergency calls, coordinating ambulance dispatch, and ensuring timely and safe patient transport [1]. Thus, the effectiveness of EMS systems depends on their ability to reach patients quickly, provide appropriate on-site medical care, and transfer them to healthcare facilities when necessary.

Effective and optimized ambulance dispatch management for emergency calls is vital for EMS systems. The primary goal of these systems is to save lives and reduce the severity of medical incidents. The performance of an EMS is typically evaluated based on response time, which is the period between receiving a call and the arrival of the first ambulance at the scene. This is a key indicator of service efficiency, quality assessment, and the overall protection of patients' lives and health [2]–[5]. Hence, reducing this response time is directly related to higher survival probabilities for patients [6], [7].

In this context, making quick and accurate decisions about which ambulance to send to the emergency is vital. Several models have been proposed to improve the efficiency of ambulance dispatch, proving to be effective [8], [9]. These models often assume the possibility of reallocating ambulances after completing a service or redirecting

them while en route [8], [10]. However, in many operational contexts (such as in Spain and other European countries) such flexibility is not feasible. In this sense, ambulances are typically assigned to fixed stations that serve as their base of operations and resource supply, and must return there after each intervention [11], [12].

Another major limitation of many existing works is the assumption of a homogeneous ambulance fleet, ignoring the operational differences between Basic Life Support and Advanced Life Support ambulances. This simplification does not reflect the clinical reality, where the medical needs of the patient must match the capabilities of the dispatched vehicle. Furthermore, most dispatching strategies assume urban environments with high ambulance density and short travel distances, making them less effective or directly inapplicable in rural or sparsely populated areas, where the resources are limited and travel times are significantly longer. Therefore, although dispatching the nearest available ambulance is a common strategy and performs reasonably well in general [13], it may not be sufficient for these scenarios.

In [14] we proposed an ambulance dispatching algorithm for type-1 (critical life-threatening) emergencies in geographically dispersed regions with static ambulance deployments, aiming to overcome some of these limitations. This algorithm considered a grid-partitioned territory based on accident rates. In this work, we extend that proposal by introducing GARD (Grid-Aware Reinforcement Dispatching), a

Please cite this article as:

C. Cubillas, J. M. Alberola, V. Sanchez-Anguix, J. Jordán. A Grid-Based Reinforcement Learning Approach for Ambulance Dispatching, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 10, no. 1, e2247, 2026, <http://doi.org/10.9781/ijimai.2026.2247>

novel dispatching strategy that combines a geographic grid-based categorization of the service area with a reinforcement learning agent. In this sense, each grid is assigned a risk level based on historical emergency data, guiding dispatch decisions to balance immediate response efficiency with long-term coverage stability. GARD represents a methodological advancement from a static risk-based rule system toward an adaptive reinforcement learning framework. While the previous work divided the territory into risk-ranked grids and applied a fixed heuristic to dispatch ambulances from lower-risk zones when possible, the current proposal uses tabular Q-learning to learn optimal dispatching policies through experience. This allows GARD to dynamically balance coverage and responsiveness based on historical patterns and spatial heterogeneity. Our results demonstrate that the strategy not only improves response times but also shows flexibility in optimizing alternative performance metrics and adapting to varying call dispersion patterns and resource availability scenarios.

The rest of the paper is organized as follows. Section II reviews relevant works on emergency dispatching, with particular emphasis on approaches that incorporate reinforcement learning. Section III formalizes the emergency dispatch problem. Section IV describes the grid-based heuristic dispatching strategy. Section V presents the architecture of the reinforcement learning agent. Then, Section VI reports the experimental evaluation of our proposal. Finally, Section VII discusses concluding remarks and outlines directions for future research.

II. RELATED WORK

The problem of ambulance dispatching has long been recognized as crucial for optimizing emergency response outcomes. Traditionally, the nearest ambulance strategy has been the standard approach in real-world applications. However, in the literature, several strategies aim to enhance response times through innovative techniques. Following we comment some of the most relevant works in this area. In [6], the Bertsekas auction method is used to minimize total travel time by iteratively matching patients to ambulances through competitive bidding. In [8], approximate dynamic programming (ADP) is used to generate solutions. A lower bound on the expected response time of any reallocation policy is developed to evaluate the quality of the solutions produced by the ADP. The proposal classifies emergencies into two priority levels, with type-1 representing high-priority emergencies. Upon receiving a high-priority emergency call, any available ambulance can be dispatched immediately, or it may wait for an ambulance that will soon be available. The work of Roa et al. [15] uses the preparedness index to evaluate and optimize the availability of ambulances in an emergency system. This index is based on the probability that at least one ambulance is available (within a time limit) when an emergency demand occurs. The arrival of emergencies is modeled as a Poisson process, and the service of ambulances as an exponential service process. The objective is to maximize this preparedness index to ensure that the ambulance system can attend to emergencies with a high probability.

In [16], a constrained Markov Decision Process (MDP) is proposed to identify optimal policies based on priority lists, formulated as a Mixed Integer Programming model. The study assumes that service rates are the same for all types of patients, which significantly reduces the state space for solving larger problems. The system allows any available ambulance to be dispatched to a received call (including the most critical), rather than immediately sending the nearest ambulance. This includes the option of not dispatching an ambulance immediately if another ambulance is expected to become available soon. In [17], a system is proposed where emergency and non-emergency ambulance fleets can share resources, allowing for dynamic modification of each fleet's size based on the system's state. The capacity-sharing policy is

designed to be flexible and adaptable to various EMS policies, but its practical implementation may present challenges due to operational differences among organizations. When an emergency call arrives, the nearest available ambulance is dispatched. A backup coverage model is used to ensure that each demand point is covered by at least two ambulances, mitigating vehicle unavailability. Other authors have approached their analysis of this issue from the perspective of smart-city and IoT systems for spatial and temporal data analysis [18].

Although the existing literature offers several strategies to reduce response times, most of them present key limitations that hinder their applicability in real-world emergency medical systems. First, many proposals are designed for general emergencies and do not focus specifically on critical life risk, type-1 (critical life-threatening) emergencies, which require the dispatch of medically-equipped ambulances. These proposals often either disregard ambulance types or default to sending the nearest available unit, which may be inadequate in scenarios with strict equipment requirements. This limitation becomes especially problematic in rural or low-density areas, where such resources are scarce and spatial demand heterogeneity is more pronounced [19].

Second, most strategies rely on assumptions that may not hold in realistic operational settings, such as the possibility of ambulance reallocation. This includes rerouting ambulances to new stations after completing a service, or even interrupting ongoing assignments to attend critical life risk calls. While these assumptions can yield performance gains (as reported in works such as [7]) real systems often enforce rigid return-to-origin policies, restricting the dispatch options available and reducing the flexibility assumed by many academic models.

Third, most dispatch algorithms are designed for dense urban settings, where the nearest-ambulance strategy is often sufficient due to short travel distances. In contrast, rural contexts introduce challenges where this approach may lead to coverage gaps or long-term inefficiencies. In such cases, incorporating spatial patterns of risk (such as using grid-based demand categorizations) can enable smarter decisions, especially in balancing short-term response efficiency with long-term system preparedness as demonstrated in our previous work [14].

Finally, dispatching strategies typically prioritize minimizing average response time. However, this objective does not always align with system-level requirements. For instance, fairness in ambulance workload distribution, consistent service quality across rural and urban areas, or the guarantee that all responses occur within a legally or medically defined time threshold can be equally or more important. In this context, minimizing the maximum or variance of response times may be more desirable than reducing the average. These limitations suggest the need for more adaptable solutions, able to optimize multiple objectives and account for spatial heterogeneity.

