

A Hybrid Evolutionary and Machine Learning Approach for Priority-Based Ambulance Routing

Anouar Haddad^{1,3}, Takwa Tlili^{1,2}, Issam Nouaouri³, Saoussen Krichen¹

Abstract—Crises such as pandemics, natural disasters, and mass casualty incidents place immense pressure on emergency medical services to ensure rapid and effective patient care. Optimized ambulance routing is essential for these operations, as it directly influences response times, patient survival rates, and the overall resilience of healthcare systems. This study proposes a hybrid approach combining the Non-Dominated Sorting Genetic Algorithm III and automated machine learning, specifically the Tree-Based Pipeline Optimization Tool, to solve the priority-oriented ambulance routing problem. The Non-Dominated Sorting Genetic Algorithm III is used to minimize treatment completion times and prioritize critically ill patients, while the Tree-Based Pipeline Optimization Tool automates the generation of predictive models for probabilistic routing factors such as patient priority and traffic conditions. By integrating the optimization capabilities of evolutionary algorithms with machine learning, the proposed method improves both operational efficiency and equitable healthcare delivery. Experimental results show that the hybrid approach outperforms traditional methods in optimizing ambulance routes and prioritizing patients effectively.

KEYWORDS

Healthcare Optimization, NSGA-III, Automated Machine Learning, Priority assignment, Emergency medical systems

I. INTRODUCTION

In emergency healthcare systems, the Ambulance Routing Problem (ARP) is crucial, as effective decision-making can significantly impact patient outcomes. Global crises such as the COVID-19 pandemic, natural disasters, and terrorist attacks have highlighted the growing importance of efficient ambulance routing to ensure timely medical intervention and optimal resource use. Medical transport, particularly during emergencies, faces a unique set of challenges. These real-time operational challenges in EMS management primarily involve dispatching and routing, both of which are critical to ensuring effective emergency responses [29]. Dispatching is essential as it ensures that the most appropriate Emergency Medical Services (EMS) vehicles are selected based on the nature and location of emergency calls, directly impacting the speed and suitability of medical intervention. Routing is equally vital, as it determines the precise path an ambulance should take to reach the patient, minimizing costs and maximizing the efficient use of resources, which can ultimately save lives. The dynamic nature of these decisions, combined

with the urgency of medical care underscores the need for optimized and adaptable solutions in EMSs.

In response, researchers have directed significant attention toward developing innovative, sustainable solutions to overcome the challenges faced in this field, aiming to make emergency healthcare operations more effective and lifesaving outcomes more accessible. Our previous research [11], conducted by A. Haddad along with T. Tlili, I. Nouaouri, and S. Krichen, contributed to the field by applying the Non-Dominated Sorting Genetic Algorithm III (NSGA-III) as a metaheuristic to address the ARP. This approach focused on the dual objectives of minimizing total travel cost and reducing treatment completion time. Building on this work, the current study expands the scope of ARP by integrating patient priority constraints, ensuring that urgency levels are factored into routing decisions. In emergencies, prioritizing patients based on their medical condition is crucial, as it directly correlates with survival rates and outcomes. Patient prioritization directly equates with survival rates, adding a vital dimension to the problem formulation. Furthermore, this research introduces the integration of NSGA-III with machine learning techniques recognizing its potential as an innovative approach in the optimization process. These contributions advance ARP solutions by addressing both operational complexity and the humanitarian demands of emergency healthcare.

II. STATE OF THE ART

Optimizing healthcare systems, particularly in emergency medical services, is crucial for improving response times, resource allocation, and patient care. Research in this area addresses challenges such as ambulance routing, patient prioritization, and cost minimization, often involving multiple conflicting objectives. This section highlights recent advancements in optimization techniques applied to EMS and healthcare settings. For instance, [1] employed a modified grouping genetic algorithm to predict EMS demand, optimizing ambulance locations and fleet sizes, thereby demonstrating the robust potential of genetic approaches in enhancing service efficiency. Similarly, [2] focused on the location-routing problem during disaster scenarios, utilizing NSGA-II combined with a shuffled frog leaping algorithm to create a robust optimization framework that minimizes relief time and costs. [3] introduced a Mixed-Integer Linear Programming (MILP) model for ambulance routing in emergencies, integrating NSGA-II and multi-objective particle swarm optimization.

¹ LARODEC Laboratory, Institut Supérieur de Gestion de Tunis, Université de Tunis, Tunis, Tunisia.

² Institut Supérieur des Arts Multimédia de la Manouba, Université de la Manouba, Manouba, Tunisia.

³ Laboratoire de Génie Informatique et d'Automatique de l'Artois (LGI2A), Université d'Artois, Béthune, France.

