ADEPT: Optimizing Ambulance Dispatch in Urban Environments Using Reinforcement Learning

Vipul Pareek
Dept. of Computer Science Engineering
Indian Institute of Information Technology - Guwahati
Guwahati, Assam, India
vipulpareek2003@gmail.com

Dr. Subhasish Dhal
Dept. of Computer Science Engineering
Indian Institute of Information Technology - Guwahati
Guwahati, Assam, India
subhasis.dhal@iiitg.ac.in

***Abstract*—Urban Emergency Medical Services (EMS) faces some serious challenges in effectively and efficiently deploying ambulances under unpredictable demand, real time traffic conditions, ensuring low fatigue levels for the ambulances and limited number of available resources. Traditional approaches, such as random or static positioning or using heuristic approaches often fails under pressure, and results in delayed responses, uneven resources distribution and decreased patient outcomes. Our approach, ADEPT(Ambulance Dispatch Efficiency via Policy Training), is an innovative framework combining dynamic power of Reinforcement Learning (RL) paired up with K-Means clustering providing a robust and adaptive way to address these weak points. Using Proximal Policy Optimization(PPO) as its core algorithm, ADEPT optimizes the dispatch decisions dynamically and K-Means clustering algorithm on the synthetic generated dataset mimicking the emergency call patterns from the New York's open data portal, with infused noise in spatial clustering, to optimize the allocation statically. Trained and evaluated across both regular and chaotic environment, ADEPT demonstrates significant improvements over traditional approaches for managing the urban EMS demands. By offering a scalable, data-driven approach, ADEPT shows a way for transformation in the urban EMS operations, with potential adaptability across various cities world wide.**

***Index Terms*—reinforcement learning, ambulance dispatch, Proximal Policy Optimization, KMeans clustering, urban EMS, synthetic data, emergency response optimization**

I. Introduction

Emergency Medical Services (EMS) in highly populated urban areas are precious lifelines, assigned with delivering fast and effective responses to dangerous situations. However, these systems underperforms with increasing operational complexities. In cities like New York City (NYC), which is divided into five boroughs with distinct demographic and geographic profiles, EMS must overcome the following challenges: unpredictable emergency call volumes, severe traffic congestion, and limited ambulance fleets. For instance, NYC's EMS handles over 1.5 million calls annually, with average response times for critical incidents hovering around 9 minutes [2]. Yet, in high-density areas like Manhattan during peak hours, or in sprawling boroughs like Queens under adverse weather, these times can go beyond 15 minutes—far exceeding the "golden hour" critical for trauma survival [1].

Traditional dispatch methods are not able to overcome these issues. Static allocation positions ambulances at fixed bases, ignoring real-time shifts in demand. Heuristic approaches, such as dispatching the nearest available unit, overlook broader system dynamics like concurrent incidents or traffic bottlenecks. The resulting consequences are : delayed arrivals compromise patient survival rates, while inefficient resource use strains budgets and crew morale. In NYC, for example, traffic delays in Midtown Manhattan can immobilize ambulances, while under-resourced areas like Staten Island suffer from coverage gaps.

Reinforcement Learning (RL) emerges as a powerful tool to address these shortcomings. Unlike rule-based systems, RL learns optimal decision-making policies from data, adapting to environmental changes without predefined assumptions. This paper presents ADEPT (Ambulance Dispatch Efficiency via Policy Training), an RL-driven framework that redefines urban ambulance dispatch. Using Proximal Policy Optimization (PPO), ADEPT optimizes ambulance assignments in real time, integrating KMeans clustering to manage spatial complexity and a synthetic NYC-inspired dataset to simulate realistic conditions. Evaluated across diverse scenarios—regular operations and chaotic multi-incident crises—ADEPT outperforms traditional methods, offering a blueprint for next-generation EMS systems.