A. Reinforcement Learning for Ambulance Dispatching

While classical ambulance dispatching strategies have introduced valuable optimizations, the use of reinforcement learning EMS has gained significant attention in recent years, driven by the increasing need to make fast, adaptive decisions in uncertain and dynamic environments. Within this context, several works have explored reinforcement learning for ambulance dispatching.

The work of Liu et al. [20] introduces a multi-agent framework in which ambulances operate as independent learning agents. The problem formulation discretizes the city into a spatial grid and divides time into 5-minute intervals, where the agents must decide how many ambulances should be pre-positioned in each cell during each time window. While this anticipatory strategy may reduce overall waiting times in the simulated environment, it assumes a level of pre-emptive mobility and flexibility that may not align with real-world

TABLE I. SUMMARY OF RELATED REINFORCEMENT LEARNING APPROACHES FOR AMBULANCE DISPATCHING

Reference	Main Contribution	Limitations	Type-1 Considered
[20]	Multi-agent RL with spatial-temporal grid pre-positioning.	Assumes pre-emptive mobility and reallocation.	No
[21]	Deep Q-Network with ambulance reallocation.	Limited area and assumes reallocation	No
[22]	MDP with post-decision states and convergence proofs.	Purely theoretical, no real data and considers reallocation.	No
[23]	Multi-objective MDP including vertical/horizontal equity.	Includes severity but not vehicle type; allows reallocation.	Partial
[24]	RL for supply dispatch with traffic-aware attention networks.	Focuses on medical supplies and not ambulance units.	No
[25]	Dynamic redeployment with DRL for mobile ambulances.	Not specially focused on dispatching.	Partial
[26]	DRL-based decision support for ambulance dispatch.	Evaluated only in simulation.	Partial
[27]	RL + LLMs + multi-objective	Focuses on LLMs integration.	No
[28]	Offline RL with real EMS data.	Not specially focused on dispatching.	No
[29]	Hybrid offline-online RL with clinical priority tiers.	Not specially focused on dispatching.	Partial

EMS protocols, where ambulances typically remain at base until an emergency call is received. In addition, this approach does not differentiate between emergency types or ambulance capabilities.

The work of Elfahim et al. [21] addresses dispatching problem by proposing a Deep Double Q-Network trained on a simulated representation of the city of Casablanca. Unlike other approaches that focus on anticipatory deployment, this work considers ambulance reallocation between calls. Resource constraints are explicitly modeled, and the agent is trained using dense-layer architectures with ϵ -greedy exploration. However, the simulation is limited to a small spatial area, and there is no evaluation of long-term system behavior or robustness under operational variability. Additionally, the model does not distinguish between types of emergencies, and its performance is not compared against baseline methods.

A more theoretical approach is presented by Hua and Zaman [22], who formulated EMS dispatching as a MDP with post-decision states to reduce computational complexity. This is one of the first works that theoretically guarantee convergence to optimal policies in this context. However, the model is purely numerical, with no real-world data or consideration of uncertain dynamics. In addition, the model considers ambulance reallocation and it does not differentiate between types of ambulances.

In contrast, Luan et al. [23] explicitly address equity in EMS dispatching by formulating a multi-objective MDP. The model presented in this paper integrates both operational efficiency and equity criteria, including vertical equity (based on patient severity) and horizontal equity (related to territorial distribution), with penalties for triage misclassification. The approach is evaluated in a simulated environment that includes both urban and rural areas, allowing assessment under varying population densities. While the model considers the severity of the emergency call to determine whether ambulance dispatch is needed and to prioritize allocation, it does not differentiate between ambulance types. Furthermore, the framework allows ambulance reallocation after service completion.

Other studies explore complementary directions. Wei et al. [24] focus on the dispatching of medical supplies rather than ambulances, using an attention-based deep reinforcement learning architecture. While innovative in terms of neural design and the use of traffic patterns, the model excludes key EMS components such as mobile units and medical prioritization. Ji et al. [25] propose a dynamic redeployment framework for mobile ambulances using deep reinforcement learning, focusing on repositioning strategies rather than real-time dispatch. A related exploratory study by Moen [26] examines deep reinforcement learning-based ambulance dispatch in a simplified setting, highlighting early attempts to address this problem with learning-based approaches. Chen et al. [27] go further by integrating reinforcement learning with Large Language Models and multi-objective optimization, enabling dispatch decisions that incorporate natural language input

from victims. Although this represents a major conceptual leap, the approach depends on large datasets, is computationally demanding, and lacks testing under realistic emergency stress conditions.

Closer to deployment, Lamarca et al. [28] apply offline reinforcement learning techniques to real EMS data from Switzerland, bypassing the need for environment interaction during training. While promising for safe deployment, the model inherits human biases and suboptimal behavior from historical data. Finally, Garriga et al. [29] propose a hybrid offline-online reinforcement learning strategy with explicit clinical prioritization. This approach accounts for dispatch penalties and medical risk tiers, yet it omits dynamic variables like weather or real-time traffic and is not validated in operational systems.

As observed, a variety of approaches have applied reinforcement learning to address ambulance dispatching problems. Most of these proposals validate improvements in response time compared to baseline strategies such as nearest-ambulance dispatch. However, a common assumption in these studies is the possibility of ambulance reallocation, whereby ambulances can be reassigned to different stations after completing a service. Nevertheless, in some European countries such as Spain, operational constraints (particularly for advanced life support ambulances that respond to type-1 emergencies) require that ambulances return to their originally assigned stations after completing a service. Moreover, the majority of these works have been designed and evaluated in urban environments, which may show different dynamics when compared to larger, more heterogeneous territories that include varying population densities, greater station dispersion, and longer travel distances. Finally, although some proposals consider the call urgency in the dispatching process (such as [23]), they do not explicitly model the distinction between different types of ambulances. This is particularly relevant in the case of critical life risk emergencies, which require ambulances equipped with specific medical resources and qualified personnel.

To better understand the positioning of our proposal within the current state of the art, Table I summarizes the main reinforcement learning-based approaches for ambulance dispatching. The table compares each work in terms of its key contributions, limitations, and whether it explicitly models the operational constraints associated with type-1 emergencies and the corresponding ambulance requirements.

According to this, we propose a dispatching strategy based on reinforcement learning tailored to the operational constraints of real EMS systems, with a particular focus on critical life risk, type-1 emergencies. Unlike many existing approaches, we do not consider ambulance reallocation and we evaluate our method in a heterogeneous geographical setting that includes both urban and rural areas. Our system divides the territory into geographic grids and assigns each grid a risk category based on historical emergency demand. This grid-based structure enables the dispatching policy to capture spatial heterogeneity and to balance short-term response efficiency with long-

term system preparedness. By integrating these practical dimensions into the decision-making process, we aim to bridge the gap between theoretical performance and real-world applicability.