Additionally, [10] presented a bi-objective model for the home care routing and scheduling problem, aiming to minimize operating costs while maximizing service levels by considering patient preferences. This study proposed a metaheuristic algorithm combining large neighborhood search with multi-directional local search. Furthermore, [22] addressed the COVID-19-induced real-time ambulance dispatching problem, focusing on minimizing response time, uncovered priority emergency calls, relocation time, and deviated ambulances. This was achieved using an improved multi-objective evolutionary algorithm with simulated annealing. [23] tackled the fair and timely distribution of medical supplies by proposing an open vehicle routing optimization model that minimizes demand satisfaction rate gaps and total delivery time. The model used a differential evolutionary algorithm with non-dominated sorting for effective Pareto optimization, validated with real-world COVID-19 data. [24] optimized ambulance dispatch, relocation, and technician allocation to improve emergency response and reduce system costs by proposing a bi-objective mathematical model. Small problems were solved using the augmented epsilon-constrained method, while larger instances were tackled using NSGA-II and multi-objective particle swarm optimization metaheuristics. [27] proposed a two-stage stochastic programming model to optimize ambulance location and routing, addressing uncertainties in emergency calls, travel times, and pathways. The model incorporates a heterogeneous fleet and prioritizes treatment golden time, using a progressive estimating algorithm. These approaches demonstrate significant advancements in tackling complex, multi-objective challenges in EMS.

Patient prioritization is critical in healthcare optimization, ensuring that critical cases receive prompt attention and improving overall system performance. Studies have emphasized the importance of tailored strategies to enhance service delivery. [4] developed a bi-objective model for the EMS vehicle routing problem, prioritizing critical patients while optimizing response times and costs using NSGA-II and a multi-objective bees algorithm. [5] tackled the home healthcare routing and scheduling problem, which shares similarities with the ARP in optimizing patient-centered transport but differs in its focus on planned care rather than emergency response. The research integrates patient satisfaction and uncertainty into a fuzzy optimization framework, solved using a new modified multi-objective version of the social engineering optimizer. [7] addressed routing and assignment in home health care services, balancing patient preferences, caregiver satisfaction, and operational costs using a mixed-integer linear programming model. Nurse scheduling was also explored by [6], optimizing visit schedules based on patient condition severity and time since the last visit using an adaptive large neighborhood search algorithm. A two-phase heuristic was introduced by [8] for a dynamic hospital dial-a-ride problem to prioritize patient convenience in real-time routing. [9] focused on a multi-objective maximum availability location problem, incorporating queuing theory to ensure population coverage reliability while minimizing

associated costs. Additionally, [12] proposed a three-phase metaheuristic combining the Longest Processing Time (LPT) algorithm for scheduling, the variable neighborhood descent algorithm for routing, and a satisfaction-maximizing heuristic to address home care service optimization, minimizing travel distances and maximizing patient satisfaction. [25] introduced a novel, model-driven decision support system for real-time ambulance relocation, maximizing demand coverage, minimizing travel time, and prioritizing demand using dynamically updated categories. These studies collectively showcase diverse approaches that prioritize patient needs in optimizing healthcare processes.

Despite the growing interest in the ambulance routing problem, there is a gap in research integrating Machine Learning (ML) techniques. Two notable studies in this area include [13], which introduces an online optimization approach that incrementally learns uncertainties in travel times, triage levels, treatment durations, and victim conditions during post-disaster scenarios, and [14], which tackles the challenge of unknown triage levels and treatment times in mass casualty incidents by employing a dynamic learning model to optimize ambulance routing and scheduling in real-time. [28] addresses the ambulance dispatch problem with a reinforcement learning-based strategy, using a multi-agent Q-network with experience replay to dynamically assign ambulances to patients, demonstrating its effectiveness through simulations and extensive experiments.

While machine learning focuses on leveraging data for predictive insights, the Internet of Things (IoT) enhances real-time decision-making by enabling seamless data collection and connectivity. For instance, [26] presents an IoT emergency service system utilizing a real-time node rank index algorithm to determine optimal ambulance routes, accounting for traffic congestion and distance, thus enhancing response times by selecting the most suitable ambulance and route. All the aforementioned studies are summarized in table I for a comparative overview.

In conclusion, researchers in EMS optimization have made substantial advancements, offering promising solutions for improving patient outcomes and healthcare efficiency. However, several challenges remain. Most studies rely on heuristic approaches with static parameters, limiting adaptability to real-time uncertainties like patient conditions, traffic dynamics, and resource availability. While some research incorporates stochastic models, the integration of emerging technologies such as artificial intelligence and real-time data analytics remains limited. Additionally, multi-objective optimization often focuses on response time and cost, overlooking patient-centric factors like service equity and dynamic prioritization. Although IoT-based solutions exist, their integration with optimization models for automated, intelligent decision-making is still underdeveloped.