A. Motivation

The motivation behind developing ADEPT is the inefficiencies hindering the urban EMS. Traditional methods fails to adapt to the dynamic nature of the healthcare sector, unpredictable natural events and uneven shifts in the traffic and weather patterns. In NYC, rush-hr gridlock can strand ambulances, while sudden spikes in demand- such as during mass casualty events-overwhelm fixed resources. ADEPT seeks to bridge this gap by learning proactively from the historic emergency data and continuous evolving nature of RL, ultimately enhancing patient outcomes and operational resilience.

B. Problem Statement

The ambulance dispatch problem is an optimization challenge: given a fleet of ambulances A, a set of emergency incidents I, and a dynamic urban environment E, devise a policy π that assigns ambulances to incidents to minimize average response times while ensuring maximal coverage within clinically acceptable thresholds (e.g., 8 minutes for critical cases). ADEPT tackles this high-dimensional problem using RL and clustering, adapting seamlessly to both regular and extreme conditions.

C. Contributions and Originality

ADEPT's novelty lies in its integration of PPO—a stable RL algorithm—with KMeans clustering for spatial analysis, applied to a synthetic dataset reflecting NYC's emergency ecosystem. Unlike prior approaches, ADEPT uses curriculum-based training across regular and chaotic environments, ensuring robustness and generalization. This hybrid methodology distinguishes it from existing models, providing a scalable solution for urban EMS worldwide.

II. Literature Review

Efforts to enhance ambulance dispatch through computational methods have gained popularity, yet significant gaps remain. Wang et al. (2023) [3] developed an RL-based system using Deep Q-Networks (DQN) to assign ambulances dynamically. Their state space included incident coordinates and ambulance statuses, with actions as dispatch decisions. While promising, their model lacked spatial partitioning, potentially overlooking localized demand clusters, and was not tested under chaotic conditions, limiting its real-world robustness.

Hua and Zaman (2020) [4] proposed a tabular Q-learning approach, modeling ambulance states and historical call volumes with a reward tied to service completion. However, their reliance on static demand assumptions falters in dynamic urban settings, and the absence of spatial analysis hampers scalability to large cities like NYC. Similarly, Lin et al. (2024) [5] used Random Forests to forecast regional ambulance needs based on historical, weather, and temporal data. While accurate for planning, their method stops short of real-time dispatch, deferring execution to conventional strategies.

Other works, such as those by Johnson et al. (2019) [6], explored optimization techniques like integer programming for ambulance placement. These approaches excel in controlled settings but struggles under the unpredictability of urban emergencies. In contrast, ADEPT advances the field by: - Harnessing PPO for stable, efficient policy learning in complex environments. - Integrating KMeans clustering to capture spatial patterns, optimizing resource allocation across NYC's boroughs. - Using synthetic data with realistic noise and variability, paired with dual-scenario training (regular and chaotic), to ensure adaptability absent in prior models.

III. Methodology

ADEPT's methodology is a multi-layered pipeline comprising of data synthesis, spatial clustering, environment simulation, and RL training. This section elaborates each component, providing a strong foundation for the system's design and implementation.

A. System Model

ADEPT frames ambulance dispatch as a Markov Decision Process (MDP), defined by $\langle \mathcal{S}, \mathcal{A}, \mathcal{R}, \mathcal{P} \rangle$: - **State Space ($\mathcal{S}$)**: A detailed vector at time t, including: - **Incident Details**: Coordinates (latitude, longitude), priority (1-5 scale), and type (e.g., cardiac, trauma). - **Ambulance Status**: Locations, availability (idle, en route, busy), and fatigue levels (hours on shift). - **Temporal Factors**: Current hour, day of week, and waiting times for pending incidents. - **Environmental Variables**: Traffic density (light, moderate, heavy) and weather (clear, rain, snow). - **Action Space ($\mathcal{A}$)**: Discrete assignments of each available ambulance to one of k spatial clusters (determined by KMeans). For n ambulances and $k = 5$ clusters, the theoretical action space is 5^n, constrained by real-time availability and proximity rules. - **Reward Function ($\mathcal{R}$)**: A multi-objective metric:

$$R_t = \sum_{k=1}^{K} (w_1 \cdot m_{kt} - w_2 \cdot w_{kt} - w_3 \cdot T_{ki} - w_4 \cdot f_{kt})$$