III. PROBLEM DESCRIPTION

To represent the emergency dispatch problem, we define a set of ambulances $A = \{a_1, a_2, \dots, a_n\}$ where each ambulance a_i is positioned at a fixed station b_i belonging to a set of stations $B = \{b_1, b_2, \dots, b_k\}$. Each ambulance is required to depart from and return to its assigned station after attending an emergency. We also define a set of emergency calls $C = \{c_1, c_2, \dots, c_m\}$ where each call c_j is assigned to an ambulance a_i . The assignment can be modeled using a binary decision variable x_{ij} where:

$$x_{ij} = \begin{cases} 1 & \text{if ambulance } a_i \text{ is assigned to emergency call } c_j \\ 0 & \text{otherwise} \end{cases}$$

Our objective is to optimize the assignment of ambulances to emergency calls in order to improve the system's responsiveness. A general formulation can be expressed as shown in Equation (1):

$$\min \sum_{i=1}^n \sum_{j=1}^m x_{ij} f(b_i, c_j) \quad (1)$$

where $f(b_i, c_j)$ represents a cost function that evaluates the assignment of ambulance a_i to call c_j . One of the classic goals is to minimize the travel time $t(b_i, c_j)$ between the station b_i where ambulance a_i is located, and the emergency site c_j . However, this function can be defined in several ways depending on the performance objective (e.g. the travel time $t(b_i, c_j)$ from the ambulance's station to the emergency location), subject to the constraints:

1. Each emergency call is assigned to exactly one ambulance as shown in Equation (2):

$$\sum_{i=1}^n x_{ij} = 1 \quad \forall j \in \{1, 2, \dots, m\} \quad (2)$$

2. Each ambulance can be assigned to at most one emergency call at a time, as shown in Equation (3):

$$\sum_{j=1}^m x_{ij} \leq 1 \quad \forall i \in \{1, 2, \dots, n\} \quad (3)$$

The requirement that ambulances return to their respective stations is implicitly satisfied by the model's assumption of fixed station assignments and the exclusion of dynamic relocation.

IV. HEURISTIC GRID-BASED DISPATCHING

The GARD strategy introduced in this work divides the service area into geographic grids and assigns each grid a category based on its historical emergency calls. This categorization supports dispatch decisions that aim to balance short-term responsiveness with long-term system coverage.

The territory is defined by its geographical boundaries (north, south, east, and west) and partitioned into a grid G , where each cell is specified by ranges of latitude and longitude. Each grid cell $g \in G$ is classified into one of three risk categories using a function $T: G \rightarrow \mathcal{T}$, where $\mathcal{T} = \{1, 2, 3\}$ represents high, medium, and low historical accident rates, respectively.

Each station b_i is associated with a grid cell $g_k \in G$, and the category $T(g_k)$ of that grid informs dispatch decisions. When multiple ambulances are available within an acceptable maximum response time t_{max} , the dispatching algorithm prioritizes those located in lower-category grids (i.e., areas with fewer historical emergencies), thus preserving high-risk zones for future calls.

Fig. 1 illustrates the overall architecture of the dispatching system. First, the territory is partitioned according to G , and training data are used to compute the category $T(g)$ for each grid cell. These categories represent the historical emergency call patterns in different regions (e.g., high, medium, or low risk areas), and they remain fixed during the simulation phase.

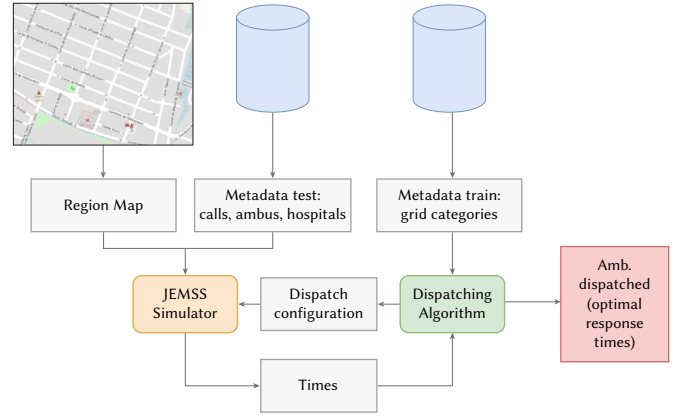


Fig. 1. Dispatching system architecture.

In the simulation phase, test data are introduced, which include information such as emergency calls, locations of ambulances, or hospital positions. The dispatching algorithm is then executed for each emergency call to assign the most appropriate ambulance and calculate the estimated response time. In our framework, the dispatching algorithm is implemented as a reinforcement learning agent that learns from simulated interactions how to assign ambulances more effectively.

Importantly, the system uses a precomputed heuristic to assess whether an ambulance can reach a call within a predefined maximum response time t_{max} . This heuristic helps simplify the simulation and improve computational efficiency by approximating real-world operational constraints, such as the possibility of an ambulance being too far to respond on time.

By using a simulator, the dynamic behavior of ambulance movement and resource allocation can be incorporated, which can be challenging to capture analytically. This provides a flexible framework for testing various dispatch strategies under different scenarios. For this purpose, we use JEMSS (Julia package for Emergency Medical Services Simulation) [30], which is a deterministic simulator designed specifically for healthcare emergencies. The JEMSS simulator allows for running various dispatch strategies under different scenarios, helping to evaluate the impact of different policies on overall system performance.

V. REINFORCEMENT LEARNING AGENT

Building on the grid-based framework, the reinforcement learning agent learns optimal dispatching policies through interaction with the simulated environment. This agent complements the heuristic by refining decisions based on long-term performance, using a tabular Q-learning approach with an ϵ -greedy exploration strategy.

As stated above, the environment is discretized into a set of grid cells G , and both emergency calls and ambulance stations are projected onto grid cells $g \in G$. A state $s \in S$ is defined as a tuple capturing the spatial configuration of the system, $s = (\mathbf{c}, \mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_N)$, where $\mathbf{c} \in \mathbb{Z}^2$ denotes the grid cell of the emergency call and $\mathbf{b}_i \in \mathbb{Z}^2$ the grid cell of the station associated with the i -th ambulance, for $i = 1, \dots, N$, with N being the total number of ambulances. This state representation encodes the spatial dynamics relevant for dispatch decisions, allowing the agent to account for the real-time positioning of emergency resources.

Due to the specific and multi-zone nature of the emergency dispatch problem, a custom reinforcement learning environment was implemented to ensure precise control over the system's dynamics, including specific reward functions such as efficiency, fairness, and critical coverage.

The action space $\mathcal{A}_s$ at a given state s is defined by the subset of currently dispatchable ambulances $\mathcal{D}_s \subseteq \{1, \dots, N\}$. The agent selects an action $a \in \mathcal{D}_s \cup \{0\}$, where $a = 0$ denotes a fallback option used when no ambulance is available, ensuring the system can still operate under such conditions. Once an ambulance is selected, the system computes the expected response time $t(s, a)$ and the agent receives a reward determined by the function shown in Equation (4):

$$R(s, a) = -100 \cdot \exp\left(\frac{t(s, a)}{T_{\max}}\right) \quad (4)$$

where $T_{\max}$ is a predefined threshold (e.g., 15 minutes). This reward function strongly penalizes long response times more severely, encouraging faster responses. The penalization constant of -100 was selected to emphasize the operational importance of minimizing response times, particularly in the critical low-range region (e.g., under 10 minutes), while ensuring numerical stability during training.

To ensure a fair comparison across experiments, the reward function remains fixed throughout, so any variation in performance can be attributed to differences in dispatch strategies. This design choice simplifies the learning dynamics while enabling a controlled comparison of behavioral policies.

The agent updates its Q-values using the classical Q-learning rule, as shown in Equation (5):

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[R(s, a) + \gamma \cdot \max_{a'} Q(s', a') - Q(s, a) \right] \quad (5)$$

where α is the learning rate, γ the discount factor, and s' the next observed state. The Q-table is lazily initialized: when a new state-action pair (s, a) is first encountered during simulation, its Q-value is initialized to zero. Optionally, to improve early-stage convergence, a domain-informed heuristic estimate based on the inverse of expected response time can be used instead of zero.

