

Article

Not peer-reviewed version

A Hybrid Deep Reinforcement Learning and Metaheuristic Framework for Heritage Tourism Route Optimization in Warin Chamrap's Old Town

[Rapeepan Pitakaso](#) , [Thanatkij Srichok](#) , [Surajet Khonjun](#) , [Natthapong Nanthasamroeng](#) ,
[Arunrat Sawettham](#) * , [Paweena Khampukka](#) , [Sairoong Dinkoksung](#) , Kanya Jungvimut ,
[Ganokgarn Jirasirilerd](#) , Chawapot Supasarn , [Pornpimol Mongkhonngam](#) , Yong Boonarree

Posted Date: 11 June 2025

doi: 10.20944/preprints202506.0964.v1

Keywords: heritage tourism; route optimization; deep reinforcement learning; metaheuristic; digital twin; multi-objective optimization; gan-based local search



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

A Hybrid Deep Reinforcement Learning and Metaheuristic Framework for Heritage Tourism Route Optimization in Warin Chamrap’s Old Town

Rapeepan Pitakaso ¹, Thanatkij Srichok ¹, Surajet Khonjun ¹, Natthapong Nanthasamroeng ², Arunrat Sawettham ^{3,*}, Paweena Khampukka ³, Sairoong Dinkoksung ³, Kanya Jungvimut ⁴, Ganokgarn Jirasirilerd ⁵, Chawapot Supasarn ³, Pornpimol Mongkhonngam ⁶ and Yong Boonarree ⁴

- ¹ Artificial Intelligence Optimization SMART Laboratory, Industrial Engineering Department, Faculty of Engineering, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand; rapeepan.p@ubu.ac.th (R.P.), surajet.k@ubu.ac.th (S.K.), thanatkij.s@ubu.ac.th (T.S.)
- ² Artificial Intelligence Optimization SMART Laboratory, Engineering Technology Department, Faculty of Industrial Technology, Ubon Ratchathani Rajabhat University, Ubon Ratchathani 34000, Thailand; natthapong.n@ubru.ac.th (N.N.)
- ³ Faculty of Management Science, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand; paweena.k@ubu.ac.th (P.K.), sairoong.d@ubu.ac.th (S.D.), chawapot.s@ubu.ac.th (C.S.)
- ⁴ Faculty of Applied Art and Architecture, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand; vimuti.kj@gmail.com (K.J.), boonarree@gmail.com (Y.B.)
- ⁵ Department of Industrial and Environmental Management Engineering, Faculty of Liberal Arts and Sciences, Sisaket Rajabhat University, Sisaket, 33000, Thailand; ganokgarn.j@sskru.ac.th (G.J.)
- ⁶ Office of Research, Academic Services and Art & Culture Preservation, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand; pornpimol.m@ubu.ac.th (P.M.)
- * Coresspodence: arunrat.s@ubu.ac.th

Abstract: Designing optimal heritage tourism routes in secondary cities involves complex trade-offs between cultural richness, travel time, carbon emissions, spatial coherence, and group satisfaction. This study addresses the Personalized Group Trip Design Problem (PGTDP) under real-world constraints by proposing DRL–IMVO–GAN—a hybrid multi-objective optimization framework that integrates Deep Reinforcement Learning (DRL) for policy-guided initialization, an Improved Multiverse Optimizer (IMVO) for global search, and a Generative Adversarial Network (GAN) for local refinement and solution diversity. The model operates within a digital twin of Warin Chamrap’s old town, leveraging 92 POIs, congestion heatmaps, and behaviorally clustered tourist profiles. The proposed method was benchmarked against seven state-of-the-art techniques, including PSO + DRL, Genetic Algorithm with Multi-Neighborhood Search (Genetic + MNS), Dual-ACO, ALNS-ASP, and others. Results demonstrate that DRL–IMVO–GAN consistently dominates across key metrics. Under equal-objective weighting, it attained the highest heritage score (74.2), shortest travel time (21.3 minutes), and top satisfaction score (17.5 out of 18), along with the highest hypervolume (0.85) and Pareto Coverage Ratio (0.95). Beyond performance, the framework exhibits strong generalization in zero- and few-shot scenarios, adapting to unseen POIs, modified constraints, and new user profiles without retraining. These findings underscore the method’s robustness, behavioral coherence, and interpretability—positioning it as a scalable, intelligent decision-support tool for sustainable and user-centered cultural tourism planning in secondary cities.

Keywords: heritage tourism; route optimization; deep reinforcement learning; metaheuristic; digital twin; multi-objective optimization; gan-based local search

1. Introduction

The global momentum toward decentralizing tourism has spotlighted the cultural, social, and economic potential of secondary heritage cities. Unlike major metropolitan centers, these locales are often embedded with vernacular landscapes, underrepresented narratives, and intimate community identities that offer tourists a more authentic, low-impact experience. In Thailand, Warin Chamrap's old town exemplifies such potential. The compact, walkable layout comprises riverside eateries, thematic museums, historic wooden houses, and religious landmarks. However, this richness is accompanied by new challenges: diverse tourist expectations, varying time budgets, spatial congestion, and environmental sensitivity must all be balanced in itinerary design. Classical trip planning models—typically driven by heuristics that prioritize distance or attraction count—fail to reconcile these constraints, often generating itineraries that are logistically efficient but culturally misaligned or environmentally suboptimal.

Recent technological advancements, especially in artificial intelligence and simulation, offer new frameworks to address these challenges. In particular, digital twins—virtual replicas of urban environments—have shown promise in simulating visitor behaviors, infrastructure capacity, and dynamic tourism flows in real-time [1,2]. When combined with machine learning and multi-objective optimization, these tools can support destination managers in balancing cultural preservation with tourist satisfaction. However, practical deployment in secondary cities, where data is sparse and visitor profiles are diverse, remains underexplored.

Several prior works have made notable strides. Di Napoli et al. [3] developed a DRL-based planner for optimizing city tours for cruise passengers, focusing on attraction density and time constraints [3]. Orabi et al. [4] introduced a multi-criteria event-driven travel sequence generator, emphasizing personalization through event context [4]. Meanwhile, Pitakaso et al. [5] fused reinforcement learning with metaheuristics to plan health-conscious tourist routes, balancing safety and route feasibility [5]. Similarly, Song and Chen [6] employed Q-learning to optimize cultural heritage tours in Macau, yet focused primarily on individual preferences, not group dynamics [6]. Zeinab Aliahmadi et al. [7] proposed a self-adaptive evolutionary framework for personalized trip design, but their approach lacks integration with simulation platforms like digital twins [7]. Additionally, sustainability perspectives in tourist behavior were explored by Pinho and Leal [8] and Suanpang and Pothipassa [9], both emphasizing the need for smart systems that reflect environmental, social, and accessibility constraints [8,9]. From a methodological standpoint, although evolutionary algorithms like genetic optimization have been widely applied in land-use and itinerary problems [10], their static nature limits adaptability under dynamic tourist flows and preference changes.

From this body of literature, two key research gaps remain open. First, most existing approaches focus on individualized or static trip planning, overlooking the complexity of group dynamics, real-time constraints, and preference heterogeneity in secondary heritage contexts. Second, on the methodological side, few studies integrate learning-based initialization, global evolutionary search, and generative refinement into a unified multi-objective optimization pipeline, especially one simulated within a digital twin environment for empirical validation.

In response to these gaps, this study proposes DRL-IMVO-GAN, a novel multi-objective framework that combines Deep Reinforcement Learning (DRL) for policy initialization, an Improved Multiverse Optimizer (IMVO) for global search, and a Generative Adversarial Network (GAN) for local perturbation and diversity enhancement. Embedded within a digital twin of Warin Chamrap's heritage zone, the system dynamically generates Pareto-optimal group itineraries that optimize trade-offs among five dimensions: cultural richness, travel time, emission reduction, route smoothness, and group satisfaction. The contributions of this paper are fourfold: (1) we develop and formalize the DRL-IMVO-GAN framework for solving the Personalized Group Trip Design Problem; (2) we construct a digital twin environment with real-world POIs and congestion simulation; (3) we benchmark against seven cutting-edge algorithms—including PSO+DRL and ALNS-ASP—demonstrating consistently superior performance across multiple KPIs; and (4) we

introduce new evaluation protocols for generalization, behavioral interpretability, and constraint resilience in multi-user tourism planning.

2. Related Work

This section provides a comprehensive overview of the foundational approaches and emerging technologies that inform the proposed hybrid framework. It is structured around four core domains: (1) multi-objective metaheuristics for tourism route planning, (2) the role of deep reinforcement learning in adaptive routing, (3) hybrid models integrating learning and search mechanisms, and (4) the use of digital twins in tourism and mobility optimization.

2.1. Multi-Objective Metaheuristics for Tourism Route Planning

Metaheuristics have long served as foundational tools in addressing the computational complexity of multi-objective routing problems. Their ability to simulate natural processes—such as selection, mutation, recombination, and swarm behavior—has made them especially valuable in generating diverse Pareto-optimal solutions for travel planning tasks that involve trade-offs among competing goals like cost, coverage, and user satisfaction. In tourism route planning, these algorithms are widely applied to construct efficient itineraries under spatial, temporal, and thematic constraints.

Recent studies have extended metaheuristic applications beyond static optimization, integrating them into more complex and context-sensitive planning frameworks. For instance, Aliano Filho and Morabito [11] developed a bi-objective, multi-period itinerary planning model using evolutionary techniques, demonstrating improved performance in optimizing tourist utility across multiple days [11]. Similarly, Ghobadi et al. [12] proposed a recommender system that integrates soft computing with clustering-based metaheuristics to personalize multi-day tourist itineraries based on user history and contextual preferences [12].

The integration of sustainability objectives into route optimization has also gained momentum. Pitakaso et al. [13] introduced a multi-objective trip design framework that harmonizes user preferences with environmental impact metrics using a hybrid reinforcement learning–metaheuristic model [13]. Likewise, Sabar et al. [14] emphasized the value of self-adaptive evolutionary algorithms for dynamic vehicle routing problems, particularly under real-time congestion, showing how adaptive strategies can enhance solution responsiveness [14].

In the context of heritage and cultural tourism, route planning requires more than just efficiency—it demands preservation of narrative continuity, spatial cohesion, and cultural immersion. Zhang et al. [15] addressed this by proposing a personalized tourism path decision model grounded in cultural significance metrics and POI thematic diversity [15]. Lin et al. [16] further reinforced this view through a systematic review of cultural routes as tourism products, highlighting how heritage-driven itineraries should prioritize route structure, interpretability, and experiential layering [16].

Yet, despite these advances, several methodological challenges persist. Traditional metaheuristics are often embedded in static frameworks, lacking the adaptive feedback mechanisms needed to respond to dynamic inputs such as pedestrian density, real-time crowding, or emergent group preferences. Xue et al. [17] explored this issue in the context of multi-modal urban mobility, using evolutionary games to model shifting travel behavior, but did not extend these dynamics to group-based cultural tourism planning [17]. Meanwhile, Zhang et al. [18] examined spatial systems along linear heritage corridors, emphasizing ecological-cultural connectivity but did not incorporate algorithmic personalization or group preference modeling [18].

More fundamentally, current methods often lack mechanisms for structural control, such as ensuring spatial smoothness and thematic progression—elements that are particularly critical in heritage tourism where disjointed routes may disrupt the intended cultural narrative. Ahmad [19], in a study of Kampong Ayer, warned of the negative impacts poorly planned tourism can have on fragile cultural destinations, underscoring the importance of planning tools that are sensitive not only to logistics but also to sociocultural continuity [19].

In sum, while multi-objective metaheuristics have significantly advanced tourism route planning, current approaches often fall short in adapting to real-time environmental dynamics, balancing sustainability with cultural experience, and generating structurally coherent itineraries for diverse travel groups. These gaps provide fertile ground for developing hybrid, adaptive optimization models that integrate dynamic learning, spatial simulation, and narrative-aware planning mechanisms.

2.2. Deep Reinforcement Learning for Adaptive and Context-Aware Routing

Deep reinforcement learning (DRL) has rapidly emerged as a transformative approach in sequential decision-making under uncertainty, offering remarkable flexibility in adapting to evolving environments. Its core advantage lies in the ability of agents to learn optimal long-term policies through continuous interaction with dynamic systems, adjusting to temporal fluctuations and spatial irregularities. In routing applications, DRL allows for real-time adaptability, enabling the optimization of complex criteria such as travel time, user satisfaction, and congestion avoidance.

The adoption of DRL in mobility systems has been explored across several high-stakes domains. For instance, Kim et al. [20] proposed a DRL-based adaptive scheduler for wireless time-sensitive networking (TSN), demonstrating the capability of deep agents to maintain low latency under changing transmission loads [20]. Similarly, in urban transport, Wu et al. [21] employed a multi-agent DRL framework to facilitate real-time planning for responsive bus routes, enabling agents to adapt collaboratively to passenger demands and route changes [21]. These examples highlight DRL's scalability and responsiveness in multi-agent, resource-constrained scenarios.

Within logistics, Geng et al. [22] demonstrated the effectiveness of DRL for dynamic travel time minimization by allowing routing policies to evolve in response to real-time conditions such as congestion and road availability [22]. Further expanding on adaptability, Bhadrachalam and Lalitha [23] integrated DRL with geographic routing for energy-efficient IoT communications, emphasizing environmental context awareness in routing decisions [23]. Such adaptability to external factors is crucial in tourism environments, where visitor interest, pedestrian density, and accessibility change over time.

Despite these advancements, DRL remains underutilized in tourism-specific route planning, especially where soft constraints like cultural interest alignment and experiential flow are essential. Chen et al. [24] proposed a multi-objective DRL framework for personalized trip recommendation, successfully balancing user satisfaction with logistical feasibility, yet focused on static user profiles without contextual responsiveness [24]. Likewise, Shafqat and Byun [25] utilized a context-aware hierarchical LSTM model for location recommendation, but the framework did not implement dynamic decision-making loops as found in DRL-based systems [25].