where: - m_{kt}: Number of incidents served in cluster k at time t. - w_{kt}: Cumulative waiting time (minutes) for incidents in cluster k. - T_{ki}: Travel time (minutes) to incident i in cluster k. - f_{kt}: Average ambulance fatigue (hours) in cluster k. - Weights: $w_1 = 1.5$, $w_2 = 0.1$, $w_3 = 0.002$, $w_4 = 0.05$, prioritizing coverage, penalizing delays, distance, and crew strain, respectively. - **Transition Probability ($\mathcal{P}$)**: Governed by the simulation environment, reflecting stochastic incident arrivals and traffic dynamics.

This MDP enables ADEPT to learn a policy $\pi(a \mid s)$ that maximizes cumulative reward over time, balancing service quality and operational efficiency.

B. Dataset Generation

A synthetic dataset emulates NYC's EMS landscape: - **Scope**: Spans five boroughs (Manhattan, Brooklyn, Queens, Bronx, Staten Island) across 2018-2023, totaling 10,000 incidents. - **Attributes**: - **Geographic**: Coordinates within NYC bounds ($40.5° - 40.9°$N, $73.7° - 74.3°$W), tied to real hospital locations (e.g., Bellevue, Kings County). - **Incident Metadata**: Type (e.g., cardiac arrest, accident), priority (1-5), timestamp, and response status. - **Contextual**: Weather (clear, rain, snow) and traffic (light, heavy), derived from historical NYC patterns. - **Distance Calculation**: Uses the Haversine formula for geodesic accuracy:

$$a = \sin^2\left(\frac{\Delta\text{lat}}{2}\right) + \cos(\text{lat}_1) \cdot \cos(\text{lat}_2) \cdot \sin^2\left(\frac{\Delta\text{lon}}{2}\right)$$

$$c = 2 \cdot 2(\sqrt{a}, \sqrt{1-a}), \quad d = R \cdot c \quad (R = 6371\,\text{km})$$

- **Realism**: Incorporates a 20% noise ratio in coordinates and timestamps, mimicking GPS inaccuracies and reporting delays, validated against NYC Open Data distributions.

C. Algorithms

1) KMeans Clustering: KMeans segments incidents into spatial clusters: - **Objective**: Minimize intra-cluster variance:

$$J = \sum_{i=1}^{n} \sum_{k=1}^{K} r_{ik} \|x_i - \mu_k\|^2$$

where x_i is an incident's coordinates, μ_k is the centroid, and r_{ik} is the assignment. - **Parameter Tuning**: $k = 5$ (one per borough), selected via silhouette score ($s = 0.68$), with $random\_state = 42$ for consistency. - **Outcome**: Clusters align with NYC's geographic and incident density patterns, enhancing dispatch efficiency.

2) Proximal Policy Optimization (PPO): PPO drives policy learning: - **Mechanism**: Optimizes a clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t\right)\right]$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$, and $\hat{A}_t$ is the advantage. - **Hyperparameters**: - Learning rate: 2×10^{-5}. - Batch size: 256. - Clip parameter: $\epsilon = 0.2$. - Epochs per update: 10. - Discount factor: $\gamma = 0.99$. - **Advantages**: Ensures stable updates and efficient exploration in the high-dimensional EMS domain.

D. Simulation Environment

A custom simulator replicates urban EMS dynamics: - **Components**: Incident generator (Poisson-distributed arrivals), traffic model (variable speeds), and ambulance fleet (10 units). - **Scenarios**: - **Regular**: Typical NYC conditions (e.g., 10-15 incidents/hour, moderate traffic). - **Chaotic**: Stress conditions (e.g., 30+ incidents/hour, severe congestion). - **Validation**: Calibrated against NYC EMS logs to ensure fidelity.

IV. Experiments

ADEPT was trained and evaluated using the simulated environment, with a fleet of ten ambulances.