TABLE I: Summary of Literature Review on EMS and Healthcare Optimization

Ref.	Approach	Technique	Problem Addressed	Key Contributions
[1]	Genetic algorithm	Modified grouping genetic algorithm	EMS demand prediction, ambulance location optimization	Improved fleet allocation and demand prediction efficiency
[2]	Metaheuristic	NSGA-II + Shuffled frog leaping algorithm	Location-routing problem in disaster scenarios	Minimized relief time and operational costs
[3]	Mathematical optimization	MILP + NSGA-II + Multi-objective PSO	Ambulance routing in emergency situations	Optimized emergency response and cost efficiency
[10]	Metaheuristic	Large neighborhood search + multi-directional local search	Home care routing and scheduling problem	Balanced operating costs and service level quality
[22]	Evolutionary algorithm	Improved MOEA + Simulated annealing	Real-time ambulance dispatching (COVID-19)	Minimized response time, uncovered priority calls, and relocation time
[23]	Evolutionary algorithm	Differential evolution + Non-dominated sorting	Fair medical supply distribution	Minimized demand satisfaction gaps and total delivery time
[24]	Mathematical optimization	Augmented epsilon-constrained method + NSGA-II + PSO	Ambulance dispatch, relocation, and technician allocation	Reduced system costs and improved response times
[27]	Stochastic programming	Two-stage stochastic programming + Progressive estimating algorithm	Ambulance location and routing under uncertainty	Considered heterogeneous fleets and treatment golden time
[4]	Metaheuristic	NSGA-II + Multi-objective bees algorithm	EMS vehicle routing problem	Prioritized critical patients while optimizing response time and cost
[5]	Fuzzy optimization	Modified MOSEO	Home healthcare routing and scheduling	Integrated patient satisfaction and uncertainty
[6]	Metaheuristic	Adaptive large neighborhood search	Nurse scheduling	Optimized visit schedules based on patient severity
[7]	Mathematical optimization	MILP	Home health care services routing	Patient preferences and caregiver satisfaction
[8]	Heuristic	Two-phase heuristic	Dynamic hospital dial-a-ride problem	Patient prioritization in real-time routing
[9]	Queueing theory	Multi-objective maximum availability location model	Population coverage in emergency response	Ensured reliability while minimizing cost
[12]	Metaheuristic	LPT + Variable neighborhood descent + Satisfaction-maximizing heuristic	Home care service optimization	Minimized travel distances, maximized patient satisfaction
[25]	Decision support system	Model-driven approach	Real-time ambulance relocation	Maximized demand coverage and minimized travel time
[13]	Machine learning	Online optimization	Post-disaster ambulance routing	Learned uncertainties to optimize response
[14]	Machine learning	Dynamic learning model	Mass casualty incident response	Optimized real-time routing and scheduling
[28]	Reinforcement learning	Multi-agent Q-network + Experience replay	Ambulance dispatch optimization	Improved assignment efficiency
[26]	IoT	Real-time node rank index algorithm	Traffic-aware ambulance routing	Optimized response time
[11]	Metaheuristic	NSGA-III	Ambulance routing problem	Optimized travel cost and completion time

III. PROBLEM DEFINITION

In the context of the ARP, ambulances are deployed from centralized medical hubs—such as hospitals or Points of Care (PoCs), to reach patients distributed across a geographical area. These patients are categorized based on the severity of their conditions. Figure 1 illustrates a representative POARP scenario. It includes:

- A central **hospital**, functioning both as the dispatch point and final destination for all ambulances;
- Multiple **patients**, with high-priority patients marked in red and lower-priority patients in green; and
- Several **ambulance routes**, each depicting how a vehicle is assigned to serve different patients.

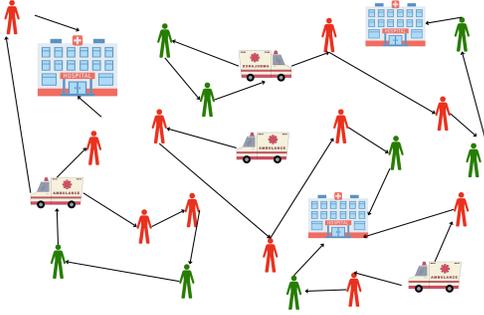


Fig. 1. Schematic illustration of the POARP

A. Mathematical Model

The mathematical model presented below builds upon our latest formulation [11]. In this updated formulation, we introduce a new objective function designed to respect patient priorities. This enhancement aligns with the goal of optimizing the routing of ambulances to patients while considering multiple objectives simultaneously. The model incorporates three distinct objectives:

Minimization of Total Travel Cost: This objective minimizes the overall travel cost incurred by all ambulances during the routing process. The goal is to reduce the cost associated with ambulance movement between patients and a PoC. This is achieved by minimizing the total distance. The objective is formulated as:

$$\min \sum_{i \in A} \sum_{j \in A} \sum_{k=1} c_{ij} x_{ijk}$$

where c_{ij} represents the travel cost between nodes i and j , and x_{ijk} is a binary decision variable indicating whether ambulance k travels from node i to node j .

Minimization of Completion Time Across Patient Priorities: The second objective considers patient priorities, introducing a weighted approach to minimize the completion time for red-code (high-priority) and green-code (lower-priority) patients. Red and green code patients are prioritized based on the severity of their medical needs, and the objective

seeks to ensure that these patients receive prompt treatment. This objective can be expressed as:

$$\min w_g \cdot E_g + w_r \cdot E_r$$

E_g and E_r represent the latest service completion times for green-code and red-code patients, respectively, and w_g and w_r are the weights given to each category.