The policy π is ϵ -greedy, selecting the action with the highest Q-value with probability $1-\epsilon$, and a random valid action with probability ϵ . The exploration parameter ϵ decays exponentially over time according to Equation (6):

$$\epsilon_{t+1} = \max(\epsilon_{\min}, \epsilon_t \cdot \delta) \quad (6)$$

where $\epsilon_{\min}$ is a minimum exploration threshold and $\delta \in (0, 1)$ is the decay rate. All dispatch decisions are logged, including states, actions, rewards, response times, and Q-values, supporting offline analysis and diagnostics. The design of the state and action spaces ensures that the agent is sensitive to the real-time spatial distribution of resources and demand.

A. Assignment Function

The reinforcement learning agent relies on an assignment function that selects the most suitable ambulance for each emergency call. This function considers both the estimated response time and the risk level of the area where each ambulance is located. In the design of this agent, different assignment functions were tested, but the final strategy adopted in this work is a risk-based assignment function, which does not only takes into account the estimated response times but also factors in the risk level of the area. This ensures that ambulances are dispatched efficiently to manage both immediate and long-term system coverage. The response time $t(s, a)$ is computed using the Open Route Service routing engine¹, which estimates travel time over the

real road network between the current location of ambulance a and the location of the emergency call in state s . This assignment function aims to preserve coverage in high-risk zones without compromising response times.

This risk-based assignment function incorporates geographical risk factors according to the three risk categories defined in Section IV. Thus, the agent prioritizes ambulances from lower-risk areas, thus preserving availability in more critical regions. Within the same risk category, the ambulance with the shortest response time is selected. If an ambulance from a higher-risk zone is within a tolerance τ of the optimal response time, it may be chosen instead. This flexibility is controlled by a threshold defined in Equation (7):

$$t_{\text{threshold}} = t_{\text{opt}} \cdot (1 + \tau) \quad (7)$$

where t_{opt} is the minimum available response time. This time threshold allows the agent to choose an ambulance from a lower-risk area if its response time is within a tolerance of the optimal response time, giving it flexibility to handle variability in ambulance availability while maintaining efficiency. Thus, this assignment strategy reflects a more nuanced decision-making process, considering both temporal and spatial factors, which can enhance overall system efficiency in high-demand environments.

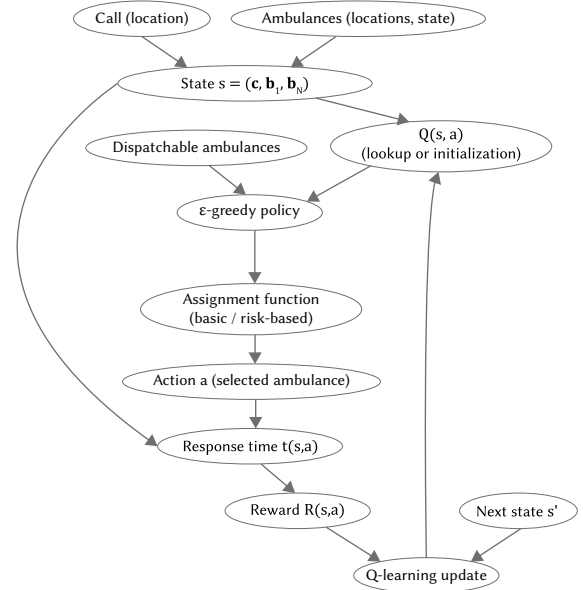


Fig. 2. Q-learning agent architecture for ambulance dispatch decision-making.

Fig. 2 illustrates the internal architecture of the Q-learning agent. The system receives the location of an emergency call and the spatial configuration of the ambulance fleet as inputs. These are combined into a discrete state s , from which the Q-values $Q(s, a)$ are retrieved or initialized. A subset of currently dispatchable ambulances $\mathcal{D}_s$ is first computed, restricting the set of actions available to the agent. This subset is then passed to the ϵ -greedy policy, which produces a candidate action. The action selected by the ϵ -greedy policy is passed to an assignment function, which may override or confirm the selected ambulance based on risk-aware heuristics. The resulting dispatch action determines the response time $t(s, a)$, which generates the reward $R(s, a)$. The system then transitions to the next state s' , which is observed, and the Q-table is updated using the Bellman equation. This update reflects the learning process, where the Q-values are adjusted based on the observed reward and the maximum future expected reward from state s' . The training is performed offline using simulated emergency call sequences generated from historical patterns, while

¹ <https://openrouteservice.org/>

TABLE II. PERFORMANCE COMPARISON OF RESPONSE TIMES RELATIVE TO THE GARD STRATEGY

Metric	GARD	Nearest amb. (NA)	Preparedness
Mean duration	15m 36s $\pm$ 16s (IQR: 5m 24s)	16m 43s $\pm$ 23s (IQR: 5m 40s)	17m 37s $\pm$ 15s (IQR: 5m 48s)
Median duration	14m 49s $\pm$ 17s (IQR: 5m 24s)	15m 08s $\pm$ 23s (IQR: 5m 40s)	15m 30s $\pm$ 17s (IQR: 5m 48s)
Mean gain vs. GARD	—	6.62% $\pm$ 1.56%	11.43% $\pm$ 1.52%
Median gain vs. GARD	—	2.08% $\pm$ 1.33%	4.46% $\pm$ 1.49%

evaluation is carried out on separate test episodes not seen during training. Each episode comprises a unique series of emergency calls, ensuring that the learned dispatch policy generalizes well across different scenarios and does not overfit to specific data.

VI. EXPERIMENTS

In this section, we present a comprehensive set of experiments aimed at evaluating the effectiveness and flexibility of our GARD dispatching strategy. We compare its performance against the traditional nearest ambulance approach, which serves as a strong baseline in emergency response systems. The experiments aim to assess not only improvements in response time, but also the robustness of the strategy under varying operational conditions, including changes in resource availability, spatial distribution of calls, and reward function priorities.

A. Setup

The scenario used for the experiments focused on the area of the province of Valencia (Spain), covering the metropolitan area of Valencia City (with a total of 45 municipalities) and other municipalities distributed throughout the region, up to a total of 270 municipalities. The total area of the province is 10,700km², while the metropolitan area covers 2,200km² (20% of the total area). The area accounts for approximately 2,600,000 inhabitants², of which about 65% are located in the metropolitan area of Valencia City. This provides a spatially and demographically heterogeneous environment, representative of both urban and rural EMS challenges. For the simulation, we employed the actual station locations for the type-1 ambulances fleet. The road network was obtained from Open Route Service.

To train and evaluate our approach, we used a full year of emergency call data from 2019. We focused specifically on type-1 calls, which typically require urgent intervention and transportation. This dataset was split into two disjoint subsets: one used to define grid risk levels across the territory (high, medium, and low-risk grids), and the other used to generate synthetic instances of emergency calls. These synthetic calls preserve the spatial and temporal structure of the real data, enabling simulation of realistic and diverse demand conditions across different times and regions.

In our experiments, we compared our GARD dispatching strategy, which uses a grid decomposition of the territory and a trained agent to dynamically assign ambulances, with the traditional nearest ambulance strategy, where the closest available unit is always dispatched. In addition, in the first experiment, we included an adapted version of the preparedness index-based dispatching algorithm proposed in [15], serving as a complementary baseline. Each configuration was tested using 10 independent experiments, each using different training and test datasets. For these evaluations, a fleet of 15 ambulances was employed in the baseline nearest-ambulance configuration, matching the resource level used by GARD to ensure a fair comparison. We set the τ value to 0.15, which allows for a 15% tolerance over the minimum response time when selecting ambulances from higher-risk zones, enabling the system to preserve coverage in critical areas without significantly compromising efficiency. It is important to note that, for

each run, all strategies were evaluated on the same test data, ensuring a fair and direct comparison.