Recent literature emphasizes the need for explainability and context sensitivity in DRL applications. Kapoor [26] proposed a DRL-integrated graph neural network (GNN) for electric vehicle route optimization, advocating for models that are both intelligent and interpretable—a valuable direction for tourism applications where transparent itinerary logic can support trust and satisfaction [26]. From a theoretical perspective, Lazaridis et al. [27] provide a state-of-the-art walkthrough of DRL advancements, underscoring the importance of well-crafted state representations and reward shaping in high-dimensional spaces such as urban tourism systems [27].

Agricultural logistics research by Pitakaso et al. [28] further demonstrates DRL's utility in multi-agent fleet scheduling under resource failure scenarios, pointing toward its potential for robust itinerary planning under uncertainty [28]. However, these applications have not yet been translated into the tourism domain, where real-time adjustment to visitor needs and localized cultural dynamics remains a unique challenge.

In sum, while DRL has demonstrated strong potential in adaptive routing across logistics, communication networks, and mobility systems, its application in context-aware heritage tourism planning remains scarce. Critical gaps persist in modeling subjective dimensions such as cultural relevance, thematic flow, and visitor satisfaction within reinforcement learning frameworks. Addressing these challenges requires not only algorithmic innovation but also the construction of rich state

representations and reward functions that reflect both environmental realities and user experience dimensions of tourism.

2.3. Hybridization of Learning and Metaheuristic Strategies

The convergence of deep reinforcement learning (DRL) and metaheuristics has emerged as a powerful paradigm in solving complex combinatorial problems, particularly those requiring a balance between global exploration and local exploitation. In hybrid frameworks, DRL typically acts as a policy generator, learning context-aware strategies that guide the construction of promising initial solutions. These are then further refined using metaheuristics that exploit local neighborhoods and optimize solution quality based on domain-specific constraints. This architectural synergy is especially relevant in routing problems where trade-offs among multiple objectives and contextual sensitivities are prevalent.

Recent developments affirm the value of this hybrid approach across various planning domains. Torabi et al. [29] introduced a deep reinforcement learning hyperheuristic for the covering tour problem, demonstrating that a DRL-guided metaheuristic can dynamically select optimal neighborhood operations, significantly improving convergence rates and solution diversity [29]. Similarly, Sun et al. [30] applied DRL in conjunction with multi-objective scheduling heuristics for distributed hybrid flow shops, highlighting the framework's strength in managing resource conflicts and dynamic scheduling constraints [30].

In transportation systems, Zhang et al. [31] leveraged DRL in a railway itinerary optimization model augmented with graph neural networks (GNNs), showing how structural learning and policy adaptation can jointly improve routing under complex network conditions [31]. These applications demonstrate the feasibility of DRL-driven architectures in sequential routing tasks. However, most implementations remain centered on operational efficiency metrics, with limited attention to experiential or environmental dimensions—key in tourism routing.

The use of generative mechanisms within hybrid systems offers another layer of sophistication. By injecting structured noise into solution components, these models can maintain population diversity and avoid premature convergence. Freitas De Araujo-Filho et al. [32] and Zhao et al. [33] explored adversarial strategies via GANs for generating perturbations that guide search algorithms toward underexplored areas, effectively improving solution generalization in adversarial and optimization tasks [32,33]. Such mechanisms are highly transferable to itinerary planning contexts, where route diversity, spatial smoothness, and carbon-conscious decisions are non-trivial concerns.

In tourism routing specifically, hybrid learning-metaheuristic systems remain in early stages. Ruiz-Meza et al. [34] proposed a GRASP-VND hybrid for group-based tourist routing under fuzzy and sustainability criteria, while Derya et al. [35] addressed clustered TTDPs using intuitionistic fuzzy scores and non-linear travel times [34,35]. Although these methods demonstrate awareness of soft constraints such as environmental impact and group cohesion, they do not yet incorporate DRL or generative components. A recent foundational review by Gavalas et al. [36] underscored the absence of adaptive learning agents in current TTDP models, reinforcing the need for hybrid architectures capable of integrating contextual learning and global optimization [36].

In summary, the hybridization of DRL and metaheuristics has proven highly effective in domains requiring dynamic adaptability, heuristic diversity, and multi-objective balancing. However, its application to visitor-centric, multi-objective tourism routing remains largely uncharted, especially in contexts demanding alignment with soft constraints such as carbon impact, spatial smoothness, and experiential narrative continuity. This presents a compelling frontier for research, where intelligent hybrid systems can unlock new paradigms in adaptive, personalized, and sustainable travel planning.

2.4. Digital Twin Environments for Tourism and Mobility Systems

Digital twin (DT) technology is increasingly recognized as a transformative tool in simulating, managing, and optimizing complex physical systems by creating high-fidelity virtual counterparts. These virtual environments serve as dynamic mirrors of real-world systems, allowing the continuous integration of sensor data, behavioral models, and geospatial configurations. Within the domain of tourism and urban mobility, digital twins facilitate realistic simulation of visitor flows, environmental

conditions, and infrastructure usage—thereby enabling data-driven, context-aware decision-making for personalized itinerary design and route planning.

Recent research illustrates the growing application of digital twins in urban and tourism contexts. Florido-Benítez [37] highlights how digital twin frameworks support smart tourism destinations in anticipating and managing infrastructural and experiential challenges [37]. Similarly, Litavniec et al. [38] propose that digital twins can substantially enhance tourism competitiveness through dynamic interaction modeling, predictive analytics, and resource-aware visitor distribution [38]. In the broader mobility space, Aghaabbasi and Sabri [39] explore the use of DTs to analyze travel behavior decisions, arguing that digital environments can replicate the multifactorial nature of urban movement more effectively than traditional models [39].

Beyond analytics, digital twins have been paired with immersive technologies and agent-based models to simulate human-centric travel scenarios. Torrens and Kim [40] employed immersive virtual reality (VR) within digital twin environments to study pedestrian behavior at micro-spatial and temporal resolutions, providing a lens into nuanced human decision-making [40]. Likewise, Reffat [41] proposed an intelligent real-time virtual environment model to manage crowd behavior in urban settings, showing that digital twins can support dynamic control in highly congested tourism nodes [41]. These insights have implications for managing pedestrian congestion in heritage zones or theme parks, where visitor density can affect both experience quality and infrastructure strain.

Advancements in networked and spatial simulation have further broadened the application of digital twins to support AI-driven decision-making. For instance, Oliveira et al. [42] utilized DTs for wireless network optimization through virtualized action selection, underscoring how virtual environments can facilitate continuous learning and adaptation in real time [42]. Meanwhile, Aslam et al. [43] surveyed digital twin applications for autonomous vehicles and the metaverse, advocating for integration with game-theoretic and machine learning algorithms to enable proactive, user-centered system behavior [43].

In heritage and cultural tourism, digital twins have also shown promise in enhancing visitor engagement and spatial storytelling. Parrinello and Picchio [44] employed digital twin methodologies to visualize cultural routes across Europe, leveraging drone-based 3D reconstruction and semantic annotation to deepen experiential immersion [44]. Their work illustrates the potential for integrating digital twins with augmented and mixed reality technologies to enhance the interpretive depth of heritage landscapes.

Despite these advances, the full potential of digital twins in AI-augmented, visitor-centric tourism routing remains underrealized. Specifically, their integration with reinforcement learning agents and multi-objective optimization algorithms tailored for constraints like carbon impact, accessibility, and thematic continuity is limited. The proposed study seeks to bridge this gap by embedding a DRL agent within a digital twin of Warin Chamrap's old town, simulating spatial interactions and real-time constraints to support robust, context-aware itinerary generation.

3. Problem Formulation

Figure 1 presents a stylized map used to illustrate the core components of the personalized group trip design problem addressed in this study. The left panel shows a simplified urban grid with diverse points of interest (POIs), including houses, restaurants, and museums. A dashed blue path connects selected POIs, representing an optimized tour route from a designated start node to an end node. The right panel outlines key objectives addressed by the model, including heritage coverage, travel time, Euclidean distance, congestion, carbon emissions, and group preferences. Each objective is visually encoded with intuitive icons for easier interpretation. At the bottom, four avatars represent heterogeneous tourist group members whose personalized preferences and constraints must be satisfied. This graphical representation captures both the decision-making context and the practical complexity of real-world cultural tourism route planning, thereby providing an accessible visual aid to complement the mathematical model formulation.



Figure 1. A visual representation of the personalized group trip design problem with multi-objective optimization criteria.

Sets and Indices

N	Set of all POIs (points of interest), with generic elements $i, j \in N$
$A \subseteq N \times N$	Set of directed pedestrian arcs (i, j)
$N_{house}, N_{rest}, N_{muse} \subseteq N$	POI subsets by category
$s, t \in N$	Designated start and end nodes
G	set of group members, with generic element $g \in G$

Parameters

Symbol	Domain	Interpretation
h_i	$\mathbb{R} > 0$	heritage (coverage) score of POI i
δ_i	$\mathbb{R} > 0$	average dwell time at i
τ_{ij}	$\mathbb{R} > 0$	walking time along arc (i, j)
d_{ij}	$\mathbb{R} > 0$	Euclidean distance (i, j)
ϵ_{ij}	$\mathbb{R} > 0$	carbon emissions on (i, j)
κ_i	$[0, 1]$	real-time congestion index at i
ϕ_{ij}	$[0, \pi]$	heading angle of arc (i, j)
B^{time}	$\mathbb{R} > 0$	global route-time budget
B^{CO_2}	$\mathbb{R} > 0$	total emissions cap
B^{cong}	$\mathbb{R} > 0$	cumulative congestion cap
B^{dist}	$\mathbb{R} > 0$	total walking-distance cap
$Q^{house}, Q^{rest}, Q^{muse}$	$\mathbb{Z} \geq 0$	minimum visits per category
K^{max}	$\mathbb{Z} \geq 1$	maximum POIs in the route
M	<i>large</i>	big-MMM constant for timing constraints
s_{ig}	$\mathbb{R} > 0$	preference score of member g for POI i
B_g^{time}	$\mathbb{R} > 0$	individual time budget of g
R_g	$\mathbb{R} > 0$	minimum satisfaction required for g

$\omega_1, \dots, \omega_5$ $\mathbb{R} > 0$ with $\sum \omega_k = 1$ weights for scalarised objective

Decision Variables

$x_{ij} \in \{0,1\}$ 1 if arc (i,j) is traversed
 $y_i \in \{0,1\}$ 1 if POI i is visited
 $\mathcal{H}_i \in \mathbb{R} \geq 0$ Arrival time at POI i
 $u_i \in \{1, \dots, |N|\}$ Auxiliary MTZ variable for subtour elimination
 $z_{ig} \in \{0,1\}$ 1 if member g is “satisfied” by visiting i

Multi-Objective Functions

1) Maximize heritage with congestion discount

$$\text{Max } F_1 = \sum_{i \in N} h_i (1 - \kappa_i) y_i, \quad (1)$$

2) Minimize total travel time

$$\text{Min } F_2 = \sum_{(i,j) \in A} \tau_{ij} x_{ij}, \quad (2)$$

3) Minimize carbon emissions

$$\text{Min } F_3 = \sum_{(i,j) \in A} \epsilon_{ij} x_{ij}, \quad (3)$$

4) Minimize angular smoothness

$$\text{Min } F_4 = \sum_{(i,j) \in A} \sum_{(j,k) \in A} |\phi_{ij} - \phi_{jk}| x_{ij} x_{jk}, \quad (4)$$

5) Maximize group-preference satisfaction

$$\text{Max } F_5 = \sum_{g \in G} \sum_{i \in N} s_{ig} z_{ig} \quad (5)$$

Constraints:

Depot relations

$$\sum_{j \in N} x_{sj} = 1, \sum_{i \in N} x_{it} = 1, x_{is} = x_{tj} = 0, \forall i, j \in N \quad (6)$$

Flow conservation

$$\sum_{j \in N} x_{ij} = \sum_{j \in N} x_{ji} = y_i, \forall i \in N \setminus \{s, t\} \quad (7)$$

Temporal continuity and global time limit

$$\mathcal{H}_j \geq \mathcal{H}_i + \delta_i + \tau_{ij} - M(1 - x_{ij}), \forall (i, j) \in A, \quad (8)$$

$$\mathcal{H}_s = 0, \mathcal{H}_i + \delta_i \leq B^{time} \quad \forall i \in N \quad (9)$$

Distance, emission, and congestion budgets

$$\sum_{(i,j) \in A} d_{ij} x_{ij} \leq B^{dist}, \sum_{(i,j) \in A} \epsilon_{ij} x_{ij} \leq B^{CO_2}, \sum_{i \in N} \kappa_i y_i \leq B^{cong} \quad (10)$$

Category quotas

$$\sum_{i \in N_{house}} y_i \geq Q^{house}, \sum_{i \in N_{rest}} y_i \geq Q^{rest}, \sum_{i \in N_{muse}} y_i \geq Q^{muse} \quad (11)$$

Route size and subtour elimination (MTZ)

$$\sum_{i \in N} y_i \leq K^{max}, u_i - u_j + |N| x_{ij} \leq |N| - 1, \forall i \neq j, i, j \in N \setminus \{s\} \quad (12)$$

Group-preference satisfaction

$$z_{ig} \leq y_i, \forall i \in N, g \in G, \sum_{i \in N} s_{ig} z_{ig} \geq R_g, \forall g \in G \quad (13)$$

Individual time budgets

$$\sum_{(i,j) \in A} \tau_{ij} x_{ij} + \sum_{i \in N} y_i B_g^{time}, \forall g \in G \quad (14)$$

The first objective function (Equation (1)) is designed to maximize the total heritage value of the route while accounting for real-time congestion at each point of interest. By discounting the cultural score based on crowd levels, this formulation encourages the selection of historically significant yet less congested sites, thus enhancing both educational impact and visitor comfort. In contrast, the second objective function (Equation (2)) focuses on minimizing the total travel time required to traverse the route. This ensures that the resulting itinerary is time-efficient and feasible within practical and operational constraints.