Minimization of Makespan: The third objective aims to minimize the makespan, which is the total time required to complete all treatments within the scheduling period. It focuses on optimizing the overall treatment efficiency for all patients. The objective is defined as follows, where T_{\max} represents the maximum completion time across all patients.

$$\min T_{\max}$$

These objectives often conflict with one another. Minimizing travel cost tends to favor shorter routes, which may increase the makespan, while focusing on giving precedence to red and green code patients might lead to higher costs and longer travel times. Moreover, prioritizing patients can increase the makespan by dedicating more resources to high-priority cases. These conflicts are resolved through Pareto-optimal solutions, where no objective can be improved without worsening another.

$$\text{Min} \sum_{i \in A} \sum_{j \in A} \sum_{k=1} c_{ij} x_{ijk} \quad (1)$$

$$\text{Min} w_g \cdot E_g + w_r \cdot E_r \quad (2)$$

$$\text{Min} T_{\max} \quad (3)$$

$$\sum_{j \in A} \sum_{k=1}^{|K|} x_{ijk} = 1, \quad \text{for all } i \in E \quad (4)$$

$$\sum_{j \in A} \sum_{k=1}^{|K|} x_{jik} = 1, \quad \text{for all } i \in E \quad (5)$$

$$\sum_{j \in E} x_{ijk} = \sum_{j \in E} x_{jik}, \quad \text{for all } 1 \leq k \leq p, i \in V \quad (6)$$

$$\sum_{i \in A} \sum_{j \in E} x_{ijk} c_{ij} \leq C_{\max}^k, \quad \text{for all } 1 \leq k \leq p \quad (7)$$

$$\sum_{i \in F} \sum_{j \in E} x_{ijk} \leq 1, \quad \text{for all } 1 \leq k \leq p \quad (8)$$

In multi-objective optimization, it is essential to balance objectives such as minimizing costs, respecting patient priorities, and optimizing treatment time. These goals are addressed simultaneously using a Pareto-based approach to manage trade-offs effectively. The mathematical model is defined through a set of objective functions and constraints, as presented in equations (1)–(8). The objective function (1) minimizes the total travel cost. Objective function (2) aims to minimize the service completion time for red-code and green-code patients,

addressing patient priorities and preferences, as described in [4]. Lastly, objective function (3) minimizes the makespan, defined as the overall time required to complete all treatments in the schedule.

The model operates under a set of constraints similar to those outlined in our earlier formulation. Constraints (4) and (5) ensure that each patient is visited exactly once by a single ambulance. Constraint (6) enforces that ambulances must depart from the same location they visit. Additionally, each route must end at a hospital, and constraint (7) ensures that the travel budget is not exceeded. Finally, constraint (8) guarantees that all ambulances begin their routes at the PoC.

B. Parameters and Variables

The problem parameters, sets, and decision variables used in the mathematical formulation are summarized in table II.

TABLE II
PROBLEM PARAMETERS AND VARIABLES DEFINITION

Sets	
$E = \{1, 2, \dots, n\}$: Set of patients
$F = \{1, 2, \dots, m\}$: Set of PoCs
$P = \{1, 2, \dots, p\}$: Set of ambulances
$A = E \cup F$: Set of patients and PoCs
Parameters	
d_{ij}	: Distance connecting the vertices i and j
c_{ij}	: Travel cost between vertices i and j
c_{\max}^k	: Maximum travel cost of ambulance k
T_{\max}	: Maximum completion time of treatment
w_r	: Priority given to red code patients
w_g	: Priority given to green code patients
Decision Variables	
$x_{ijk} \in \{0, 1\}$: Binary decision variable showing if edge (i, j) is traversed by ambulance k
E_r	: Latest service completion time among all red code patients
E_g	: Latest service completion time among all green code patients

IV. PROPOSED APPROACH

This section introduces our methodology for solving Priority- Oriented Ambulance Routing Problem (POARP). We leverage NSGA-III as a metaheuristic optimization framework to minimize total travel cost, treatment completion time, and patient prioritization penalties.

A. NSGA-III Overview

The non-dominated sorting genetic algorithm III is an advanced evolutionary algorithm tailored for many-objective optimization problems, offering improved diversity handling compared to its predecessor, NSGA-II. Like its predecessor, it operates on a population of solutions that evolve over multiple generations through selection, crossover, and mutation. However, NSGA-III introduces a significant improvement in diversity preservation. Instead of relying on the crowding distance mechanism used in NSGA-II, NSGA-III employs a reference direction approach to maintain diversity across multiple objectives. This approach uses a predefined set of evenly distributed reference directions (or vectors) to guide the algorithm. These reference directions partition the objective

space, and each solution is associated with its nearest direction based on proximity in the objective space. This strategy ensures a well-distributed set of solutions across the objective space.

The key steps of NSGA-III, as outlined in its pseudo-code 1, include initialization, non-dominated sorting, and selection based on proximity to reference directions. By leveraging this reference direction-based selection strategy, the algorithm achieves a balanced trade-off between objectives, such as minimizing travel cost, completion time, and penalties for patient prioritization. The algorithm starts with an initial population of random solutions. Each solution represents a complete ambulance routing schedule. Fitness is evaluated across three objectives (as shown in steps 2–6), and solutions are ranked using non-dominated sorting. The reference direction mechanism is then applied to preserve diversity among trade-offs. At each generation, offspring are generated using crossover and mutation, followed by evaluation and selection. This tailored use of NSGA-III allows for finding well-balanced solutions that account for both logistical efficiency and patient urgency.