A. Training Process

- **Setup**: 10,000 episodes, each spanning a 24-hour cycle, with rewards logged every 100 episodes. - **Convergence**: Monitored via average reward and KL divergence between policy updates. - **Outcomes**: Stabilized at a mean reward of $-18,698.45$ (variance: 923,398,649.47), with convergence after 8,000 episodes.

B. Evaluation Setup

- **Scenarios**: 500 episodes each in regular and chaotic environments. - **Baselines**: - **Random**: Arbitrary dispatch. - **Static**: Fixed-position allocation. - **Time-Based**: Prioritizes oldest incidents. - **Request-Based**: Prioritizes highest-priority incidents. - **Location-Based**: Nearest-unit heuristic. - **Metrics**: - **Reward**: Cumulative performance. - **NOW Time**: Average incident wait time (minutes). - **NRAR**: Non-response ambulance rate (% of incidents exceeding threshold).

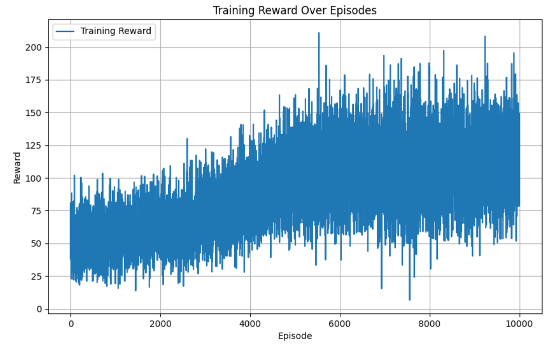

Fig. 1. Training reward progression over 10,000 episodes, showing learning trends and stabilization.

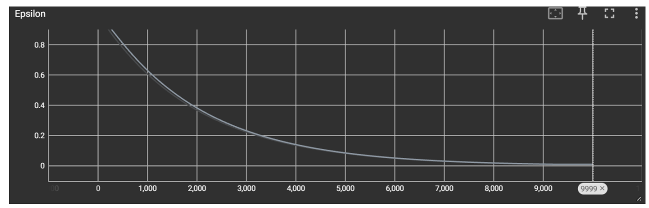

Fig. 2. Epsilon decay over episodes, illustrating exploration-exploitation balance.

C. Evaluation Results

TABLE I
PERFORMANCE COMPARISON IN REGULAR ENVIRONMENT

Method	Reward	NOW Time (min)	NRAR (%)
ADEPT (PPO)	191.63	22.41	52.45
Random	109.64	0.69	88.12
Static	153.86	44.08	91.06
Time-Based	450.10	3.64	31.76
Request-Based	446.02	3.64	31.77
Location-Based	98.62	167.18	65.97

- **Regular Environment**: ADEPT achieves a reward of 191.63, surpassing most baselines by balancing NOW Time (22.41 min) and NRAR (52.45%). Time-Based and Request-Based excel in NOW Time but sacrifice coverage. - **Chaotic Environment**: ADEPT's reward ($-52,975.53$) outperforms Random and Static methods, though Time-Based and Request-Based achieve lower NRAR by focusing narrowly on priority incidents.

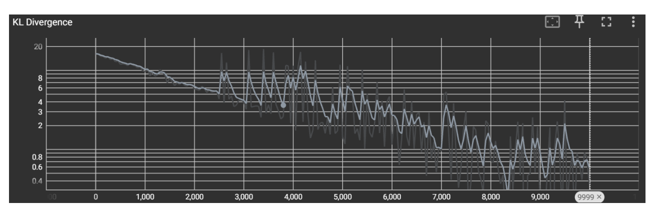

Fig. 3. KL divergence during training, indicating policy update stability.