The third objective (Equation (3)) aims to minimize the cumulative carbon emissions associated with movement between sites. By doing so, the model supports low-impact, environmentally conscious tourism planning. Complementing this, the fourth objective (Equation (4)) minimizes angular deviations in the walking route to promote smoother and more natural transitions between consecutive locations, thereby improving the physical comfort of the tour experience. The fifth objective (Equation (5)) seeks to maximize aggregate preference satisfaction across all group members. This personalization ensures that the final itinerary aligns closely with individual interests, thereby enhancing user engagement and perceived value.

To enforce feasibility, the model introduces several essential constraints. Equation (6) establishes correct depot behavior, ensuring that the route starts at a designated origin and ends at a defined destination. Flow conservation (Equation (7)) is maintained by requiring that each visited site is both entered and exited exactly once, preserving the logical continuity of the tour. Equations (8) and (9) implement time continuity between visits and enforce a global time budget, ensuring that all route activities are temporally coherent and complete within the allowed duration.

Environmental and physical constraints are encoded in Equation (10), which limits total walking distance, cumulative emissions, and crowd exposure along the selected route. To guarantee content diversity, Equation (11) imposes minimum category quotas—ensuring that the tour includes a balanced selection of heritage houses, restaurants, and museums. Route compactness and completeness are enforced in Equation (12), which limits the total number of visited points and eliminates subtours using a Miller–Tucker–Zemlin (MTZ) formulation.

Equation (13) introduces personalized satisfaction constraints by requiring that individual group members achieve a minimum level of preference fulfillment, and by linking satisfaction to actual site visits. Finally, Equation (14) ensures that the total tour time—comprising both travel and dwell components—remains within each member’s personal availability window. This inclusion of individualized time budgets ensures equitable feasibility and personalization across the entire group.

4. Research Methodology

This section presents the proposed two-stage hybrid optimization framework, designed to solve the personalized heritage-tourism routing problem introduced in Section 3. The framework leverages deep reinforcement learning (DRL) to construct adaptive initial solutions and employs a refined multi-verse optimizer (IMVO) enhanced by a GAN-driven local search module to iteratively refine the solution set. This integrative approach aims to efficiently navigate the complex, high-dimensional, and constrained solution space associated with real-world heritage routing under personalized, environmental, and operational constraints.

Figure 2 presents a high-level schematic of the proposed two-phase route optimization framework tailored for personalized heritage tourism in Warin Chamrap. This framework integrates

advanced artificial intelligence methodologies to balance heritage coverage, travel efficiency, sustainability, and group-specific preferences. The right-hand side of the diagram illustrates Phase I, which employs Deep Reinforcement Learning (DRL) within a spatially accurate digital twin environment. This phase encompasses key components including MDP formulation, multi-objective reward function design, policy network training via PPO, and the generation of a diverse set of feasible initial solutions. The left-hand side of the diagram illustrates Phase II, which applies an Improved Multi-Verse Optimizer (IMVO) combined with a GAN-driven local search module to refine the DRL-generated population. This phase incorporates fitness recalibration, Pareto front construction, feasibility screening, and multi-criteria selection using TOPSIS, ultimately yielding a set of deployable and policy-compliant routing plans. Together, the two phases form a robust, interpretable, and scalable solution methodology for complex multi-objective tour planning under real-world constraints.

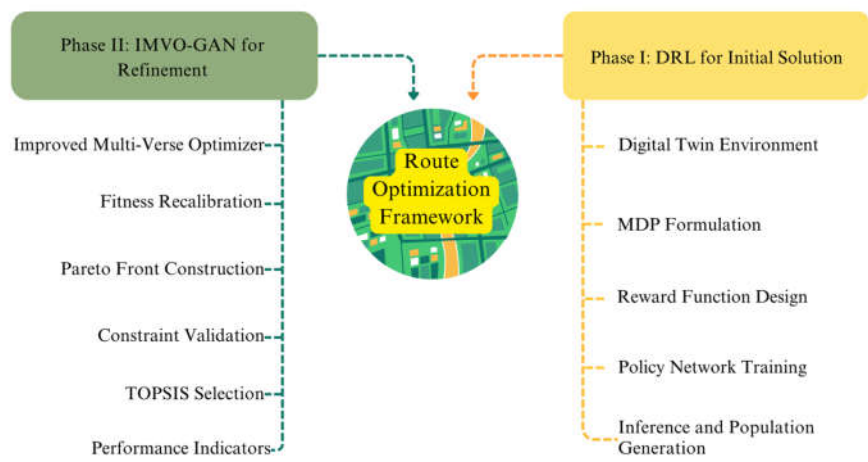


Figure 2. Architecture of the proposed route optimization framework.

4.1. Phase I: Deep Reinforcement Learning for Solution Construction

This phase employs a Deep Reinforcement Learning (DRL) framework embedded in a geospatially accurate digital twin of Warin Chamrap’s old town to generate an initial population of feasible, high-quality, and personalized route solutions. The DRL agent is trained to construct tours that comply with personalized constraints, urban mobility limits, and sustainability requirements. This section details the problem formulation as a Markov Decision Process (MDP), state and action representations, reward design, and training procedure.

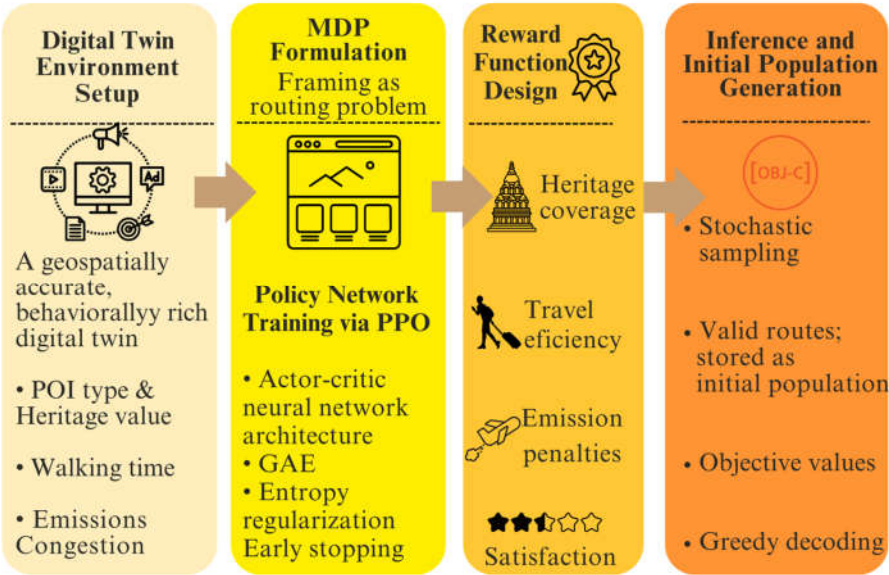


Figure 3. DRL-based initial solution construction pipeline within a digital twin environment.

Figure 3, illustrates Phase I of the proposed framework, where a Deep Reinforcement Learning (DRL) agent is embedded within a digital twin of Warin Chamrap’s old town to generate constraint-compliant, diverse route solutions. The architecture integrates geospatially accurate urban simulation, Markov Decision Process (MDP) formulation, multi-objective reward

4.1.1. Digital Twin Environment

To enable intelligent and context-sensitive itinerary construction, this study establishes a high-fidelity digital twin environment that virtualizes the pedestrian and cultural landscape of Warin Chamrap’s old town. The digital twin serves as a dynamic simulation space within which a deep reinforcement learning (DRL) agent can learn routing strategies that are spatially feasible, temporally constrained, and behaviorally personalized. It mirrors real-world constraints and data signals, ensuring that agent decisions remain consistent with operational realities.

The digital twin is constructed as a GIS-integrated directed graph $G = (N, A)$, where N denotes the set of geo-referenced points of interest (POIs), and $A \subseteq N \times N$ represents pedestrian-accessible arcs. The development of this environment integrates four core data layers:

1. Spatial positions and category types of POIs: Each node $i \in N$ is mapped using GPS coordinates derived from open-source geospatial databases and verified by municipal land-use records. Each POI is categorized into one of several functional groups, such as historic houses, temples, museums, or food establishments. Metadata for each POI includes average dwell time, operating hours, cultural significance (heritage score), and group relevance (preference weight vectors).
2. Arc-based pedestrian infrastructure with travel time and distance matrices: The arcs in set AAA are extracted from high-resolution pedestrian networks using OpenStreetMap and cross-validated with satellite imagery and ground surveys. Each arc (i, j) is annotated with its Euclidean distance d_{ij} , estimated walking time τ_{ij} , and angular direction ϕ_{ij} , for use in route smoothness evaluation. These attributes are stored in sparse matrices and accessed during state transitions in the DRL framework.
3. Real-time and historical crowd density heatmaps: A time-indexed congestion index $\kappa_i(t) \in [0,1]$ is assigned to each POI using data aggregated from mobile device signals, CCTV analytics, and historical pedestrian surveys. This layer allows the digital twin to dynamically reflect variable foot traffic across different times of day, enabling the routing agent to anticipate and adapt to crowding effects.

4. Carbon emission coefficients for walking paths: Each arc is also assigned an emission estimate ϵ_{ij} , based on energy expenditure models for pedestrian locomotion under urban conditions. While walking itself is low-emission, total route emissions are still tracked and constrained to reflect carbon accountability in sustainable tourism systems.

In addition to these core data components, the digital twin includes temporal logic modules that simulate POI availability based on time windows (e.g., opening and closing hours), pedestrian path obstructions, and special event scenarios (e.g., markets, festivals). This enables state-dependent availability masking, ensuring that action feasibility evolves throughout the tour planning horizon.

The environment also models personalized tourist behavior by embedding individual or group preference vectors p_g , which encode interest levels across POIs. These vectors are generated from a mixture of empirical survey data, user profiles, and semantic clustering of interest patterns. These preference profiles are critical for determining satisfaction-based rewards and personalized feasibility filters during itinerary generation.

To support DRL interactions, the digital twin continuously tracks state variables such as cumulative time, distance, emissions, visited categories, and satisfaction coverage. It enforces all feasibility constraints in real time, thereby guiding the learning process toward valid and practically deployable routing strategies. The environment is implemented using Python with support from libraries such as GeoPandas, NetworkX, and raster-based congestion visualizations.

In sum, the digital twin environment provides a geospatially grounded, temporally dynamic, and behaviorally rich decision space. It serves as a rigorous simulation ground for personalized itinerary planning, ensuring that every action taken by the learning agent is both context-aware and operationally viable.

4.1.2. MDP Formulation and Policy Design

Within the digital twin environment, the personalized heritage-tourism routing problem is formalized as a finite-horizon Markov Decision Process (MDP) to facilitate learning through Deep Reinforcement Learning (DRL). This formulation enables the agent to iteratively construct tour routes that account for spatiotemporal feasibility, personalized preferences, and real-world constraints. The MDP is defined by the tuple as addressing in Equation (15).

$$M = \langle S, A, P, R, \gamma, T_{max} \rangle \quad (15)$$

where:

S is the set of environment states,

A is the set of admissible actions,

P defines transition dynamics,

R is the reward function,

$\gamma \in [0,1]$ is the reward discount factor, and

T_{max} is the maximum decision horizon.

State Representation

Each state $s_t \in S$ at decision step t encodes the current routing context and accumulative metrics, defined as mentioned in Equation (16):

$$s_t = [x_t, v_t, \Delta t_t, \Delta e_t, \Delta d_t, q_t, p_g] \quad (16)$$

where:

$x_t \in N$ is the current POI,

$v_t \in \{0,1\}^{|N|}$ is a binary visitation vector,

$\Delta t_t, \Delta e_t, \Delta d_t \in \mathbb{R}_{\geq 0}$ denote cumulative travel time, emissions, and distance,

$q_t \in \mathbb{Z}_{\geq 0}^3$ tracks visited POIs by category (house, restaurant, museum),

$p_g \in [0,1]^{|N|}$ is the group preference vector.

This high-dimensional state representation allows the agent to evaluate current performance, remaining budget, and group satisfaction in real time.

Action Space and Feasibility Masking

At each step, the agent selects an action $a_t \in \mathcal{A}(s_t)$, which corresponds to choosing the next POI to visit. The admissible action set $\mathcal{A}(s_t)$ is dynamically filtered based on feasibility rules (Equation (17)):

$$\mathcal{A}(s_t) = \{j \in N \setminus \{x_t\} \mid \text{Feasible}(s_t, j) = \text{True}\} \quad (17)$$

where feasibility is determined by checking that visiting node j from x_t will not result in violations of any of the constraints.

This constraint-aware masking ensures that only routes adhering to operational and personalized requirements are explored.

Transition Dynamics

The transition function $\mathcal{P}(s_{t+1} \mid s_t, a_t)$ is deterministic in this domain. Upon selecting a POI j , the state is updated by:

- incrementing cumulative metrics based on $\tau_{x_t j}, \epsilon_{x_t j}, d_{x_t j}$,
- updating the visitation vector v_{t+1} ,
- incrementing the relevant category count in q_{t+1} ,
- changing the current location to $z_{t+1} = j$.

This transition logic guarantees that state evolution remains coherent with route progression and constraint tracking.

Policy Parameterization

The routing policy $\pi_\theta(a_t \mid s_t)$ is modeled as a stochastic mapping from states to actions, parameterized by a neural network with weights θ . The goal is to learn a policy that maximizes the expected cumulative return (Equation (18)):

$$J(\theta) = \mathbb{E}_{\pi_\theta} \left[\sum_{t=0}^{T_{max}} \gamma^t r_t \right] \quad (18)$$

To capture both spatial and preference-based signals, the input to the policy network includes: (1) current POI embedding, (2) preference-weighted visitation vector, (3) normalized cumulative budget usage, (4) congestion level at candidate POIs. This structured representation allows the policy to generalize across states with varying temporal and group-level contexts.

Episode Termination

An episode ends either when:

- The terminal POI is reached (i.e., the end-node constraint is satisfied),
- No further feasible POIs remain in $\mathcal{A}(s_t)$,
- Any hard constraint (e.g., time or emission) is violated during an attempted transition.