B. Results

In an initial exploratory phase, we tested different values for key hyperparameters to optimize the performance of our dispatching strategy. The focus was not on benchmarking against other strategies, but rather on determining the most effective configuration for our method. After conducting multiple evaluations, we settled on the following set of parameters for the final experiments: $\alpha = 0.1$, $\gamma = 0.9$, $\epsilon = 0.7$, $\epsilon_{min} = 0.01$, and $\delta = 0.995$. These values were selected based on their ability to achieve a balance between exploration and exploitation, allowing the reinforcement learning agent to converge efficiently while adapting to different scenarios.

Next, to assess the effectiveness of our proposed strategy, we compared its performance against the nearest ambulance strategy. In this first experiment, we included a baseline adapted from the algorithm presented in [15], which uses the preparedness index to evaluate the availability of ambulances in the system. For our comparison, we excluded the ambulance reallocation component in order to isolate the impact of the dispatch policy. This performance was evaluated based on the time it took for the assigned ambulance to reach the emergency location. The results, including mean and median response times, along with relative improvements, are shown in Table II.

Although absolute durations depend on simulator-specific parameters (e.g., vehicle speed), all the strategies were executed under identical experimental conditions, ensuring that the reported percentage improvements are objective and comparable. Our strategy achieves a mean improvement of 6.62% and a median improvement of 2.08% over the nearest ambulance baseline and a mean improvement of 11.43% a median improvement of 4.46% over the Preparedness baseline. The lower median compared to the mean suggests that a majority of response times were shorter, while a smaller number of longer cases increased the average. This pattern indicates that the strategy not only enhances overall performance but also reduces the frequency and severity of extreme delays. As a result, it delivers more consistent and robust outcomes across a wide variety of scenarios. To further support this, we report the interquartile range (IQR) of response times. The GARD strategy exhibits a narrower IQR compared to both baselines, indicating reduced variability and fewer extreme cases. This reinforces the conclusion that GARD provides not only faster, but also more reliable response times. Since the nearest ambulance strategy outperforms the preparedness-based approach in terms of both average and median response times, we focus exclusively on comparisons with the nearest ambulance strategy in the subsequent experiments.

C. Flexibility Through Alternative Reward Functions

While our strategy demonstrated improvements in average response time, one of its key strengths lies in its flexibility to optimize alternative performance metrics. In particular, the reward function can be tailored to reflect specific operational objectives. A relevant example is the percentage of emergencies served within a predefined time threshold, which may be more operationally meaningful than merely reducing the overall average.

² Data from the 2021 census.

TABLE III. PERCENTAGE OF CALLS SERVED WITHIN DIFFERENT TIME THRESHOLDS: COMPARISON BETWEEN THE GARD STRATEGY AND THE NEAREST AMBULANCE (NA) STRATEGY

Threshold	GARD	NA	Abs. improvement (pp)	Rel. improvement (%)
10 min	24.19% $\pm$ 1.49%	22.92% $\pm$ 1.35%	1.27	5.54%
12 min	33.17% $\pm$ 1.86%	31.74% $\pm$ 1.58%	1.43	4.50%
15 min	50.92% $\pm$ 2.36%	49.06% $\pm$ 2.06%	1.86	3.79%
20 min	77.26% $\pm$ 1.57%	74.03% $\pm$ 1.36%	3.23	4.36%

To explore this capability, we trained our agent using a function that rewarded serving calls within a given time limit. We tested four different time thresholds: 10, 12, 15, and 20 minutes. Table III shows the percentage of calls served within each threshold by our strategy compared to the nearest ambulance strategy. As shown, our method outperforms the baseline, with improvements ranging from 1.27 to 3.23 percentage points depending on the threshold. These gains are particularly notable for longer thresholds, but remain meaningful even under stricter time constraints, indicating the strategy's capacity to adapt its behavior to different optimization targets. Standard deviations remained low in all scenarios, indicating stable performance across the repetitions. In addition to absolute differences, the relative improvements (ranging from 3.79% to 5.54%) offer further insight into the practical impact of the gains, especially under tighter time limits.

Fig. 3 complements the table by illustrating the distribution of the percentage of served calls. The plot confirms the reliability of our approach, showing consistently higher service rates across all thresholds. This supports the robustness of the method when optimized for operational constraints such as time limits.

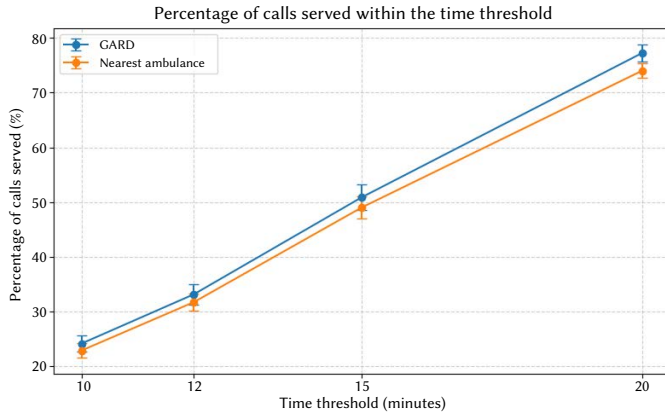


Fig. 3. Distribution of the percentage of calls served below the threshold.

D. Robustness Under Varying Emergency Distributions

In this experiment, we tested six different configurations of emergency call distribution between the metropolitan area and the surrounding province: 30-70, 40-60, 50-50, 60-40, 70-30, and 80-20 — where, for example, 80-20 indicates that 80% of the calls originate in the metropolitan zone and 20% in the province. For each configuration, we measured the percentage reduction in response time using our strategy, considering both average and median values, across three geographic scopes: general (all calls), metropolitan, and provincial.

Table IV summarizes the results. The values show the percentage reduction in response time achieved by our strategy with respect to the nearest ambulance strategy. Across all configurations and zones, our method outperforms the baseline. The most notable overall improvement is observed in the 40-60 configuration, which shows the highest general average reduction (7.78%), along with the best metropolitan gain (8.41%) and a strong improvement in the province (6.99%). This suggests that a mild preference for allocating more units

to the metropolitan area provides the best balance of efficiency across both regions.

TABLE IV. PERCENTAGE OF RESPONSE TIME SAVED BY THE GARD STRATEGY COMPARED TO THE NEAREST AMBULANCE, BROKEN DOWN BY GEOGRAPHIC ZONE AND CONFIGURATION

Metric	80-20	70-30	60-40	50-50	40-60	30-70
General						
Mean	5.94	5.29	6.44	5.66	7.78	6.53
Median	3.07	1.29	2.28	2.24	4.07	1.94
P90	6.42	3.85	7.20	4.13	7.87	5.10
Metro						
Mean	6.59	5.44	7.23	5.75	8.41	5.24
Median	3.51	1.81	2.23	1.63	3.79	1.52
P90	5.54	4.52	7.39	4.16	9.07	5.01
Province						
Mean	4.86	5.04	5.31	5.56	6.99	7.76
Median	2.83	2.02	2.24	3.16	3.59	2.85
P90	6.27	2.67	6.12	3.26	6.60	5.34

In general, extremely unbalanced deployments (such as 30-70 or 80-20) resulted in reduced performance in the under-resourced zone and did not significantly outperform more balanced configurations. For instance, the 30-70 setup shows in the lowest metropolitan median improvement (1.52%), while 80-20 reduced provincial performance (4.86%) in exchange for a modest metropolitan gain. The 50-50 configuration offered stable but suboptimal improvements across all zones, indicating that uniform deployment does not yield optimal system-wide performance.