Algorithm 1 NSGA-III for POARP

Require: Cost matrix c , Priority weights w_g, w_r , Number of generations N_{gen} , Population size Pop_{size} , Reference directions Ref_{dirs}

Ensure: Pareto-optimal set of solutions

- 1: Initialize population P with random solutions
 - 2: Evaluate fitness values for each solution in P :
 - 3: F_1 : Total travel cost
 - 4: F_2 : Treatment completion time
 - 5: F_3 : Weighted patient prioritization penalties
 - 6: Perform non-dominated sorting to rank solutions into fronts
 - 7: Assign solutions to reference directions for diversity preservation
 - 8: **while** Current generation $< N_{\text{gen}}$ **do**
 - 9: Select parent solutions using tournament selection
 - 10: Apply crossover and mutation to generate offspring
 - 11: Evaluate offspring on all objectives
 - 12: Combine parent and offspring populations
 - 13: Perform non-dominated sorting and select the best solutions
 - 14: Update reference directions with new solutions
 - 15: **end while**
 - 16: **return** the final non-dominated Pareto-optimal set
-

B. Patient Priorities

In the ARP, patient prioritization is critical due to the varying levels of medical urgency among individuals needing care. Patients are categorized into different urgency levels to ensure efficient resource allocation. Specifically, they are classified into degree 1 (red-code) and degree 2 (green-code) urgency levels, based on the severity of their conditions:

- **Degree 1 (Red-Code Patients):** These patients require immediate medical intervention due to life-threatening

conditions such as severe trauma, cardiac arrest, or critical respiratory distress. Any delay in treatment significantly increases the risk of death or severe complications.

- **Degree 2 (Green-Code Patients):** These individuals have non-life-threatening conditions, such as minor fractures, mild respiratory issues, or stable injuries. While medical attention is still necessary, delays in treatment primarily result in discomfort or dissatisfaction rather than critical health deterioration.

To ensure that high-urgency cases receive priority in the routing process, we incorporate a weighted penalty objective that penalizes delays differently based on patient category. Specifically, the objective minimizes the maximum treatment completion time for each group, weighted by urgency level. The use of the maximum treatment completion time, rather than average, ensures that the worst-case delays—especially for high-risk patients—are minimized, aligning with life-critical response goals.

Eq. 1: Priority Objective Function

$$F = w_g \cdot \max(\text{completion time of green-code patients}) + w_r \cdot \max(\text{completion time of red-code patients}),$$

where w_g and w_r are priority weights assigned to green-code and red-code patients, respectively. The "completion time" is defined as the time from dispatch to the end of treatment for each patient. By assigning a higher weight to red-code patients ($w_r > w_g$), the algorithm implicitly prioritizes more urgent cases. This approach draws inspiration from the study [16], which emphasizes the role of triage in dispatch and fleet optimization. Triage-based classifications, like the red/green code system, are widely used in emergency medical research to manage resource allocation effectively. To set up this objective in our optimization routine, we implement the following procedure to calculate the weighted penalty value F :

Algorithm 2 Priority Objective Calculation (F)

Require: Binary solution vector x , Priority weights w_g, w_r , Number of PoCs N_{PoCs} , Number of patients N_{patients}

Ensure: Weighted priority objective F

- 1: Split x into:
 - 2: PoC section: $x[0 : N_{\text{PoCs}}]$
 - 3: Patient section: $x[N_{\text{PoCs}} :]$
 - 4: Compute completion times:
 - 5: Red patients: max service time for red patients
 - 6: Green patients: max service time for green patients
 - 7: Apply priority weights:
 - 8: $F = w_g \cdot \max(\text{green}) + w_r \cdot \max(\text{red})$
 - 9: **return** F
-

However, in real-world scenarios, the urgency and contextual importance of each patient group may evolve over time. To address this, we introduce a machine learning-based dynamic adjustment of priority weights in the next section.

C. Machine Learning Integration

In our case, the POARP involves prioritizing patients based on dynamically changing criteria such as urgency, location, and historical patterns. To enhance realism and adaptability in this priority assignment process, we integrated Automated Machine Learning (AutoML) into our framework. AutoML automates key tasks such as model selection, hyperparameter tuning, and pipeline optimization, allowing us to leverage data-driven insights without requiring extensive manual intervention. This is critical for our problem, as emergency response scenarios demand real-time decision-making informed by both historical data and current conditions.

In our problem, AutoML is used to dynamically adjust the priority weights for patients categorized as red-code and green-code. By training the AutoML model on historical priority data, the framework predicts weights that reflect the evolving importance of different patient categories during the optimization process. This real-time adaptability enables our routing algorithm to align resource allocation with situational demands, thus improving the overall efficiency and responsiveness of the problem. The versatility of AutoML in automating model selection and parameter optimization [17] makes it an ideal choice for our routing optimization framework. By reducing reliance on manual tuning, AutoML streamlines decision-making and ensures consistent, high-quality predictions. A reference study [18] demonstrates the successful use of NSGA-III combined with AutoML for energy-saving renovations, highlighting AutoML's ability to process large datasets, improve prediction accuracy, and make advanced optimization techniques accessible. Similarly, in our framework, AutoML analyzes complex relationships between patient urgency, resource availability, and historical data to inform dynamic priority adjustments.