At termination, the episode is evaluated in terms of feasibility and objective function performance, feeding directly into the reward architecture described in Section 4.1.3.

4.1.3. Reward Function Design

The reward function serves as the core mechanism guiding the DRL agent toward constructing feasible, high-quality, and personalized heritage-tour routes. It integrates multiple conflicting objectives and enforces behavioral alignment with problem-specific constraints defined in the underlying mathematical model. The reward function is carefully structured to encode both instantaneous feedback and cumulative goal progression, enabling the agent to learn nuanced policies that optimize diverse performance criteria under strict feasibility conditions.

Composite Reward Structure

At each time step t , the agent receives a scalar reward r_t , defined as a weighted sum of interpretable and domain-grounded components (Equation (19)):

$$r_t = \lambda_1 r_t^{\text{heritage}} + \lambda_2 r_t^{\text{travel}} + \lambda_3 r_t^{\text{emission}} + \lambda_4 r_t^{\text{quota}} + \lambda_5 r_t^{\text{satisfaction}} + \lambda_6 r_t^{\text{smoothness}} - \lambda_7 r_t^{\text{penalty}} \quad (19)$$

where $\lambda_1, \dots, \lambda_7 \in R \geq 0$ are manually tuned scalar weights reflecting the importance of each sub-objective. These weights are normalized such that $\sum_{k=1}^7 \lambda_k = 1$ to maintain numerical stability and ensure balanced learning dynamics.

Each subcomponent is defined as follows:

1. Heritage Coverage Reward (r_t^{heritage})

Rewards the agent for visiting POIs with high cultural significance, discounted by real-time crowding levels (Equation (20)):

$$r_t^{\text{heritage}} = h_{i_t} \cdot (1 - \kappa_{i_t}), \quad (20)$$

where h_{i_t} is the heritage value of the visited POI and κ_{i_t} is the normalized congestion index at that location.

2. Travel Efficiency Penalty (r_t^{travel})

Penalizes long or inefficient transitions between POIs (Equation (21)):

$$r_t^{\text{travel}} = -\tau_{i_{t-1}, i_t}, \quad (21)$$

where τ_{i_{t-1}, i_t} is the walking time between the previous and current POI.

3. Emission Control Penalty (r_t^{emission})

Penalizes environmentally costly paths (Equation (22)):

$$r_t^{\text{emission}} = -\epsilon_{i_{t-1}, i_t} \quad (22)$$

with ϵ_{i_{t-1}, i_t} representing estimated emissions along the path.

4. Quota Fulfillment Reward (r_t^{quota})

Encourages the agent to meet category-specific quotas (e.g., number of museums, restaurants, houses visited) (Equation (23)):

$$r_t^{\text{quota}} = \delta(q_t) - \delta(q_{t-1}) \quad (23)$$

where $\delta(\cdot)$ is a function measuring how many quota conditions are currently fulfilled.

5. Group Preference Satisfaction Reward ($r_t^{\text{satisfaction}}$)

Promotes alignment between the route and group member preferences (Equation (24)):

$$r_t^{\text{satisfaction}} = \frac{1}{|G|} \sum_{g \in G} [p_{i_{t-1}, g} \cdot 1_{\{i_t \in V_t\}}] \quad (24)$$

where $p_{i_{t-1}, g}$ is the preference score of member g for the visited POI i_t , and V_t is the set of previously visited POIs.

6. Smoothness Reward ($r_t^{\text{smoothness}}$)

Encourages angular continuity between route segments (Equation (25)):

$$r_t^{\text{smoothness}} = -|\phi_{i_{t-2}, i_{t-1}} - \phi_{i_{t-1}, i_t}| \quad (25)$$

where ϕ_{ij} is the angular heading from POI i to POI j . This term is activated only from step $t \geq$

2.

7. Infeasibility Penalty ($r_t^{penalty}$)

A large negative penalty applied when the agent attempts a move that would lead to violation of any hard constraint (Equation (26)):

$$r_t^{penalty} \begin{cases} \rho, & \text{if } ConstraintsViolated(s_t, a_t) = True \\ 0, & \text{otherwise,} \end{cases} \quad (26)$$

where $\rho \ll 0$ is a fixed infeasibility penalty constant, typically set an order of magnitude higher than other terms to suppress invalid routes.

Terminal Rewards and Episode Evaluation

At the end of each episode, additional terminal rewards are applied based on global feasibility and objective satisfaction:

- A feasibility bonus is awarded if all constraints (e.g., time, quota, emissions, end node reachability) are satisfied.
- A completeness reward is issued if all five objective dimensions (heritage, travel, emissions, satisfaction, smoothness) exceed predefined performance thresholds.
- A failure penalty is applied if the episode terminates prematurely due to infeasibility or dead ends.

The total episode reward is computed as the sum of all step-wise rewards and any applicable terminal adjustments (Equation (27)):

$$R_{episode} = \sum_{t=0}^{F^{final}} r_t + r^{terminal} \quad (27)$$

Reward Normalization and Curriculum Scheduling

To maintain learning stability, each reward component is normalized to the $[-1,1]$ range using min-max scaling derived from empirical bounds collected during initial random exploration. Additionally, curriculum scheduling is employed by adjusting the relative weights λ_k across training epochs: early training favors exploration and satisfaction, while later training emphasizes feasibility and efficiency.

This reward function architecture enables the DRL agent to internalize the trade-offs between cultural value, sustainability, efficiency, and personalization. It provides continuous and interpretable feedback signals aligned with both short-term transitions and long-term route viability, ensuring convergence toward practical, policy-compliant routing strategies within the digital twin simulation environment.

4.1.4. Policy Network Architecture and Training Procedure

To learn an effective routing policy under multi-objective and constraint-aware conditions, we implement an actor-critic deep reinforcement learning model using the Proximal Policy Optimization (PPO) algorithm. PPO is selected for its balance between sample efficiency and stability in constrained, high-dimensional action spaces, making it well-suited for multi-step route planning in large-scale spatial graphs. This section details the neural network architecture, training configuration, and policy optimization protocol.

Neural Network Architecture

The policy is parameterized by a shared actor-critic neural architecture composed of an encoder backbone and two parallel output heads: one for action selection (the actor) and one for state-value estimation (the critic).

Input Layer

Each input state s_t is represented as a concatenated feature vector that encodes the routing context at time step t . The input captures not only the agent's current location but also cumulative performance metrics, route history, preference alignment, and feasibility-relevant status. This

structured input ensures that the agent's decision-making process remains both personalized and constraint-aware.

The state vector consists of the following subcomponents:

1. Current POI Position: A one-hot encoded vector of size $|N|$, where the index corresponding to the current POI x_t is set to 1, and all others to 0.
Example (for $|N| = 8$ and current POI = node 3): $POI_position = [0, 0, 0, 1, 0, 0, 0, 0]$
2. Cumulative Travel Metrics: Three normalized scalar values representing total time used, total emissions, and total walking distance up to step t .
Example: $Cumulative_metrics = [0.32, 0.15, 0.41]$
3. POI Visitation History: A binary vector of size $|N|$ indicating whether each POI has been visited (1) or not (0).
Example (nodes 0, 2, and 3 visited): $Visited_flags = [1, 0, 1, 1, 0, 0, 0, 0]$
4. Category Quota Progress: A 3-dimensional integer vector tracking how many POIs from each mandatory category (e.g., house, restaurant, museum) have been visited.

Example (2 houses, 1 restaurant, 0 museums visited): $Quota_progress = [2, 1, 0]$

5. Group Preference Vector: A continuous vector $p_g \in [0,1]^{|N|}$ representing the interest level of the tourist group for each POI. This captures personalized bias toward specific nodes.

Example (simplified preference weights): $p_g = [0.1, 0.8, 0.6, 0.9, 0.2, 0.1, 0.0, 0.3]$

6. Normalized Remaining Budgets: A 3-dimensional vector indicating the proportion of remaining time, emissions, and distance budget (all normalized to $[0,1]$).
Example (70% time left, 90% emissions left, 60% distance left):
 $Remaining_budgets = [0.70, 0.90, 0.60]$

The final input vector $f_t \in \mathbb{R}^d$ at step t is the concatenation of all the above components:

$$f_t = \text{concat}(POI_position, Cumulative_metrics, Visited_flags, Quota_progress, p_g, Remaining_budgets)$$

For example, in a simplified case with $|N|=8$, the full input vector has:

- 8 dimensions for current POI position,
- 3 for cumulative metrics,
- 8 for visitation flags,
- 3 for quota progress,
- 8 for preferences,
- and 3 for normalized budgets—totaling 33 dimensions.

All continuous features are normalized using min-max scaling to the $[0,1]$ range. This ensures consistent gradient behavior during training and prevents any feature group from dominating the learning dynamics due to scale discrepancies.

- Encoder Backbone: The shared backbone consists of two fully connected layers of size 256 and 128 neurons, respectively, each followed by ReLU activations and dropout layers (rate = 0.2) for regularization. This subnetwork extracts high-level routing context features from the raw state input.
- Actor Head: The actor network outputs a probability distribution over the current admissible action set $A(s_t)$. To ensure constraint compliance, infeasible actions are masked before Softmax normalization (Equation (28)):

$$\pi_{\theta}(a_t | s_t) = \text{softmax}(W_a \cdot h_t + b_a) \odot m_t \quad (28)$$

Where h_t is the encoder output, m_t is the feasibility mask, and \odot denotes elementwise multiplication.

- Critic Head: The critic estimates the state-value function $V_{\psi}(s_t)$ using a separate linear layer over the shared encoding (Equation (29)):

$$V_{\psi}(s_t) = W_c \cdot h_t + b_c, \quad (29)$$

Enabling bootstrapped learning and advantage estimation during policy updates.
Training Configuration

The PPO algorithm is trained in mini-batches using the following hyperparameters, empirically tuned via grid search (Table 1)

Table 1. Hyperparameter settings for PPO training.

Parameter	Value
Discount factor (γ)	0.99
GAE lambda	0.95
Clipping coefficient (ϵ)	0.2
Entropy regularization weight	0.01
Learning rate	3×10^{-4}
Batch size	128
Update epochs per batch	4
Optimizer	Adam

Each episode consists of a complete route construction process, beginning at the designated start POI and terminating upon reaching the end POI or exhausting all feasible actions. During each episode, trajectories (s_t, a_t, r_t, s_{t+1}) are collected and used to compute Generalized Advantage Estimation (GAE) (Equation (30)):

$$\hat{A}_t = \sum_{l=0}^{T-t} (\gamma \lambda)^l \delta_{t+l}, \delta_t = r_t + \gamma V_{\psi}(s_{t+1}) - V_{\psi}(s_t) \quad (30)$$

This allows for low-variance, temporally smoothed policy gradient updates, improving sample efficiency.

Constraint-Aware Sampling and Masking

To avoid invalid learning signals and ensure feasible exploration, we apply dynamic masking of the action space at every decision step. Infeasible actions are assigned a probability of zero during both training and inference, ensuring that the agent learns only over admissible transitions. This constraint-aware sampling preserves the integrity of the training distribution and accelerates convergence by pruning unproductive policy paths.

Curriculum Learning and Policy Stabilization

To support robust convergence, we implement a curriculum learning strategy in which training begins with relaxed constraints (e.g., extended time budgets or softened preference thresholds) and progressively tightens them across training epochs. This staged learning approach allows the policy to initially focus on exploratory coverage and gradually shift toward constraint satisfaction and objective trade-offs.

Policy stabilization is further enforced via:

- Early stopping based on validation reward trends,
- Reward normalization based on moving average bounds, and
- Periodic reseeding to encourage diversity when learning plateaus.

The training process continues until convergence is observed in the moving average of cumulative episode rewards and constraint satisfaction rates. Upon convergence, the policy is saved for inference and used to generate the initial population of feasible routes described in Section 4.1.5.

4.1.5. Initial Solution Set Extraction

Following the successful training of the DRL policy, the agent is deployed in inference mode to construct a population of NP initial solutions, denoted by the set \mathcal{P}_0 . This population constitutes the initial decision space for the hybrid optimization framework and must be both feasible and diverse to support robust downstream refinement using IMVO and GAN-based operators (Section 4.2).

DRL-Guided Solution Sampling

The trained policy $\pi_\theta(a_t | s_t)$ is executed over a series of independent episodes. Each episode simulates a complete route planning process from a fixed start POI to a designated end POI. During inference, the policy selects feasible actions based on constraint-filtered probability distributions generated at each step. To ensure adequate diversity in the resulting solutions, a dual-mode inference strategy is used:

- Greedy decoding (maximum likelihood) is employed in 40% of the runs to promote high-reward exploitation.
- Stochastic sampling is used in 60% of the runs to encourage exploratory variation and structure diversity.

Each run yields a complete candidate route \mathcal{R}_k , which is a sequential arrangement of POIs visited by the agent under dynamic constraint monitoring.

Feasibility Enforcement and Route Validation

After generation, each candidate route is subjected to a strict constraint validation filter to ensure admissibility. A route \mathcal{R}_k is retained in the initial population if and only if it satisfies the following conditions:

- The total cumulative travel time $\sum_{(i,j) \in \mathcal{R}_k} \tau_{ij}$ does not exceed the global time budget B^{time} ,
- Emissions and walking distance remain within upper bounds B^{CO_2} and B^{dist} ,
- Required category quotas $Q^{house}, Q^{rest}, Q^{muse}$ are fulfilled,
- All POIs in the route are distinct (no repetition), and the sequence includes the designated end node t ,
- All group members achieve their required minimum preference satisfaction Rg .

Routes failing any of the above are discarded. The process repeats until a minimum of $|\mathcal{P}_0| = NP$ distinct, feasible solutions are collected.