Interestingly, the improvements were generally more pronounced in the average response time than in the median, highlighting the strategy's ability to mitigate extreme delays. This is particularly relevant in emergency response scenarios, where reducing outliers can have life-saving consequences.

This pattern is also evident in the P90 metric, where our strategy consistently improves the 90th percentile response time across all configurations and zones. The most significant P90 gains appear in intermediate configurations such as 40-60 (up to 9.07% in the metropolitan area), suggesting that GARD is especially effective in reducing late responses under moderately balanced deployments. These improvements in the upper tail of the distribution further reinforce the strategy's robustness under stress.

To further investigate how the system responds to varying demand distributions,

Fig. 4 presents the average response times in the provincial area. A clear trend is evident: although both approaches exhibit increasing response times as calls become more geographically dispersed, the relative improvement of the GARD strategy over the nearest ambulance strategy becomes more pronounced. This suggests that GARD scales more effectively in dispersed or decentralized scenarios, particularly by better handling emergencies in less densely populated regions.

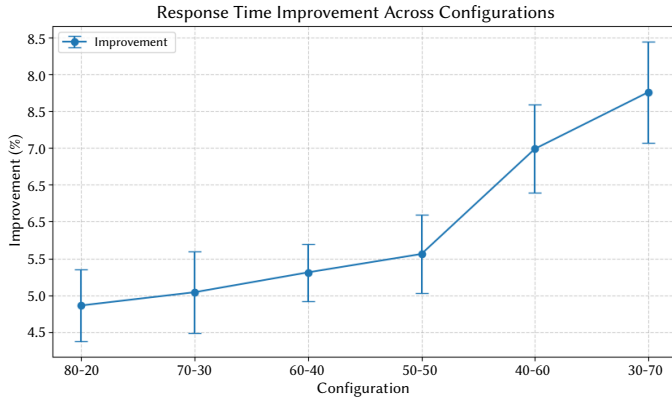


Fig. 4. Percentage of response time improvement for each configuration.

This behavior can be attributed to the grid-based design of our method, which facilitates more balanced territorial coverage and dispatching. In contrast, the nearest ambulance approach becomes less effective in rural or peripheral zones, where longer travel distances and fewer available units increase the likelihood of delays. These findings confirm the robustness of our strategy and its suitability for regions with high spatial variability in emergency demand.

E. Scalability Under Varying Resource Availability

To evaluate the effectiveness of the proposed dispatch strategy under different resource conditions, we performed a series of tests using varying numbers of available ambulances: 10, 12, 14, 16, 18, and 20 units. Table V presents the detailed numerical results of these tests, showing the relative improvements in both mean and median response times, along with the associated standard deviations. This provides a more precise comparison between configurations and highlights the variability of the strategy's performance across different levels of resource availability. Additionally, the table includes improvements in the 90th percentile (P90) of response times, which reflect the strategy's impact on the slowest 10% of cases. These results show consistent gains across all ambulance levels, indicating that the strategy not only improves average performance but also reduces delays in worst-case scenarios, which is critical in emergency contexts.

TABLE V. RELATIVE IMPROVEMENT (%) IN AVERAGE AND MEDIAN RESPONSE TIMES OF THE STRATEGY COMPARED TO THE CLOSEST-UNIT DISPATCH, FOR EACH NUMBER OF AMBULANCES

# Amb.	Mean improv. (%)	Median improv. (%)	P90 improv. (%)
10	2.53 ± 1.50	0.65 ± 2.33	2.02
12	3.15 ± 0.99	0.89 ± 1.22	6.22
14	5.45 ± 1.08	1.79 ± 0.97	4.44
16	6.73 ± 1.06	1.99 ± 1.38	6.27
18	6.88 ± 0.40	3.29 ± 1.28	5.61
20	7.23 ± 0.93	4.14 ± 1.50	6.00

Fig. 5 visually illustrates these improvements. The results suggest that with only 10 or 12 ambulances, the system may be operating under saturation, as reflected by limited or highly variable gains. However, as the number of units increases, the GARD strategy outperforms the baseline, with clear gains particularly from 14 units onward. This suggests that the benefits of the strategy become more pronounced under adequately dimensioned systems, where unit availability enables the strategy to make more meaningful trade-offs beyond proximity.

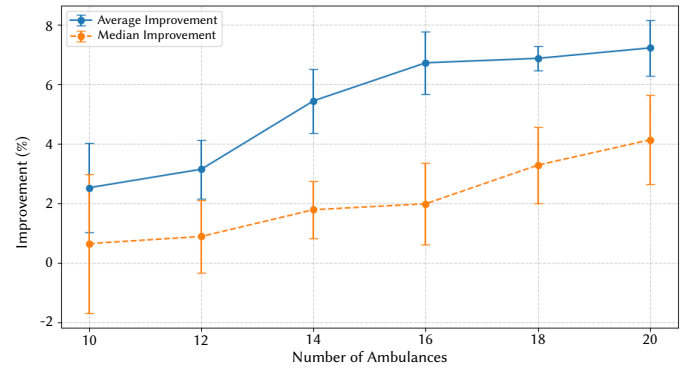


Fig. 5. Percentage improvements in mean and median response times.

Overall, the experimental results demonstrate that our reinforcement learning-based dispatch strategy outperforms the traditional closest-unit approach across a wide range of scenarios. Beyond reducing average response times, it proves to be flexible and robust under different demand distributions and resource constraints, and can be easily adapted to optimize alternative performance indicators, such as service level thresholds. These findings validate the versatility of our approach and its potential to support more informed and effective decision-making in emergency response systems.

VII. CONCLUSIONS

One of the most critical components of EMS operations is the dispatching of ambulances, particularly for high-priority (type-1) emergencies where delays can significantly affect survival outcomes. While numerous dispatching strategies have been proposed in the literature, many rely on assumptions that do not hold in real-world settings. These include the ability to reallocate ambulances dynamically between stations, the omission of ambulance type distinctions and an excessive focus primarily on urban areas with dense and predictable emergency patterns. In addition, most approaches prioritize minimizing average response time, often at the expense of long-term system coverage and equitable service distribution in geographically dispersed or resource-constrained environments.

To address these challenges, this work has introduced a novel dispatching framework that explicitly considers operational constraints typical of static EMS deployments, such as those found in Spain and other European systems. Our proposal, called GARD, integrates a spatial risk modeling component with a reinforcement learning agent that adapts dispatching decisions based on historical patterns and real-time availability. Specifically, the service area is divided into grid cells, each of which is assigned a risk category derived from historical emergency call data. These categories inform a dispatching heuristic that restricts feasible assignments and guides the learning process of the agent. The agent is implemented as a tabular Q-learning model, trained through simulation to balance immediate response effectiveness with the preservation of future system coverage. GARD maintains operational realism by adhering to real-world EMS constraints, while offering a principled approach to improve dispatch decisions in heterogeneous territories. This evolution enables more robust and adaptive decisions, especially under resource limitations, and distinguishes our work from existing models that rely on reallocation or assume uniform geographic risk.