For our problem, we selected the Tree-based Pipeline Optimization Tool (TPOT) as our AutoML platform. Among other well-known AutoML techniques, such as H2O AutoML and Auto-sklearn, TPOT stood out for its efficiency and unique features. As noted in [19], TPOT explores the search space faster and achieves superior performance on larger datasets. Additionally, TPOT integrates a multi-objective optimization mechanism, which is particularly relevant to our multi-objective ARP. This allows TPOT to simultaneously optimize several performance metrics during model training, such as prediction accuracy and runtime efficiency. TPOT's reliance on genetic programming to optimize machine learning pipelines [20] aligns with the evolutionary nature of our optimization approach. By employing both TPOT and NSGA-III, our methodology benefits from a cohesive framework where evolutionary principles guide both the routing optimization and the AutoML-driven priority adjustment.

In our case, the dynamic priority adjustment is directly incorporated into the optimization process. The AutoML model predicts a weight for green-code and red-code patients based on extracted features from the solution space. This prediction dynamically updates the priority weights w_g and w_r , influencing the third objective of the optimization problem,

which aims to minimize the weighted preference objective. This integration ensures that patient urgency is accurately reflected in the routing decisions, aligning the algorithm’s output with real-world emergency response priorities.

The combined use of NSGA-III and TPOT in our framework highlights the innovative synergy between optimization algorithms and machine learning. While NSGA-III handles the multi-objective routing optimization, AutoML enriches the framework by dynamically aligning optimization priorities with real-time predictions based on historical and contextual data. This hybrid approach exemplifies the potential of integrating advanced AI techniques into operational research problems, particularly in critical domains like healthcare and emergency management.

V. EXPERIMENTS AND RESULTS

For the experimental evaluation of the developed algorithms, testing was conducted on a system with the following hardware configuration: an Intel(R) Celeron(R) N4000 processor running at 1.10 GHz, coupled with 4 GB of RAM. This computational environment was utilized to assess the performance of the algorithms under various conditions.

A. Setup and Metrics

The feasibility of the proposed model was validated using Gurobi which is an exact optimization solver. This approach ensured that the three main objectives were effectively balanced while meeting all constraints. The choice of solver and parameter settings aligns with our previous study, enabling a direct comparison with earlier findings.

To assess the performance of NSGA-III, we adopted parameter values consistent with those used in [15] and our prior work. Table III summarizes the key problem parameters for large-scale instances. The number of ambulances (p) varies between 3 to 50, reflecting typical deployment scenarios in both small and large cities. The number of patients (n) in each instance is scaled to reflect both low and high-demand environments, with up to 120 patients to simulate urban and suburban areas. The maximum patient treatment time was determined by the severity of the patient’s condition, consistent with [15]. For non-critical patients, treatment times follow a normal distribution $N(10, 40)$ minutes, while for critical patients, treatment durations range from 25 to 120 minutes. This design ensures compatibility with real-world scenarios and facilitates comparisons with previous results.

TABLE III
PROBLEM CONSTANT PARAMETERS

Parameter	Description	Value
m	Number of hospitals	Up to 10
p	Number of ambulances	3 to 50
n	Number of patients	Up to 120
C_k^{\max}	Ambulance capacity	4

B. Performance Analysis

The performance of the proposed model was analyzed based on the Pareto front distribution and the hypervolume

convergence over generations. These metrics provide insights into the algorithm’s ability to explore the objective space and improve solution quality over iterations.

The 3D Pareto front, as shown in Figure 2, illustrates the trade-offs between the three objectives. The solutions are well-distributed across the objective space, indicating the algorithm’s effectiveness in capturing diverse trade-offs. Notably, lower total travel costs often correspond to slightly higher completion times or preference values, highlighting the inherent conflicts among objectives. The spread along the preference axis reflects the algorithm’s ability to balance all objectives, with mid-range solutions offering an efficient trade-off between cost and time, making them suitable for practical applications. Figure 3 presents the hypervolume convergence

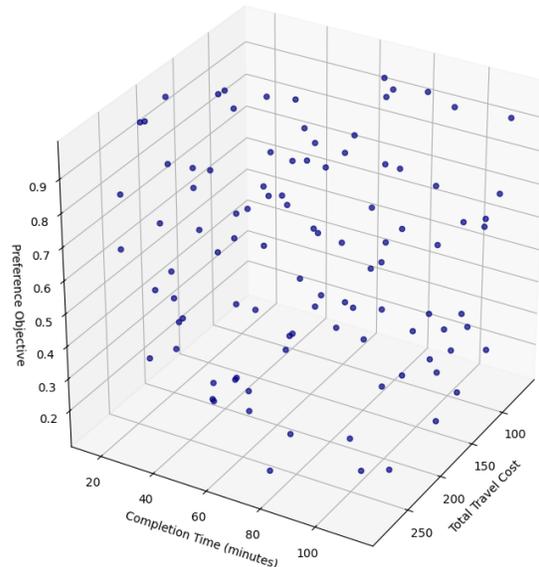


Fig. 2. Pareto Front Illustration: Trade-off between cost, time, and priorities

plot, which measures the algorithm’s approximation of the Pareto front over generations. The hypervolume quantifies the volume dominated by the solutions in the objective space, reflecting both their proximity to the optimal Pareto front and their diversity. A steady increase in hypervolume indicates consistent improvement in the quality of solutions. After multiple generations, the growth rate diminishes, suggesting convergence toward the true Pareto front. The hypervolume improvement plateauing after the 50th generation suggests near-optimal convergence with minimal trade-offs in solution diversity. This behavior confirms the algorithm’s capability to optimize all objectives simultaneously.