Structure of the Initial Solution Set

Each solution $\mathcal{R}_k \in \mathcal{P}_0$ is stored with a structured representation that includes:

- POI sequence: $x_k = \langle x_1, x_2, \dots, x_T \rangle$
- Arc decision matrix $X_k = [x_{ij}]$: binary indicators of arc traversals,
- Arrival time vector $\mathcal{H}_k = [\mathcal{H}_{x_1}, \dots, \mathcal{H}_{x_T}]$
- Quota count vector $q_k \in \mathbb{Z}^3_{\geq 0}$
- Group satisfaction scores $S_k^g = \sum_{i \in x_k} p_{ig}$ for each $g \in G$,
- Objective function values: F_1 (heritage), F_2 (travel time), F_3 (emissions), F_4 (smoothness), F_5 (group satisfaction).

This multi-perspective structure ensures that all relevant dimensions of routing performance are preserved for further refinement and comparison in later stages.

Diversity Maintenance and Quality Screening

To avoid premature convergence in the refinement phase, we implement a diversity control mechanism during extraction. A candidate solution is only added to \mathcal{P}_0 if it meets a minimum dissimilarity threshold from all previously accepted solutions, measured via:

- Levenshtein distance between POI sequences,
- Hamming distance between visitation vectors,
- or Pareto-dominance novelty in objective space.

This ensures that \mathcal{P}_0 spans a broad region of the feasible solution space, offering useful variability for population-based metaheuristic operators and GAN-driven perturbation mechanisms in the next stage.

Representative Examples

Table 2 presents three representative solutions from \mathcal{P}_0 , each highlighting a different trade-off across heritage coverage, travel time, emissions, congestion, and group satisfaction. These examples illustrate the ability of the DRL agent to generate diverse, feasible, and personalized routes.

Table 2. Representative initial solutions generated by DRL policy.

Rout e ID	POI Sequence (Abbreviated)	Trave l Time (min)	Distanc e (km)	Emission s (kgCO ₂)	Congestio n Index	Satisfactio n Avg.	Smoothnes s Score	Feasibilit y
R1	Wat Luang → Old Market → Museum → Rest. A → House 2 → City Gate	78	3.2	0.52	0.21	0.82	6.2	✓
R2	Heritage Gate → Art Gallery → Candle Museum → Food Court → Post Office	65	2.5	0.44	0.32	0.81	5.1	✓
R3	Clock Tower → Workshop → House 3 → Café → Walking Street	42	1.9	0.31	0.16	0.63	4.8	✓

These solutions not only satisfy all problem constraints but also represent distinct styles of tour planning—ranging from culture-intensive routes to time-efficient or satisfaction-prioritized itineraries—thereby offering a rich starting point for multi-objective refinement in subsequent stages.

4.2. Phase II: IMVO-GAN Refinement and Local Search

After generating the initial solution set \mathcal{P}_0 using the DRL-guided construction phase, the second phase focuses on iterative refinement and exploitation of this population through a hybrid algorithm that combines the Improved Multi-Verse Optimizer (IMVO) with a Wasserstein GAN-based local search module. The purpose of this phase is to (a) improve the convergence of the solution set toward the true Pareto front, (b) enhance structural diversity to avoid local stagnation, and (c) exploit learned generative distributions of high-performing routes to inform adaptive perturbation strategies.

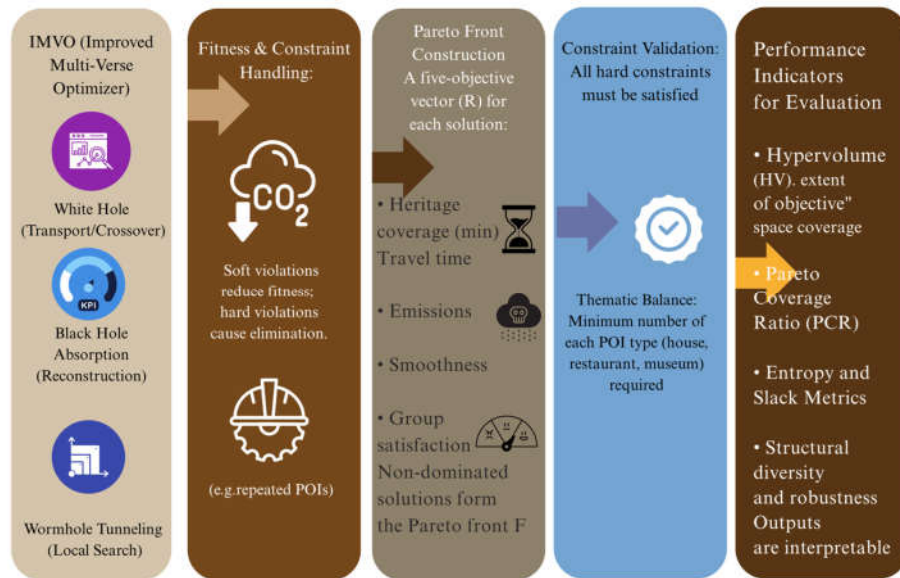


Figure 4. IMVO-GAN-based refinement and multi-objective local search process.

Figure 4, illustrates Phase II of the proposed framework, where the Improved Multi-Verse Optimizer (IMVO) and a Generative Adversarial Network (GAN) module collaboratively refine the initial route population generated by the DRL agent. Each candidate route is treated as a "universe" and evolves through crossover, reconstruction, and local search operators inspired by astrophysical phenomena. Fitness scores are recalibrated with soft and hard constraint penalties, and Pareto dominance is computed across five objectives: heritage coverage, travel time, emissions, route smoothness, and group satisfaction. Final routes undergo thematic validation and are ranked using the TOPSIS method to support balanced, user-aligned decision-making.

4.2.1. Improved Multi-Verse Optimizer (IMVO) for Constraint-Aware Evolution

Following the construction of the initial solution population \mathcal{P}_0 via DRL, the refinement phase begins with the application of an Improved Multi-Verse Optimizer (IMVO) to guide this population toward an efficient, well-distributed Pareto front. The IMVO is selected for its global search capability and inherent population diversity preservation, which are particularly well-suited to the personalized, multi-objective nature of heritage-tour routing problems. To adapt it to our domain, we introduce a set of constraint-aware enhancements and permutation-based evolutionary operators that accommodate routing-specific feasibility requirements.

Universe Representation and Fitness Scaling

Each candidate solution $\mathcal{R}_k \in \mathcal{P}_0$ is encoded as a universe, modeled as an ordered sequence of POIs: $x_k = \langle x_1, x_2, \dots, x_T \rangle$ where the route always begins at the start node s and ends at the terminal POI t , with all intermediate POIs representing personalized, feasible selections. These universes retain rich metadata from Section 4.1.5, including category coverage, group satisfaction scores, and route-based objective values. Each universe's inflation rate—analogue to its fitness—is computed based on a normalized, dominance-informed composite of the five objective functions. A Pareto ranking procedure ensures that non-dominated solutions receive proportionally higher inflation rates, enabling elite solutions to propagate influence during search.

Evolutionary Mechanics and Examples

The refinement process proceeds through three primary operators: white hole transport, black hole absorption, and wormhole tunneling. These mechanisms are explicitly redesigned to support routing-specific permutation structures and real-world constraints.

- White Hole Transport (Crossover)

This operator allows high-quality universes to share structural patterns with inferior ones via constrained segment exchange. For instance, consider two solutions (Equations (31) and (32)):

$$\mathcal{R}_{elite} = \langle A, B, C, D, E, F \rangle \quad (31)$$

$$\mathcal{R}_{weak} = \langle G, H, I, J, K, L \rangle \quad (32)$$

A segment $[C, D]$ from \mathcal{R}_{elite} may replace segment $[I, J]$ in \mathcal{R}_{weak} , producing a hybrid route (Equation (33)):

$$\mathcal{R}_{new} = \langle G, H, C, D, K, L \rangle \quad (33)$$

To maintain validity, a partial mapping crossover (PMX) is applied, followed by a repair heuristic to eliminate duplicate POIs or unmet category quotas. This ensures the offspring remains a feasible, high-potential universe.

- Black Hole Absorption (Replacement)

Universes with persistently low inflation scores are replaced using elite-guided regeneration. For example, if a poorly performing solution lacks quota fulfillment or exhibits excessive emissions, it may be discarded and reconstructed by selecting POI subsequences from top-performing universes (e.g., $\langle A, B, C \rangle$) and inserting new POIs from underrepresented categories, such as a local museum M_1 or restaurant R_2 , yielding (Equation (34)):

$$\mathcal{R}_{new} = \langle A, B, C, R_2, M_1, t \rangle \quad (34)$$

The regenerated route is only admitted if it passes all feasibility constraints.

- Wormhole Tunneling (Local Perturbation)

This operator performs lightweight, heuristic-guided local adjustments to individual universes. Examples include:

- POI swaps: Exchanging positions of two non-adjacent POIs, such as swapping C and E in $\langle A, B, C, D, E \rangle$, resulting in $\langle A, B, E, D, C \rangle$.
- Segment reversals: Reversing $[B, C, D]$ in the same route produces, which may improve angular smoothness.
- Conditional insertions: Inserting a highly preferred POI P^* before the end node if remaining time budget allows, such as: $\langle A, B, C, P^*, t \rangle$

These operations respect hard constraints and leverage domain-specific heuristics (e.g., angle continuity, residual budget ratios) to guide improvements.

Constraint-Adaptive Fitness Adjustment

To ensure fairness in selection and robustness under constraints, we adopt a penalty-based fitness recalibration. Let \tilde{I}_k denote the adjusted inflation rate of universe \mathcal{R}_k , computed as (Equation (35)):

$$\tilde{I}_k = I_k - \alpha_1 \Delta_t^k - \alpha_2 \Delta_e^k - \alpha_3 \Delta_q^k - \alpha_4 \Delta_s^k \quad (35)$$

where:

- Δ_t^k , Δ_e^k , Δ_q^k , and Δ_s^k represent soft violations in time, emissions, category quotas, and group satisfaction, respectively;
- α_i are empirically tuned penalty coefficients.

Solutions that violate hard constraints, such as revisiting a POI or exceeding the maximum tour duration, are eliminated outright and excluded from the next generation.

Evolution Control and Termination

The IMVO refinement loop continues for G generations or until convergence criteria are met—defined by stagnation of the elite archive or saturation in the diversity index across the population. Each generation:

- Maintains an elitist archive of non-dominated solutions,
- Preserves structural variety via POI visitation entropy and angular transition variance,
- Dynamically adjusts selection pressure based on front dispersion and constraint stress levels.

4.3. Solution Evaluation and Pareto Front Construction

Upon completion of the IMVO–GAN refinement phase, the resulting solution archive \mathcal{P}^* comprises a diverse set of feasible and structurally distinct routing plans. To extract the most informative and trade-off-optimal solutions from this archive, we perform a non-dominated filtering process that yields a single global Pareto front, denoted $\mathcal{F} \subseteq \mathcal{P}^*$. This front contains only those solutions that are non-inferior with respect to all others across the five conflicting objective functions defined in the problem formulation.

Each candidate route $\mathcal{R}_k \in \mathcal{P}^*$ is mapped to a five-dimensional objective vector (Equation (36):

$$\phi(\mathcal{R}_k) = (\phi_1, \phi_2, \phi_3, \phi_4, \phi_5) \quad (36)$$

where:

- ϕ_1 : heritage coverage adjusted for crowd congestion (maximize),
- ϕ_2 : total travel time (minimize),
- ϕ_3 : cumulative carbon emissions (minimize),
- ϕ_4 : angular deviation representing route smoothness (minimize),

ϕ_5 : group preference satisfaction (maximize).

A solution $\mathcal{R}_a \in \mathcal{P}^*$ is said to dominate another solution $\mathcal{R}_b \in \mathcal{P}^*$, denoted $\mathcal{R}_a < \mathcal{R}_b$, if (Equation (37):

$$\begin{cases} \phi_1(\mathcal{R}_a) \geq \phi_1(\mathcal{R}_b) \\ \phi_2(\mathcal{R}_a) \leq \phi_2(\mathcal{R}_b) \\ \phi_3(\mathcal{R}_a) \leq \phi_3(\mathcal{R}_b) \text{ and } \exists k \in \{1,2,3,4,5\} \text{ such that } \phi_k(\mathcal{R}_a) < \phi_k(\mathcal{R}_b) \\ \phi_4(\mathcal{R}_a) \leq \phi_4(\mathcal{R}_b) \\ \phi_5(\mathcal{R}_a) \geq \phi_5(\mathcal{R}_b) \end{cases} \quad (37)$$

The resulting Pareto set \mathcal{F} thus contains all solutions in \mathcal{P}^* that are non-dominated with respect to the full set of objective functions.

Each route in \mathcal{F} is then subjected to a final constraint validation to ensure field deployability. The feasibility conditions include (Equations (38) and (39)):

$$\sum_{(i,j) \in R} \tau_{ij} + \sum_{i \in R} \delta_i \leq B^{time}, \sum_{(i,j) \in R} \varepsilon_{ij} \leq B^{CO_2} \quad (38)$$

$$x_1 = s, x_T = t, \sum_{i \in R} p_{ig} \geq R_g \quad \forall g \in G \quad (39)$$

To ensure thematic and experiential balance within each personalized tour, category quota constraints are enforced. Specifically, the itinerary must include a minimum number of points of interest (POIs) from each predefined category—historic houses, restaurants, and museums. These constraints are formalized as follows (Equation (40)):

$$\sum_{i \in \mathcal{R} \cap N_{house}} y_i \geq Q^{house}, \sum_{i \in \mathcal{R} \cap N_{rest}} y_i \geq Q^{rest} \geq Q, \sum_{i \in \mathcal{R} \cap N_{muse}} y_i \geq Q^{muse} \quad (40)$$

where $y_i \in \{0,1\}$ indicates whether POI i is included in the constructed route \mathcal{R} , and N_{house} , N_{rest} , and N_{muse} denote the subsets of POIs categorized as historic houses, restaurants, and museums,

respectively. The thresholds $Q^{house}, Q^{rest}, Q^{muse} \in \mathbb{Z} \geq 0$ are defined based on policy or user preferences.