Experimental results confirm the effectiveness of the proposed method. We compared GARD against the classic nearest ambulance strategy, showing that GARD reduced average response times and highlighted its capacity to mitigate extreme delays while improving general efficiency. Beyond average performance, further experiments

demonstrated the adaptability of our approach. First, we trained the agent with alternative reward functions targeting service-level objectives, such as the percentage of calls served within specific time thresholds. GARD consistently outperformed the nearest ambulance strategy across thresholds from 10 to 20 minutes. We also evaluated robustness under varying geographic distributions of emergencies (urban vs. rural) and found that GARD retained or even increased its advantage in dispersed scenarios, thanks to its grid-aware allocation. Finally, scalability was assessed by varying the number of ambulances. While the benefits were marginal in resource-scarce settings, GARD achieved increasingly significant improvements as more units became available, especially from 14 units onward. These results confirm that GARD is not only more effective on average but also more robust, adaptable, and scalable than traditional strategies, offering a promising tool for EMS systems facing diverse operational challenges.

As observed, a key advantage of the proposed reinforcement learning approach lies in its capacity to learn context-specific dispatching policies. Unlike traditional strategies (such as the nearest-ambulance rule), which apply a fixed and context-insensitive heuristic, our agent can be trained under different operational conditions and constraints, enabling it to adapt its behavior accordingly (scenarios characterized by high system saturation, low ambulance availability, or varying spatial distributions of demand). Similarly, separate models could be tailored to different geographic zones, such as densely populated metropolitan areas versus sparsely populated rural regions, each with their own response time dynamics and operational trade-offs. This modular and data-driven adaptability significantly enhances the strategic flexibility of the system, allowing it to respond more effectively to heterogeneous environments and evolving emergency care needs.

Although the results obtained are promising, there are relevant limitations that must be acknowledged. First, the model is trained and evaluated within a simulated environment using historical data, meaning its real-time performance may be affected by external factors not captured in the current setup—such as real-time traffic, special events, or abrupt changes in resource availability. Second, the grid structure and risk categorization rely on historical patterns that might not reflect dynamic or emerging behaviors in certain different regions. Finally, the integration of the system into actual EMS workflows requires careful coordination to avoid disrupting existing protocols.

As future work, several directions could enhance the current approach. First, extending the agent to support continuous state representations or more complex policy models (e.g., deep reinforcement learning) could improve scalability to larger urban areas. Second, incorporating dynamic ambulance repositioning policies (while maintaining realistic operational constraints) could further increase system efficiency. Lastly, deploying the framework with real EMS data and collaborating with medical authorities would enable evaluation in operational settings and facilitate the translation of these methods into practice.

CREDiT AUTHORSHIP CONTRIBUTION STATEMENT

Carlos Cubillas: Conceptualization; Software; Methodology; Data curation; Formal analysis; Investigation; Visualization; Validation; Writing - original draft.

Juan M. Alberola: Supervision; Conceptualization; Methodology; Project administration; Writing – review and editing.

Victor Sanchez-Anguix: Supervision; Conceptualization; Methodology; Project administration; Validation; Writing - review and editing.

Jaume Jordan: Resources; Validation; Writing - review and editing.

DATA STATEMENT

The dataset used in this work consists of confidential emergency service call data and cannot be publicly released due to privacy and institutional restrictions. Access is restricted to authorized research use. Methodological details are provided to ensure reproducibility of the results.

DECLARATION OF CONFLICTS OF INTEREST

No conflict of interest exists.

ACKNOWLEDGMENT

This work was partially supported by MINECO/FEDER RTI2018-095390-B-C31 project of the Spanish government, partially financed with FEDER funds under the project INNVA1/2024/91 of Agencia Valenciana de la Innovación, DIGITAL2022 CLOUDAI02/S8760000 from the European Commission, PID2021-123673OB-C31 and PID2021-124975OB-I00 funded by the Spanish Ministry of Science and Innovation, PDC2022-133161-C32 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”, and Cátedra Telefónica Smart Inteligencia Artificial.

REFERENCES

- [1] A. D. Hanchate, M. K. Paasche-Orlow, W. E. Baker, M.-Y. Lin, S. Banerjee, J. Feldman, “Association of race/ethnicity with emergency department destination of emergency medical services transport,” *JAMA Network Open*, vol. 2, no. 9, 2019, doi: <https://doi.org/10.1001/jamanetworkopen.2019.10816>
- [2] H. Andersson, T. A. Granberg, M. Christiansen, E. S. Aartun, H. Leknes, “Using optimization to provide decision support for strategic emergency medical service planning—three case studies,” *International Journal of medical informatics*, vol. 133, p. 103975, 2020, doi: <https://doi.org/10.1016/j.ijmedinf.2019.103975>
- [3] N. Pulsiri, R. Vatananan-Thesenvitz, T. Sirisamutr, P. Wachiradilok, “Save lives: a review of ambulance technologies in pre-hospital emergency medical services,” in *International Conference on Management of Engineering and Technology*, 2019, pp. 1–10, https://doi.org/10.1007/978-3-031-77738-7_15
- [4] C. Gallego, J. M. Alberola, V. Sánchez-Anguix, F. Villa, V. Julián, V. Botti, “A metaheuristic algorithm guided by simulation for optimizing the static allocation of emergency medical vehicles,” in *International Conference on Soft Computing Models in Industrial and Environmental Applications*, 2024, pp. 295–305, Springer, https://doi.org/10.1007/978-3-031-75013-7_28
- [5] V. Sánchez-Anguix, Y. Karpova, J. M. Alberola, F. Villa, V. Julián, V. Botti, “A local-search metaheuristic informed by simulation for optimizing response time in static ambulance location problems,” *Logic Journal of the IGPL*, In Press.
- [6] H. Billhardt, M. Lujak, V. Sánchez-Brunete, A. Fernández, S. Ossowski, “Dynamic coordination of ambulances for emergency medical assistance services,” *Knowledge-Based Systems*, vol. 70, pp. 268–280, 2014, doi: <https://doi.org/10.1016/j.knosys.2014.07.006>
- [7] L. Aboueljainane, Z. Jemai, E. Sahin, “Reducing ambulance response time using simulation: The case of val-de-marne department emergency medical service,” in *Proceedings of the 2012 Winter Simulation Conference (WSC)*, 2012, pp. 1–12, <https://doi.org/10.1109/WSC.2012.6465018>
- [8] A. Nasrollahzadeh, A. Khademi, M. Mayorga, “Real-time ambulance dispatching and relocation,” *Manuf. & Service Operations Management*, vol. 20, no. 3, pp. 467–480, 2018, doi: <https://doi.org/10.1287/msom.2017.0649>
- [9] C. J. Jagtenberg, S. Bhulai, R. D. van der Mei, “Dynamic ambulance dispatching: is the closest-idle policy always optimal?,” *Health care management science*, vol. 20, pp. 517–531, 2017, doi: <https://doi.org/10.1007/s10729-016-9368-0>
- [10] V. Bélanger, Y. Kergosien, A. Ruiz, P. Soriano, “An empirical comparison of relocation strategies in real-time ambulance fleet management,”

Computers & Industrial Engineering, vol. 94, pp. 216–229, 2016, doi: <https://doi.org/10.1016/j.cie.2016.01.023>