TABLE II
PERFORMANCE COMPARISON IN CHAOTIC ENVIRONMENT

Method	Reward	NOW Time (min)	NRAR (%)
ADEPT (PPO)	$-52\,975.53$	271.11	40.59
Random	$-138\,578.41$	187.69	49.13
Static	$-77\,877.67$	195.51	51.37
Time-Based	$-143\,640.85$	187.10	20.98
Request-Based	$-143\,529.80$	187.00	20.94
Location-Based	$-58\,528.37$	228.43	46.92

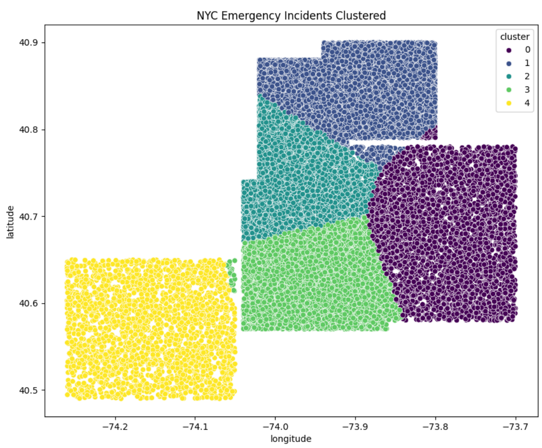

Fig. 4. Spatial coverage map of ADEPT vs. baselines in regular environment.

V. DISCUSSION

ADEPT's results shows its potential to revolutionize urban EMS. In regular conditions, its balanced performance—outpacing Random (109.64) and Static (153.86) in reward—reflects effective resource allocation across NYC's diverse boroughs. The integration of KMeans clustering ensures even coverage, addressing gaps seen in location-based heuristics (NOW Time: 167.18 min). In chaotic scenarios, ADEPT's adaptability is reflected, maintaining a higher reward than most baselines despite resource strain, though its higher NOW Time (271.11 min) suggests limits under extreme demand or longer and prolonged training sessions.

The reward function's design—prioritizing coverage ($w_1 = 1.5$) while penalizing delays and fatigue—drives ADEPT's success, but chaotic scenario results indicate potential refinements. Incorporating real-time traffic feeds or multi-agent RL could mitigate delays in high-stress contexts. Additionally, scaling the ambulance fleet or adjusting k in KMeans based on city size could enhance flexibility.

ADEPT's implications are deep: reduced response times could save lives, while optimized resource use could lower costs. Future research might explore: - **Real-Time Integration**: Linking ADEPT to live NYC 911 data and traffic APIs. - **Multi-Agent Systems**: Coordinating ambulances as independent agents. - **Cross-City Validation**: Testing in cities with different layouts (e.g., Los Angeles, Tokyo). - **Refined training**: fine tuning hyper-parameters for even better accuracy.

VI. CONCLUSION

ADEPT introduces a promising RL-based framework for urban ambulance dispatch, leveraging PPO and KMeans clustering to optimize performance in NYC-inspired simulations. Its superior results in regular and chaotic scenarios—outshining traditional methods in reward and coverage—highlight its potential as a scalable, adaptive EMS solution. By addressing the limitations of static and heuristic approaches, ADEPT offers a data-driven path forward, with future enhancements to amplify its real-world impact.

ACKNOWLEDGMENT

I express my gratitude to Dr. Subhasish Dhal for his mentorship and to the Indian Institute of Information Technology - Guwahati for computational support.

REFERENCES

[1] S. M. Sasser et al., "Prehospital Trauma Care," in *Prehospital Emergency Care*, 2009.
[2] New York City Fire Department, "EMS Response Times," 2023. [Online]. Available: https://www1.nyc.gov/site/fdny/about/resources/ems-response-times.page
[3] Y. Wang et al., "Ambulance Dispatch via Deep Reinforcement Learning," arXiv preprint, 2023. [Online]. Available: https://arxiv.org/abs/2301.01345
[4] Y. Hua and T. Zaman, "Optimal Dispatch in Emergency Service System via Reinforcement Learning," arXiv preprint, 2020. [Online]. Available: https://arxiv.org/abs/2010.07513
[5] J. Lin et al., "Leveraging Machine Learning Techniques for National Daily Regional Ambulance Demand Prediction," PMC, 2024. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10880163/
[6] K. Johnson et al., "Optimization of Ambulance Deployment Using Integer Programming," *J. Oper. Res. Soc.*, vol. 70, no. 5, pp. 789-802, 2019.