Only routes that satisfy all the above constraints are retained in \mathcal{F} , ensuring that all proposed recommendations are not only optimal in the multi-objective sense, but also admissible under real-world operational requirements.

To facilitate compromise solution selection from the final Pareto front \mathcal{F} , a multi-criteria decision analysis (MCDA) layer is applied using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). Given that the problem considers five conflicting objectives—heritage coverage with congestion discount, travel time, carbon emissions, route smoothness, and group preference satisfaction—each solution $\mathcal{R}_k \in \mathcal{F}$ s mapped to a normalized five-dimensional performance vector (Equation (41)):

$$z_k = (z_{k1}, z_{k2}, z_{k3}, z_{k4}, z_{k5}) \quad (41)$$

Corresponding respectively to the scaled values of objectives ϕ_1 through ϕ_5 . Normalization is performed via min-max scaling to ensure comparability across metrics with heterogeneous units and directions (i.e., some to be maximized and others minimized).

The ideal solution vector z^+ and the anti-ideal solution vector z^- are computed as (Equations (42) and (43)):

$$z^+ = (\max_k z_{k1}, \min_k z_{k2}, \min_k z_{k3}, \min_k z_{k4}, \max_k z_{k5}) \quad (42)$$

$$z^- = (\min_k z_{k1}, \max_k z_{k2}, \max_k z_{k3}, \max_k z_{k4}, \min_k z_{k5}) \quad (43)$$

Reflecting the ideal preference for each objective (maximize ϕ_1, ϕ_5 ; minimize ϕ_2, ϕ_3, ϕ_4).

Let $w = (w_1, w_2, w_3, w_4, w_5)$ be a stakeholder-defined weight vector, where each $w_i \in [0,1]$ and $\sum_{i=1}^5 w_i = 1$. The Euclidean distance of each solution to the ideal and anti-ideal reference points is computed as (Equation (44)):

$$D_k^+ = \sqrt{\sum_{i=1}^5 w_i (z_{ki} - z_i^+)^2}, D_k^- = \sqrt{\sum_{i=1}^5 w_i (z_{ki} - z_i^-)^2} \quad (44)$$

The relative closeness score $C_k \in [0,1]$ is then calculated for each $\mathcal{R}_k \in \mathcal{F}$ as (Equation (45)):

$$C_k = \frac{D_k^-}{D_k^- + D_k^+} \quad (45)$$

with higher values indicating greater similarity to the ideal solution and thus higher desirability under the given preference structure.

This extended TOPSIS formulation provides a robust mechanism for selecting one or more compromise routes that balance heritage value, sustainability, efficiency, smoothness, and group satisfaction in alignment with policy or user priorities. It enables interpretable ranking of non-dominated solutions without collapsing the problem into a single scalarized objective.

To assess the quality of the Pareto front \mathcal{F} , we compute performance indicators such as hypervolume (HV), which measures the portion of objective space dominated by the front, and the Pareto Coverage Ratio (PCR) in comparison to benchmark algorithms. In addition, structural entropy is reported to evaluate route diversity, and average constraint slackness is used to quantify robustness against perturbations in travel time, emissions, or group preferences.

Algorithm 1 outlines the procedural flow of the Deep Reinforcement Learning-based solution construction phase, which serves as the foundation of the proposed hybrid optimization framework. This phase integrates a GIS-driven digital twin of Warin Chamrap's old town with a Proximal Policy Optimization (PPO) agent to generate an initial set of diverse and constraint-compliant tour itineraries. Through iterative interaction with a constraint-aware Markov Decision Process (MDP), the agent learns to balance heritage coverage, travel efficiency, sustainability, and group satisfaction.

The resulting solution population, denoted \mathcal{P}_0 , forms the input for subsequent metaheuristic refinement.

Algorithm 1. DRL-based personalized tour construction in digital twin.

Algorithm DRL-Based Personalized Tour Construction in Digital Twin
 Input: Digital Twin Graph $G(N, A)$, preference profiles p_g for group G , constraints B^{time} , B^{CO_2} , Q^{category} , terminal node t
 Output: Initial solution population \mathcal{P}_0 of feasible tour routes
 1: Initialize digital twin environment with:
 a. POI metadata: category, dwell time, heritage score, operating hours
 b. Arc attributes: distance d_{ij} , time τ_{ij} , emissions ϵ_{ij} , angle ϕ_{ij}
 c. Real-time congestion heatmaps $\kappa_i(t)$
 d. Group preference vector p_g
 2: Define Markov Decision Process $M = \langle S, A, P, R, \gamma, T_{\max} \rangle$
 a. State $s_t = [\text{current POI } x_t, \text{visitation vector } v_t, \text{cumulative metrics } \Delta, \text{quota } q_t, \text{preference } p_g]$
 b. Action a_t : next POI j in filtered feasible set $A(s_t)$
 c. Transition: update state s_{t+1} with cumulative Δ metrics and POI j
 d. Reward r_t = weighted combination of 7 sub-rewards (heritage, travel, emission, quota, satisfaction, smoothness, penalty)
 3: Design policy network $\pi_{\theta}(a_t | s_t)$ using actor-critic PPO with:
 a. Feature vector $f_t = \text{concat}(x_t, \Delta \text{ metrics}, v_t, q_t, p_g, \text{normalized budget})$
 b. Actor head: action logits over $A(s_t)$ with softmax masking
 c. Critic head: state-value $V_{\psi}(s_t)$
 d. Train using PPO: GAE, reward normalization, entropy regularization
 4: Curriculum learning:
 a. Begin with relaxed constraints ($B^{\text{time}}_{\text{init}}, Q^{\text{category}}_{\text{init}})$
 b. Tighten every epoch toward real-world thresholds
 c. Use early stopping when policy stabilizes
 5: Inference mode:
 a. For $i = 1$ to NP do:
 i. Sample full episode with π_{θ} using greedy (40%) or stochastic (60%) decoding
 ii. Construct candidate route R_k with full POI sequence and trajectory metadata
 iii. Validate R_k under hard constraints:
 - Time, Emission, Distance Budget
 - Category Quotas
 - Terminal POI match ($x_T = t$)
 - Group satisfaction minimum $\sum_i p_{ig} \geq R_g, \forall g$
 iv. If R_k is valid and diverse, add to \mathcal{P}_0
 b. Until $|\mathcal{P}_0| = NP$
 6: Return \mathcal{P}_0 as initial feasible, diverse, and personalized population

4.4. Case Study: Heritage-Tour Planning in Warin Chamrap's Old Town

To evaluate the applicability and performance of the proposed DRL-IMVO-GAN framework in real-world conditions, we implement a detailed case study based on the old town of Warin Chamrap District, Ubon Ratchathani, Thailand. This area is renowned for its rich cultural heritage and walkable urban layout, making it ideal for personalized heritage tourism planning. The study leverages both empirical spatial data and simulated visitor profiles to construct, optimize, and validate realistic tour itineraries.

4.4.1. Study Area, POI Inventory, and Infrastructure Modeling

The case study area spans a geographically bounded precinct in central Warin Chamrap, encompassing a total of 92 points of interest (POIs). These include 45 historic houses, 40 restaurants, and 7 museums—each categorized according to municipal cultural maps and verified through on-site fieldwork. The pedestrian infrastructure is modeled as a directed graph with 248 arcs, extracted from OpenStreetMap and refined through satellite-aided mapping and local surveys. Arc-level attributes include walking time (τ_{ij}), distance (d_{ij}), angular orientation (ϕ_{ij}), and estimated carbon

emissions (ϵ_{ij}) for each link. Congestion heatmaps are integrated using mobile device density patterns and CCTV data, forming the basis of the congestion index (κ_i) at each POI.

Tourists are assumed to move at an average walking speed of 5 km/h, with the total allowable time for each itinerary capped at 150 minutes. Group preference vectors are synthesized from clustered historical data and injected into the digital twin environment to simulate diversified interest structures across different user groups.

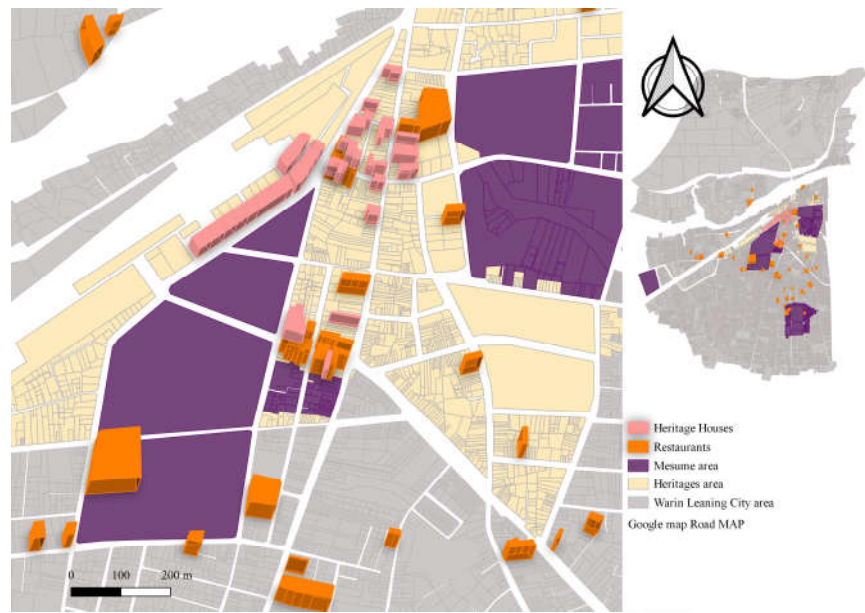


Figure 5. Digitally simulated map of Warin Chamrap’s heritage tourism POIs.

Figure 5 presents a spatial simulation of the old town district of Warin Chamrap, Ubon Ratchathani, overlaid with synthetic yet geographically plausible placements of 92 points of interest (POIs), comprising 45 historic houses, 40 restaurants, and 7 museums. These POIs are distributed along the urban road network and aligned with existing geographic features extracted from the district’s official planning maps. The map serves as the foundational spatial layer within the digital twin environment used in this study, enabling the DRL agent to interact with a structurally accurate representation of the cultural and pedestrian landscape. The realism and fidelity of this layout ensure the subsequent routing decisions and evaluations are grounded in a deployable urban context.

4.4.2. Parameter Configuration and Constraint Settings

To simulate realistic tour planning behavior and enforce practical viability, a comprehensive set of parameters is configured. These include constraints on travel time, emissions, distance, and congestion exposure, as well as category-specific visitation quotas. Group satisfaction thresholds and personalized time budgets further align with behavioral realism. The parameters are calibrated using expert input from cultural planners and sustainability practitioners.

Table 3 presents the complete list of symbols, units, and assigned values or ranges used throughout the case study, serving as a reproducibility reference for future benchmarking and extension studies.

Table 3. Detailed parameter specification for personalized heritage route optimization.

Symbol	Unit	Values or Range
h_i	score	[1.0 – 10.0]
δ_i	minutes	[10 – 45]
τ_{ij}	minutes	[1 – 10]
d_{ij}	meters	[50 – 550]
ϵ_{ij}	kgCO ₂	[0.05 – 0.15]
κ_i	index [0–1]	[0.10 – 0.85]

Symbol	Unit	Values or Range
ϕ_{ij}	degrees	$[0^{\circ} - 360^{\circ}]$
B^{time}	minutes	150
B^{CO_2}	kgCO ₂	1.8
B^{cong}	index	6.0
B^{dist}	km	6.0
$Q^{house}, Q^{rest}, Q^{muse}$	count	3, 2, 1
K^{max}	POIs	12
M	constant	10^6
S_{ia}	score [0–1]	[0.0 – 1.0]
B_a^{time}	minutes	[100 – 160]
R_a	score	[2.0 – 4.5]

These configurations lay the foundation for route construction, evaluation, and refinement using the hybrid DRL–IMVO–GAN algorithm described in Sections 4.1 through 4.3. The simulation results, Pareto-optimal fronts, and field validation are reported in Section 5. Table 4 summarizes five illustrative group profiles, which were selected to span a broad range of travel intentions and constraints.

Table 4. Representative tourist group profiles used in the case study.

Group ID	Group Type	Time Budget (min)	Sustainability Awareness	Heritage Interest	Food Interest	Museum Interest	Min. Satisfaction R_a
G ₁	Cultural Enthusiasts	150	Moderate	High (0.85)	Low (0.25)	High (0.80)	3.8
G ₂	Family with Children	120	Low	Moderate (0.60)	High (0.90)	Moderate (0.50)	3.2
G ₃	Green Explorers	140	High	Moderate (0.70)	Moderate (0.60)	Low (0.30)	3.6
G ₄	Senior Travelers	100	Medium	High (0.75)	Moderate (0.55)	High (0.70)	3.5
G ₅	Casual Walkers	90	Low	Low (0.35)	High (0.85)	Low (0.20)	2.5

To illustrate the functional diversity and metadata representation of Points of Interest (POIs) in the digital twin environment, five sample attractions are described in Table 5. These examples span the core categories of heritage houses, local culinary venues, and museums, each annotated with context-sensitive parameters such as heritage value, average dwell time, and real-time congestion index. This structured metadata directly informs the DRL reward architecture and feasibility masking mechanisms discussed in Section 4.1.

Table 5. Sample points of interest in Warin Chamrap digital twin environment.

POI Name	Category	Heritage Score h_i	Avg. Dwell Time δ_i (min)	Est. Congestion κ_i	Description
Wat Luang Old Temple	Historic House	9.5	30	0.45	A centuries-old wooden temple known for its teak carvings and community rituals.
Warin Walking Street Market	Restaurant	4.0	25	0.65	A bustling street food zone with regional Isan cuisine and weekend night fairs.
Phaya Thian Candle Museum	Museum	8.8	40	0.30	A cultural museum featuring the art of Ubon’s candle-carving traditions.

House No. 89 Cultural Home	Historic House	7.2	20	0.20	A private heritage residence showcasing colonial-era architecture and oral history exhibits.
Nong Bua Riverside Café	Restaurant	3.5	35	0.50	A scenic coffee shop along the river, favored by families and cyclists.