- [11] P. Berg, J. T. Essen, “Comparison of static ambulance location models,” *International Journal of Logistics Systems and Management*, vol. 32, no. 3–4, pp. 292–321, 2019, doi: <https://doi.org/10.1504/IJLSM.2019.098321>
- [12] J. J. Boutilier, T. C. Chan, “Ambulance emergency response optimization in developing countries,” *Operations Research*, vol. 68, no. 5, pp. 1315–1334, 2020, doi: <https://doi.org/10.1287/opre.2019.1969>
- [13] G. M. Carter, J. M. Chaiken, E. Ignall, “Response areas for two emergency units,” *Operations Research*, vol. 20, no. 3, pp. 571–594, 1972, doi: <https://doi.org/10.1287/opre.20.3.571>
- [14] C. Cubillas, J. M. Alberola, V. Sanchez-Anguix, “A grid-based approach for ambulance dispatch in critical emergencies within static systems,” in *International Conference on Intelligent Data Engineering and Automated Learning*, 2024, pp. 171–183, Springer, https://doi.org/10.1007/978-3-031-77738-7_15
- [15] J. Roa, J. Escobar, C. Moreno, “An online real-time matheuristic algorithm for dispatch and relocation of ambulances,” *International Journal of Industrial Engineering Computations*, vol. 11, no. 3, pp. 443–468, 2020, doi: <https://doi.org/10.5267/j.ijiec.2019.11.003>
- [16] L. A. Albert, “A mixed-integer programming model for identifying intuitive ambulance dispatching policies,” *Journal of the Operational Research Society*, vol. 74, no. 11, pp. 2300–2311, 2023, doi: <https://doi.org/10.1080/01605682.2022.2139646>
- [17] Y. Kergosien, V. Bélanger, A. Ruiz, “A capacity sharing approach to manage jointly transportation and emergency fleets at ems organisations,” *International Journal of Production Research*, vol. 61, no. 3, pp. 880–897, 2023, doi: <https://doi.org/10.1080/00207543.2021.2018138>
- [18] K. K. Mohbey, “An efficient framework for smart city using big data technologies and internet of things,” *AISC*, vol. 714, pp. 319–328, 2018, doi: https://doi.org/10.1007/978-981-13-0224-4_29
- [19] S. Dunnett, J. Leigh, L. Jackson, “Optimising police dispatch for incident response in real time,” *Journal of the Operational Research Society*, vol. 70, no. 2, pp. 269–279, 2019, doi: <https://doi.org/10.1080/01605682.2018.1434401>
- [20] K. Liu, X. Li, C. C. Zou, H. Huang, Y. Fu, “Ambulance dispatch via deep reinforcement learning,” *Proceedings of the 28th International Conference on Advances in Geographic Information Systems, SIGSPATIAL GIS 2020*, 2020, doi: <https://doi.org/10.1145/3397536.3422204>
- [21] O. Elfahim, E. M. B. Laoula, Y. Mohamed, O. Barakat, M. Mohammed, “Deep reinforcement learning approach for emergency response management,” *2022 International Conference on Intelligent Systems and Computer Vision (ISCIV)*, 2022, doi: <https://doi.org/10.1109/ISCIV54655.2022.9806108>
- [22] C. Hua, T. Zaman, “Optimal dispatch in emergency service system via reinforcement learning,” *INFORMS International Conference on Service Science*, 2020, doi: https://doi.org/10.1007/978-3-030-75166-1_3
- [23] S. Luan, F. Pan, X. Chen, Q. Bing, J. Yang, X. Jin, “Markov-decision-process-based reinforcement learning for emergency vehicles of dispatch and reallocation strategies through the lens of efficiency and fairness: evidence from chinese emergency medical services system,” *Transportation Research Record*, vol. 2679, no. 4, pp. 1015–1030, 2025, doi: <https://doi.org/10.1177/03611981241299747>
- [24] Q. Wei, W. Dai, C. Tang, Z. Guo, F. Guo, K. Li, “A deep reinforcement learning approach for solving emergency medical supplies dispatching problem,” *Emergency Medicine and West China Biomedical Big Data Center, West China Hospital, Sichuan University*, 2024, doi: <https://doi.org/10.2139/ssrn.4913956>
- [25] S. Ji, Y. Zheng, Z. Wang, T. Li, “A deep reinforcement learning-enabled dynamic redeployment system for mobile ambulances,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 1, 2019, doi: <https://doi.org/10.1145/3314402>
- [26] J. E. Moen, “Deep reinforcement learning for supporting ambulance dispatch decisions,” *Norwegian University of Science and Technology*, 2022, <https://doi.org/10.48550/arXiv.2307.07508>
- [27] S. Chen, G. Luo, L. Chen, “Intelligent urban emergency response: Integrating large language models, multi-objective optimization, and reinforcement learning,” *IEEE Smart World Congress*, 2024, doi: <https://doi.org/10.1109/SWC62898.2024.00226>
- [28] E. Lamarca, “Offline reinforcement learning for ambulance dispatch,” *School of Engineering University Rovira Virgili*, 2023.

[29] A. Garriga, “Applying reinforcement learning techniques for emergency resource dispatch,” *School of Engineering University Rovira Virgili*, 2023.

[30] S. Ridler, A. J. Mason, A. Raith, “A simulation and optimisation package for emergency medical services,” *EJOR*, vol. 298, no. 3, pp. 1101–1113, 2022, doi: <https://doi.org/10.1016/j.ejor.2020.10a1r6l/oj.esjoCr.u20b2i1l.10a7.s038>



Carlos Cubillas

Carlos Cubillas is a PhD candidate in Computer Engineering and a strategic-technological consultant with more than 15 years of experience, specializing in solving complex business optimization problems across diverse industries. His current research lies at the intersection of multi-agent systems, reinforcement learning, simulation, and advanced optimization techniques, with a particular focus on developing intelligent agent-based models for real-time decision-making in emergency management and other complex dynamic environments.



Juan M. Alberola

Juan M. Alberola is an Associate Professor at the Universitat Politècnica de València (UPV), and serves as Director of the Smart AI Chair, focused on research and collaboration with industry in the application of AI to smart cities. He has participated in more than 20 national and international research projects and has authored over 70 publications in international journals and conferences. He has also served on the organizing and program committees of several international conferences and workshops and acts as a reviewer for multiple JCR-indexed journals. His research interests include multi-agent systems, mobility, optimization, and intelligent assistants, with a particular focus on the application and transfer of AI solutions to domains such as smart cities, healthcare, and education.



Jaume Jordán

Victor Sanchez-Anguix is an Associate Professor at the Universitat Politècnica de València (UPV), Associate Head of Student Affairs at the Computing School, and a researcher at the Instituto de Tecnología Informática. He received his Ph.D. in Artificial Intelligence from UPV in 2013, earning the Outstanding Thesis Award. His research focuses on multi-agent systems and applied optimization, particularly within the healthcare and education sectors. Dr. Sanchez-Anguix has an extensive publication record, including high-impact JCR journals and several renowned artificial intelligence conferences. He has led and participated in numerous national, international, and technology transfer projects. Additionally, he has previously worked as a data scientist in the private sector.



Victor Sanchez-Anguix

Dr. Jaume Jordán is an Associate Professor at the Universitat Politècnica de València (UPV) and a researcher at the Valencian Research Institute for Artificial Intelligence (VRIN). He received his Ph.D. in Artificial Intelligence from UPV with Outstanding Cum Laude and International Mention. His research focuses on Artificial Intelligence, particularly Multi-Agent Systems and optimization techniques. He has participated in 19 R&D projects, serving as principal investigator on several. He has authored more than 60 publications at international conferences and in JCR journals, and he actively collaborates as a reviewer, editor, and program committee member at international AI venues.