C. Machine Learning Evaluation

To evaluate the performance of our learned model, we chose the Mean Squared Error (MSE) as the primary evaluation metric. The MSE is widely used in regression tasks to measure the average squared difference between the predicted values and the actual values. A perfect model achieves an MSE of 0, indicating that the predictions match the actual values

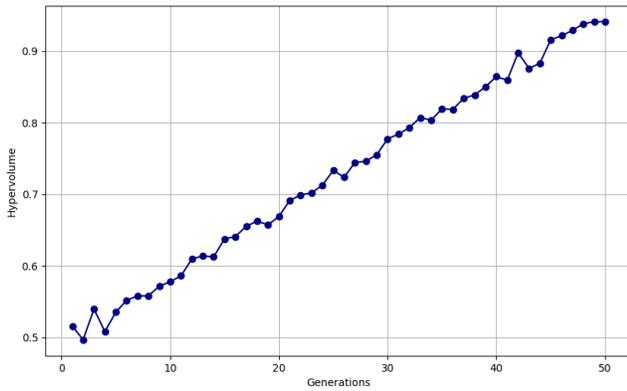


Fig. 3. Hypervolume convergence over generations

perfectly. As the MSE increases, it reflects greater errors in the predictions. The significance of the MSE metric is further highlighted in the literature, which notes that an MSE of 0 occurs when the linear regression model fits the data perfectly, while positive values reflect an imperfect fit [21].

In our evaluation, the MSE was calculated to be 0.0736. This result indicates that, on average, the model’s predictions deviate by approximately 0.0736 units from the actual values. When expressed as a percentage relative to the scale of the data, this corresponds to an average prediction error of about 7.36%. A lower MSE value, like the one obtained here, signifies that the model’s errors are relatively small. Besides, the relatively low MSE value of indicates that the AutoML model is providing fairly accurate predictions for the priority weights w_r and w_g . This, in turn, ensures that the NSGA-III optimization process effectively incorporates real-time patient priority adjustments, leading to more accurate and efficient solutions. This evaluation highlights that while the model demonstrates relatively small errors, further improvements could be made to enhance its predictive capabilities.

VI. CONCLUSION

This study introduced a hybrid approach that integrates NSGA-III with AutoML techniques to address the Priority-Oriented Ambulance Routing Problem. The evaluation of the algorithm’s performance demonstrated its ability to balance these objectives, providing solutions that span the trade-off space effectively. The model showed consistent improvement in solution quality over iterations, reflecting progress toward optimal outcomes. Additionally, the machine learning component of the model exhibited reliable predictive capabilities, though there remains room for refinement in its accuracy.

Future work could focus on refining the machine learning models to achieve even lower prediction errors and further enhancing the integration of NSGA-III with AutoML methods for more dynamic routing adjustments in real-time scenarios. Additionally, while the model showed consistent improvement in solution quality over iterations, the convergence behavior and computational efficiency require more thorough investigation, especially in real-time emergency scenarios where rapid decision-making is critical. Moreover, incorporating

more complex features, such as traffic conditions or varying emergency response priorities, could further improve the robustness of the system.