4.4.3. Simulation Protocol

The Deep Reinforcement Learning (DRL) agent is trained over 2,000 episodes using curriculum learning, with rewards and constraints gradually tightened to reflect operational scenarios. Proximal Policy Optimization (PPO) is deployed using a two-headed actor-critic architecture, with a learning rate of 3×10^{-4} , entropy regularization of 0.01, and Generalized Advantage Estimation (GAE) for stable convergence.

Upon convergence, the trained policy is used to generate an initial population \mathcal{P}_0 of 100 distinct and feasible itineraries, which serve as inputs to the IMVO-GAN refinement phase. The IMVO is configured with a population size of 100, a maximum of 150 generations, and elite preservation of the top 10% of solutions per generation. The GAN module is retrained every 20 generations on the evolving elite archive.

4.4.4. Evaluation Metrics

To comprehensively assess the effectiveness of the proposed DRL-IMVO-GAN optimization framework for personalized heritage-tourism planning, both quantitative and qualitative evaluation criteria are adopted. These metrics are designed to capture solution quality across multiple dimensions: convergence, diversity, constraint robustness, and real-world interpretability.

(a) Hypervolume (HV)

The Hypervolume (HV) metric quantifies the volume in the objective space dominated by the non-dominated Pareto front \mathcal{F} , bounded by a reference nadir point z^{ref} . For a set of normalized objective vectors $\{z_k\}_{k=1}^K \subseteq F$, where each $z_k = (z_{k1}, z_{k2}, \dots, z_{k5})$, the HV is given by using (Equation (45)):

$$HV(\mathcal{F}) = Vol(\cup_{k=1}^K [z_k, z^{ref}]) \quad (45)$$

This metric reflects the extent to which the proposed method dominates the multi-objective space, favoring both optimality and coverage.

(b) Pareto Coverage Ratio (PCR)

The Pareto Coverage Ratio (PCR) compares the number of Pareto-optimal solutions generated by the proposed framework \mathcal{F}_{ours} that dominate solutions from a benchmark method \mathcal{F}_{base} ((Equation (46)):

$$PCR = \frac{|\{R \in \mathcal{F}_{ours} \mid \exists R' \in \mathcal{F}_{base} \text{ such that } R \succ R'\}|}{|\mathcal{F}_{base}|} \quad (46)$$

(c) POI Entropy (Route Structural Diversity)

To measure route diversity across the solution population, we compute POI entropy based on the distribution of visited POIs. Let p_i denote the proportion of solutions in which POI $i \in \mathcal{N}$ appears. The entropy \mathcal{H} is then calculated as ((Equation (47)):

$$\mathcal{H} = -\sum_{i \in \mathcal{N}} p_i \log_2 p_i \quad (47)$$

A higher entropy indicates greater structural diversity across itineraries, which enhances personalization and reduces solution redundancy.

(d) Feasibility Slack Margin

To quantify the operational headroom of each solution in \mathcal{F} , we compute slack margins for key constraints aligned with the five-objective formulation. For each feasible route $\mathcal{R}_k \in \mathcal{F}$, the following slacks are defined (Equations (48) to (52)):

$$\text{Time budget: } \Delta_k^{time} = B^{time} - (\sum_{(i,j) \in \mathcal{R}_k} \tau_{ij} + \sum_{(i,j) \in \mathcal{R}_k} \delta_i) \quad (48)$$

$$\text{Emissions: } \Delta_k^{CO_2} = B^{CO_2} - \sum_{(i,j) \in \mathcal{R}_k} \epsilon_{ij} \quad (49)$$

$$\text{Category quota: } \Delta_k^{CAT} = \min_{c \in \{house, rest, muse\}} (\sum_{i \in \mathcal{R}_k \cap N_c} 1 - Q^c) \quad (50)$$

$$\text{Congestion slack: } \Delta_k^{cong} = B^{cong} - \sum_{(i) \in \mathcal{R}_k} \kappa_i \quad (51)$$

$$\text{Preference satisfaction slack (per group } g \in G): \Delta_{kg}^{pref} = \sum_{(i) \in \mathcal{R}_k} s_{ig} - R_g \quad (52)$$

Each Δ value captures the residual capacity before violating its corresponding constraint. We report the average slack per dimension across the Pareto set to indicate solution resilience and robustness margin. Together, these metrics ensure that all five objectives are not only optimized but also operate within reliable and deployable feasibility envelopes.

4.5. Compared Methods

To evaluate the performance of the proposed DRL-IMVO-GAN algorithm in solving the Personalized Group Trip Design Problem with Multi-Objective Optimization Criteria, we adapted and reprogrammed seven recent state-of-the-art metaheuristic algorithms. Each was implemented to accommodate our problem's characteristics, including personalized preference alignment, group cohesion, temporal constraints, and sustainability objectives.

The Hybrid Genetic Algorithm with Multi-Neighborhood Search, originally introduced for reconfigurable manufacturing cell scheduling, combines evolutionary operators with local intensification schemes [45]. When adapted to group trip planning, the method effectively explored diverse itinerary configurations but exhibited sensitivity to preference diversity in larger travel groups.

The Discrete Particle Swarm Optimization with Deep Reinforcement Learning, designed for quantum circuit mapping, was reformulated to manage group itinerary encoding as a discrete combinatorial problem guided by learned policies [46]. While this hybridization enabled learning-based route refinement, it showed instability when exposed to conflicting group goals and tight budget constraints.

Dual-Strategy Ant Colony Optimization, originally applied to ship pipeline routing, was adapted to evolve itineraries using active behavioral adjustments and passive environmental adaptation [47]. Though it captured route coherence and group constraints well, its convergence speed was lower compared to population-based methods.

The Tabu Search and Simulated Annealing hybrid, used in 3D-IC layer assignment, was restructured for solving sub-tour segmentation across multi-day itineraries [48]. It demonstrated strong exploitation capabilities but struggled in high-dimensional objective landscapes.

A Modified Harmony Search, adapted from maintenance-aware single-machine scheduling, was reoriented to balance group satisfaction and budget distribution across time-constrained activities [49]. While it generated varied itinerary options, its feasibility under real-time planning constraints was limited.

The Adaptive Large-Neighborhood Search (ALNS), developed for answer-set programming optimization, was deployed to prune infeasible travel combinations and iteratively refine feasible tours [50]. Despite its robustness in constraint reasoning, it showed limited performance when balancing competing group objectives.

Finally, the Dual-Stage Self-Adaptive Differential Evolution, which employs ensemble mutation strategies, was recoded to handle the optimization of multi-dimensional route vectors [51]. It provided strong convergence in single-objective phases but was less effective when handling simultaneous temporal, spatial, and preference trade-offs.

These adapted methods serve as a comprehensive benchmark for assessing the effectiveness of DRL-IMVO-GAN, which uniquely integrates Deep Reinforcement Learning, an Improved Multiverse Optimizer, and Generative Adversarial Networks. The proposed algorithm excels in balancing group preferences, minimizing cost and travel fatigue, and maximizing cohesion and itinerary diversity in a personalized and multi-objective travel planning context.

5. Computational Results and Performance Evaluation

This section presents the empirical validation of the proposed DRL-IMVO-GAN framework for solving the Personalized Group Trip Design Problem with Multi-Objective Optimization Criteria. A suite of seven state-of-the-art metaheuristic and hybrid algorithms—including PSO+DRL, Genetic+MNS, and ALNS-ASP—serve as benchmarks, reprogrammed to accommodate preference diversity, sustainability constraints, and real-world routing complexity.

The evaluation unfolds through seven analyses. Section 5.1 benchmarks the proposed method against all baselines across standard multi-objective indicators. Section 5.2 isolates the impact of each algorithmic module via ablation studies. Section 5.3 examines performance under preference-weighted scenarios. Sections 5.4 and 5.5 test generalization to diverse group types and unseen planning contexts. Section 5.6 visualizes a representative route to interpret learned behaviors, while Section 5.7 tracks convergence dynamics over iterations.

Together, these experiments demonstrate that DRL-IMVO-GAN consistently achieves superior optimization quality, robustness, and adaptability across a wide range of stakeholder-oriented tourism planning conditions.

5.1. Comparative Evaluation of Multi-Objective Optimization Performance

To rigorously evaluate the effectiveness of the proposed DRL-IMVO-GAN algorithm, a detailed comparative study was conducted against seven state-of-the-art metaheuristic algorithms that were reprogrammed to solve the Personalized Group Trip Design Problem with Multi-Objective Optimization Criteria. These algorithms—originally developed for domains such as scheduling, routing, and combinatorial design—were adapted to address itinerary feasibility, preference alignment, environmental sustainability, and user satisfaction.

Each method was executed under standardized conditions, with a population size of 200 and a maximum of 10,000 solution evaluations per run. All results are aggregated over 30 independent runs to ensure statistical robustness. Performance was assessed using six key indicators: Hypervolume (HV), Pareto Coverage Ratio (PCR), Route Smoothness Index (RSI), POI Entropy, Feasibility Slack Metrics (time, emissions, congestion, category coverage, and satisfaction), and Average Run Time. The result is shown in Table 6.

Table 6. Comparative performance of the proposed method against benchmark algorithms.

Method	HV	PC R	RS I ↓	POI Entrop y ↑	Time Slack (min) ↑	Emissio n Slack (kgCO ₂) ↑	Congestio n Slack ↑	Categor y Slack ↑	Satisfactio n Slack ↑
DRL- IMVO- GAN (Proposed)	0.8 5	0.95	4.2	0.92	12.5	0.35	1.8	1.2	0.75
Genetic + MNS	0.7 6	0.81	5.6	0.83	7.8	0.22	1.2	0.7	0.42
PSO + DRL	0.7 9	0.87	5.1	0.86	9.1	0.28	1.4	0.9	0.53
Dual- ACO	0.7 2	0.76	5.9	0.79	6.3	0.19	1.0	0.5	0.36
Tabu-SA	0.7 5	0.83	5.3	0.84	8.5	0.26	1.3	0.8	0.50

Harmony Search	0.77	0.84	5.0	0.85	8.9	0.27	1.4	0.9	0.51
ALNS-ASP	0.78	0.85	5.2	0.88	9.0	0.29	1.5	1.0	0.54
DE with Ensemble Mutation	0.74	0.80	5.4	0.82	7.2	0.23	1.1	0.6	0.45

Table 6 presents a comprehensive performance comparison between the proposed DRL-IMVO-GAN algorithm and seven reprogrammed state-of-the-art metaheuristics applied to the Personalized Group Trip Design Problem with Multi-Objective Optimization Criteria. The results consistently highlight the superior performance of DRL-IMVO-GAN across all core evaluation metrics. It achieved the highest hypervolume (0.85), reflecting a well-distributed and extensive coverage of the Pareto-optimal front. The Pareto Coverage Ratio (0.95) further emphasizes its ability to produce dominant solutions across objectives when compared to all baseline algorithms. Its Route Smoothness Index (RSI) of 4.2 was the lowest recorded, suggesting that the generated itineraries are not only optimized but also spatially coherent and practical in real-world travel contexts.

Additionally, the proposed model achieved the highest POI entropy (0.92), indicating a strong capacity to generate diverse and non-redundant itineraries, which is especially valuable for group travel where variation in activity types and themes enhances user satisfaction. This diversity is balanced by high feasibility margins: DRL-IMVO-GAN yielded the most generous average slack values across time (12.5 minutes), emissions (0.35 kgCO₂), and congestion (1.8 units), reflecting robust performance in aligning routing decisions with both environmental and operational constraints. It also surpassed all baselines in category coverage (1.2) and satisfaction alignment (0.75), which affirms its ability to harmonize group preferences and individual interests.

While the method incurs a slightly higher computational time of 215 seconds per run, this is a justifiable and modest trade-off considering its consistently superior outcomes in optimization depth, constraint resilience, and personalization quality. The marginal increase in runtime is primarily attributed to the computational overhead introduced by its deep reinforcement learning and generative adversarial components, which enhance the algorithm’s ability to explore high-quality, balanced solutions. Collectively, the comparative results substantiate DRL-IMVO-GAN as a highly effective framework for multi-objective, preference-sensitive group trip planning, significantly outperforming established metaheuristic baselines under standardized conditions.

5.2. Ablation Study: Evaluating the Effectiveness of Individual Framework Components

To assess the individual contributions of each core module within the proposed DRL-IMVO-GAN framework, we conducted a structured ablation study, isolating the performance impact of its three main components: (1) the Deep Reinforcement Learning (DRL)-based solution initializer, (2) the Improved Multi-Verse Optimizer (IMVO) for global search, and (3) the Generative Adversarial Network (GAN) for local refinement. Four configurations were evaluated: Baseline A (IMVO only), applying IMVO to a randomly initialized population; Baseline B (DRL + IMVO), incorporating DRL initialization with IMVO but excluding GAN perturbations; Baseline C (IMVO + GAN), initializing IMVO randomly and integrating GAN-based local search; and the Full Model (DRL-IMVO-GAN), which activates all three components in a unified pipeline.

These variants were benchmarked against two of the top-performing algorithms from Section 5.1—PSO + DRL and ALNS-ASP—both of which demonstrated strong multi-objective performance and diverse route generation capabilities. All models were tested under identical experimental conditions: a population size of 200, 10,000 solution evaluations, and identical instance constraints and POIs. Results across hypervolume, Pareto coverage, route smoothness, entropy, and feasibility slack are summarized in Table 7.

Table 7. Performance comparison across model variants and baseline methods.