REFERENCES

- [1] S. Sasaki, A. J. Comber, H. Suzuki, and C. Brunson, "Using genetic algorithms to optimise current and future health planning - the example of ambulance locations," *Int. J. Health Geographics*, vol. 9, 2010.
- [2] H. Adarang, A. Bozorgi-Amiri, K. Khalili-Damghani, and R. Tavakkoli-Moghaddam, "A robust bi-objective location-routing model for providing emergency medical services," *J. Humanitarian Logistics Supply Chain Manag.*, vol. 10, 2020.
- [3] M. Rabbani, N. Oladad-Abbasabady, and N. Akbarian-Saravi, "Ambulance routing in disaster response considering variable patient condition: NSGA-II and MOPSO algorithms," *J. Ind. Manag. Optim.*, vol. 18, 2022.
- [4] E. Talebi, M. Shaabani, and M. Rabbani, "Bi-Objective Model for Ambulance Routing for Disaster Response by Considering Priority of Patients," *Int. J. Supply Oper. Manag.*, vol. 9, 2022.
- [5] A. M. Fathollahi-Fard, A. Ahmadi, F. Goodarzian, and N. Cheikhrouhou, "A bi-objective home healthcare routing and scheduling problem considering patients’ satisfaction in a fuzzy environment," *Appl. Soft Comput.*, vol. 93, 2020.
- [6] A. Cinar, F. S. Salman, and B. Bozkaya, "Prioritized single nurse routing and scheduling for home healthcare services," *Eur. J. Oper. Res.*, vol. 289, 2021.
- [7] L. En-nahli, H. Allaoui, and I. Nouaouri, "A Multi-objective Modelling to Human Resource Assignment and Routing Problem for Home Health Care Services," *IFAC-PapersOnLine*, vol. 48, 2015.
- [8] A. Beaudry, G. Laporte, T. Melo, and S. Nickel, "Dynamic transportation of patients in hospitals," *OR Spectrum*, vol. 32, 2008.
- [9] S. I. Harewood, "Emergency ambulance deployment in Barbados: a multi-objective approach," *J. Oper. Res. Soc.*, vol. 53, 2002.
- [10] K. Braekers, R. F. Hartl, S. N. Parragh, and F. Tricoire, "A bi-objective home care scheduling problem: Analyzing the trade-off between costs and client inconvenience," *Eur. J. Oper. Res.*, vol. 248, 2016.
- [11] A. Haddad, T. Tlili, I. Nouaouri, and S. Krichen, "Solving the multi-objective ambulance routing problem using NSGA-III," in *Proc. 2024 10th Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, 2024.
- [12] B. Issaoui, I. Zidi, E. Marcon, and K. Ghedira, "New Multi-Objective Approach for the Home Care Service Problem Based on Scheduling Algorithms and Variable Neighborhood Descent," *Electron. Notes Discrete Math.*, vol. 47, 2015.
- [13] D. Shiri, V. Akbari, and S. Salman, "Online algorithms for ambulance routing in disaster response with time-varying victim conditions," *OR Spectrum*, vol. 46, 2024.
- [14] D. Shiri, V. Akbari, and H. Tozan, "Online optimisation for ambulance routing in disaster response with partial or no information on victim conditions," *Comput. Oper. Res.*, vol. 159, 2023.
- [15] P. P. Repoussis, D. C. Paraskevopoulos, A. Vazacopoulos, and N. Hupert, "Optimizing emergency preparedness and resource utilization in mass-casualty incidents," *Eur. J. Oper. Res.*, vol. 255, 2016.
- [16] R. Tavakkoli-Moghaddam, P. Memari, and E. Talebi, "A Bi-Objective Location-Allocation Problem of Temporary Emergency Stations and Ambulance Routing in a Disaster Situation," in *Proc. 4th Int. Conf. Optimization Applications (ICOA)*, 2018.
- [17] X. He, K. Zhao, and X. Chu, "AutoML: A Survey of the State-of-the-Art," *Knowl.-Based Syst.*, vol. 212, 2021.
- [18] Y. Shi and P. Chen, "Energy retrofitting of hospital buildings considering climate change: An approach integrating automated machine learning with NSGA-III for multi-objective optimization," *Energy Build.*, vol. 319, 2024.
- [19] L. Parmentier, O. Nicol, L. Jourdan, and M.-E. Kessaci, "TPOT-SH: A Faster Optimization Algorithm to Solve the AutoML Problem on Large Datasets," in *Proc. IEEE 31st Int. Conf. Tools Artif. Intell. (ICTAI)*, 2019.
- [20] T. Agrawal, "Optuna and AutoML," in *Hyperparameter Optimization in Machine Learning*, 2021.
- [21] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation," *PeerJ Comput. Sci.*, 2021.

- [22] M. Hemici, D. Zouache, B. Brahmi, A. Got, and H. Dria, "A decomposition-based multiobjective evolutionary algorithm using Simulated Annealing for the ambulance dispatching and relocation problem during COVID-19," *Appl. Soft Comput.*, vol. 141, 2023.
- [23] F. Zheng, L. Du, X. Li, J. Zhang, B. Tian, and R. Jallad, "Multi-objective medical supplies distribution open vehicle routing problem with fairness and timeliness under major public health emergencies," *Manag. Syst. Eng.*, vol. 2, 2023.
- [24] F. Ravandi, A. Fathi Heli Abadi, A. Heidari, M. Khalilzadeh, and D. Pamucar, "A bi-objective model for location, dispatch and relocation of ambulances with a revision of dispatch policies," *Int. J. Cybernet. Syst. Manag. Sci.*, 2024.
- [25] M. Hajiali, E. Teimoury, M. Rabiee, and D. Delen, "An interactive decision support system for real-time ambulance relocation with priority guidelines," *Decis. Support Syst.*, vol. 155, 2022.
- [26] O. A. Mahdi, J. A. Eleiwy, Y. Rafid, B. A. Al-Mayouf, and B. AL-Attar, "An efficient node selection algorithm in the context of IoT-based vehicular ad hoc network for emergency service," *J. Intell. Syst.*, 2024.
- [27] F. Khoshghebari and S. M. J. Mirzapour Al-e-Hashem, "Ambulance location routing problem considering all sources of uncertainty: Progressive estimating algorithm," *Comput. Oper. Res.*, vol. 160, 2023.
- [28] K. Liu, X. Li, C. C. Zou, H. Huang, and Y. Fu, "Ambulance Dispatch via Deep Reinforcement Learning," in *Proc. 28th Int. Conf. Adv. Geogr. Inform. Syst.*, 2020.
- [29] R. Aringhieri, M. E. Bruni, S. Khodaparasti, and J. T. van Essen, "Emergency medical services and beyond: Addressing new challenges through a wide literature review," *Comput. Oper. Res.*, vol. 78, 2017.