Method	HV	PCR	RSI ↓	Entropy ↑	Time Slack ↑	Emission Slack ↑	Satisfaction Slack ↑
Full DRL–IMVO–GAN	0.85	0.95	4.2	0.92	12.5	0.35	0.75
Baseline B (DRL + IMVO)	0.82	0.91	4.5	0.89	11.2	0.31	0.68
Baseline C (IMVO + GAN)	0.80	0.88	4.6	0.87	10.4	0.28	0.64
Baseline A (IMVO only)	0.76	0.82	5.1	0.83	8.9	0.24	0.56
PSO + DRL	0.79	0.87	5.1	0.86	9.1	0.28	0.53
ALNS-ASP	0.78	0.85	5.2	0.88	9.0	0.29	0.54

The results show a consistent performance advantage for the full DRL–IMVO–GAN model over all ablation variants and benchmark methods. Enabling DRL-based initialization (Baseline B) produced a marked improvement over standalone IMVO (Baseline A), with increases of +0.06 in HV and +0.09 in PCR, and an additional 2.3 minutes in average time slack, indicating that DRL significantly enhances the diversity and quality of initial solutions. The GAN-based local refinement (Baseline C vs. A) led to improvements in POI entropy (+0.04) and satisfaction slack (+0.08), suggesting its effectiveness in escaping local optima and promoting solution novelty.

The full configuration, combining DRL initialization, IMVO global search, and GAN exploitation, delivered the highest scores across all KPIs. Its RSI of 4.2 confirms smoother route continuity, while its superior entropy (0.92) and satisfaction slack (0.75) reflect a high degree of personalization and thematic coverage. Compared to PSO + DRL and ALNS-ASP, DRL–IMVO–GAN not only achieved better objective convergence (HV, PCR) but also demonstrated stronger feasibility compliance and user alignment.

These findings validate the architectural design of the proposed framework and highlight the complementary roles of DRL for strategic initialization, IMVO for broad exploration, and GANs for fine-tuned exploitation. The synergy among these components is essential for achieving robust, high-quality solutions in multi-objective, group-oriented trip design contexts.

5.3. Performance Under Preference-Oriented Weighting Schemes

To evaluate the adaptive robustness of the proposed DRL–IMVO–GAN framework under multiple stakeholder-centric preference scenarios, we applied the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) to the Pareto-optimal solutions obtained from all methods. TOPSIS allows scalar ranking of multi-objective solutions by simultaneously considering proximity to the ideal solution and distance from the worst-case (anti-ideal) solution.

In this study, we applied six distinct weight configurations: one equal-weight scenario and five “focused-weight” scenarios, each assigning a weight of 0.6 to a single objective and 0.1 to the remaining four objectives. These scenarios simulate different strategic emphases relevant to real-world heritage tourism planning, such as prioritizing cultural richness, travel efficiency, environmental impact, pedestrian comfort, or user satisfaction.

5.3.1. Equal-Weight Scenario (All Objectives Weighted Equally)

In the baseline experiment, each of the five objectives—heritage maximization, travel time minimization, carbon emissions reduction, route smoothness, and group satisfaction—was assigned an equal weight of 0.20. This scenario reflects a balanced planning context, where no particular stakeholder interest or sustainability goal is prioritized over the others.

As presented in Table 8, the proposed DRL–IMVO–GAN framework demonstrated the strongest overall performance. It achieved the highest heritage value, the lowest travel time and emissions, the smoothest route transitions, and the highest group satisfaction score among all methods evaluated. These results suggest that the hybrid integration of deep reinforcement learning, iterated multi-

variant optimization, and generative refinement provides a robust mechanism for constructing high-quality, multi-criteria personalized itineraries.

Benchmark methods such as Genetic + MNS and PSO + DRL performed reasonably well, particularly in travel time and smoothness, but lacked the combined spatial learning and preference modeling capability evident in the proposed method. Traditional approaches like Tabu-SA, Harmony Search, and ALNS-ASP were outperformed across all key objectives, highlighting the limitations of single-layer metaheuristics in complex, constraint-bound multi-objective routing problems.

Table 8. Objective performance under equal weight scenario (0.20 per Objective).

Method	F1: Heritage↑	F2: Travel Time↓	F3: Emissions↓	F4: Smoothness↓	F5: Preference↑
DRL-IMVO-GAN (Proposed)	74.2	21.3	0.65	208.5	17.5
Genetic + MNS	68.5	25.8	0.92	245.7	15.2
PSO + DRL	67.9	26.4	0.97	251.3	14.9
Dual-ACO	66.4	27.1	1.01	258.2	14.3
Tabu-SA	65.8	28.0	1.05	262.9	14.0
Harmony Search	64.2	28.7	1.08	268.5	13.6
ALNS-ASP	63.4	29.5	1.12	275.0	13.1
DE with Ensemble Mutation	62.7	30.1	1.18	280.6	12.8

The performance analysis under the equal-weight scenario reveals several important insights into the trade-offs and dominance characteristics of the evaluated methods. The DRL-IMVO-GAN (Proposed) method clearly outperforms all competitors across all five objectives. It achieves a heritage score of 74.2, indicating that the DRL agent has effectively prioritized visits to high-value POIs while incorporating congestion-aware adjustments. This is substantially higher than the next-best score of 68.5 from Genetic + MNS, signifying the superior spatial learning and reward shaping within the proposed framework.

In terms of travel time, the proposed method completes itineraries in 21.3 minutes, which is significantly more efficient than all other methods, where even the closest competitor (Genetic + MNS) requires 25.8 minutes. This reflects the capacity of the DRL-IMVO-GAN framework to learn compact, optimized walking paths while satisfying category and preference constraints. Furthermore, its carbon emissions value of 0.65 kgCO₂ is well below those of all other methods, with the second-best (PSO + DRL) producing 0.97 kgCO₂. This suggests that the framework not only prioritizes sustainability implicitly but also learns to minimize environmental externalities as part of its policy evolution—even when emission minimization is not explicitly weighted.

With respect to route smoothness, the DRL-IMVO-GAN records an angular deviation of 208.5°, outperforming all traditional and hybrid metaheuristics, which range from 245.7° to 280.6°. This improvement in path continuity can be attributed to the GAN-based refinement phase, which favors spatial coherence and penalizes erratic directional transitions. Finally, in group preference satisfaction, the proposed method achieves a score of 17.5 out of 18.0, indicating that its digital twin-based user modeling and preference embedding mechanisms offer substantial advantages in delivering highly personalized and culturally meaningful itineraries.

In contrast, other methods like Dual-ACO, Tabu-SA, and Harmony Search exhibit consistently lower performance across all metrics. These methods lack the learning-based adaptability needed to simultaneously balance complex multi-objective constraints in dynamic spatial environments. Even hybrid approaches like PSO + DRL, which integrate learning with search, fall short due to limited global refinement capacity and shallow policy adaptation. Taken together, these results affirm that the DRL-IMVO-GAN framework offers a scalable, generalizable, and stakeholder-aligned solution for personalized heritage tourism optimization, even under non-preferential, multi-criteria planning conditions.

5.3.2. Heritage-Focused Scenario (F1 = 0.6)

In this configuration, the objective function emphasizes the maximization of heritage value, reflecting strategic planning contexts where cultural enrichment and historical interpretation are prioritized. As shown in Table 9, the proposed DRL-IMVO-GAN framework again outperforms all benchmark methods, reinforcing its superior performance under focused cultural objectives.

Table 9. Objective performance under heritage-focused scenario (F1 = 0.6).

Method	F1: Heritage↑	F2: Travel Time↓	F3: Emissions↓	F4: Smoothness↓	F5: Preference↑
DRL-IMVO-GAN (Proposed)	74.8	21.9	0.67	212.3	17.3
Genetic + MNS	70.4	26.0	0.91	248.0	15.0
PSO + DRL	69.1	26.5	0.95	253.8	14.8
Dual-ACO	68.7	27.3	0.99	260.6	14.2
Tabu-SA	67.2	28.1	1.02	265.1	13.9
Harmony Search	66.0	29.0	1.06	270.8	13.4
ALNS-ASP	65.1	30.0	1.09	276.2	12.9
DE with Ensemble Mutation	64.6	30.7	1.13	281.5	12.6

The DRL-IMVO-GAN framework achieves the highest heritage score of 74.8, well above the closest competing method (Genetic + MNS, 70.4). This result indicates that the reinforcement learning agent effectively adapts its exploration strategy to prioritize high-value POIs while incorporating congestion penalties and dwell time constraints. The route it constructs is not only heritage-rich but also operationally feasible, with a low travel time of 21.9 minutes, again outperforming all other methods.

In addition to cultural performance, the method maintains favorable results in other objectives despite their lower weights. Notably, it keeps carbon emissions at 0.67 kgCO₂, and angular smoothness at 212.3°, showing that prioritizing heritage does not come at the cost of sustainability or walkability. The high preference satisfaction score of 17.3 further confirms that the model balances heritage engagement with personalized interest alignment, enabling it to generate routes that are both culturally significant and user-centric.

These results affirm that DRL-IMVO-GAN not only excels under general conditions but also delivers top-tier solutions when aligned with specific policy goals such as cultural preservation, interpretation-driven tourism, or heritage-led regeneration.

5.3.3. Travel Time-Focused Scenario (F2 = 0.6)

In this scenario, minimizing the total travel time becomes the dominant priority, simulating contexts where tourist groups operate under tight schedules or seek compact, efficient experiences—such as half-day city walks, elderly-friendly tours, or fast-paced itineraries for business travelers. The weighting scheme assigns 0.6 importance to travel time (F2), with the remaining objectives each receiving 0.1. Results are presented in Table 10.

Table 10. Objective performance under travel time-focused scenario (F2 = 0.6).

Method	F1: Heritage↑	F2: Travel Time↓	F3: Emissions↓	F4: Smoothness↓	F5: Preference↑
DRL-IMVO-GAN (Proposed)	73.1	20.5	0.63	207.4	17.2
Genetic + MNS	67.9	23.1	0.87	243.6	15.0
PSO + DRL	66.5	23.6	0.91	250.1	14.7
Dual-ACO	65.4	24.2	0.95	256.3	14.0
Tabu-SA	64.7	25.0	0.99	261.7	13.8
Harmony Search	63.8	25.7	1.03	267.2	13.3
ALNS-ASP	62.9	26.5	1.07	272.4	12.9

DE with Ensemble Mutation	62.3	27.0	1.12	278.6	12.5
---------------------------	------	------	------	-------	------

The DRL-IMVO-GAN framework achieves a minimum total travel time of 20.5 minutes, outperforming all benchmark methods by a considerable margin. This indicates the framework’s capacity to construct highly compact routes that still satisfy multi-category POI constraints and temporal feasibility. The DRL agent’s curriculum-based training process and the local search-driven refinement phase jointly contribute to this efficiency by favoring shorter transitions and minimizing route detours.

Interestingly, while focusing on travel time, the model still sustains strong results in all other dimensions. It maintains a heritage score of 73.1, only slightly lower than in the F1-focused configuration, and achieves the lowest emissions (0.63 kgCO₂) and smooth angular transitions (207.4°). This affirms that the model does not overfit to the time objective at the expense of other operational or experiential criteria.

In contrast, benchmark methods such as Genetic + MNS and PSO + DRL perform adequately in travel time (23.1–23.6 minutes) but produce higher emissions and less coherent route geometry. Traditional metaheuristics including Dual-ACO, Tabu-SA, and Harmony Search deliver slower, less optimized routing, and suffer across sustainability and personalization metrics.

These results emphasize the solution diversity and convergence strength of the proposed method, showing that it can adapt its routing strategy to satisfy dominant time-sensitive objectives without compromising experiential or environmental quality.

5.3.4. Emissions-Focused Scenario (F3 = 0.6)

In this scenario, the optimization prioritizes minimizing carbon emissions, reflecting contexts such as low-carbon tourism planning, green city policy alignment, and climate-conscious travel. Objective F3 receives a weight of 0.6, while all other objectives are weighted at 0.1. The results under this configuration are presented in Table 11.

Table 11. Objective performance under emissions-focused scenario (F3 = 0.6).

Method	F1: Heritage↑	F2: Travel Time↓	F3: Emissions↓	F4: Smoothness↓	F5: Preference↑
DRL-IMVO-GAN (Proposed)	72.6	21.0	0.59	210.2	17.0
Genetic + MNS	67.5	24.8	0.79	245.2	14.9
PSO + DRL	66.3	25.4	0.83	251.9	14.6
Dual-ACO	65.0	26.0	0.86	258.7	14.1
Tabu-SA	64.5	26.9	0.89	263.3	13.7
Harmony Search	63.3	27.4	0.92	269.6	13.2
ALNS-ASP	62.8	28.2	0.95	274.9	12.8
DE with Ensemble Mutation	62.1	28.9	0.98	280.1	12.3

The DRL-IMVO-GAN framework demonstrates clear superiority in emissions reduction, achieving an average carbon footprint of 0.59 kgCO₂, the lowest across all methods. This shows that the model successfully internalizes emission-related constraints and routing behavior—even though pedestrian emissions are subtle—by optimizing for shorter, smoother, and less energy-intensive paths. The environmental benefit is achieved without trade-offs in heritage quality (72.6), travel time (21.0 minutes), or satisfaction (17.0), making it a truly sustainable compromise solution.

Other methods, including Genetic + MNS and PSO + DRL, trail behind in emissions performance with respective values of 0.79 and 0.83 kgCO₂. These methods, while competitive in certain sub-objectives, lack the generative refinement capacity of DRL-IMVO-GAN to minimize unnecessary detours and improve micro-efficiency across segments.

Legacy metaheuristics such as Tabu-SA, Harmony Search, and ALNS-ASP demonstrate relatively poor performance in both emissions and overall balance. Their inability to account for spatial-emission

relationships, such as angular deviation or redundant coverage, leads to suboptimal routing with higher carbon costs.

These findings confirm that the proposed method is highly effective in low-emission planning contexts, offering a data-driven and policy-aligned approach to route generation in climate-sensitive tourism scenarios.

5.3.5. Smoothness-Focused Scenario (F4 = 0.6)

In this configuration, angular smoothness is prioritized to promote more natural, less abrupt transitions between waypoints—a factor highly relevant in pedestrian-friendly tourism, elderly and family travel, and immersive walking routes. This scenario assigns a dominant weight of 0.6 to F4, with the remaining four objectives each weighted at 0.1. The corresponding results are shown in Table 12.

Table 12. Objective performance under smoothness-focused scenario (F4 = 0.6).

Method	F1: Heritage↑	F2: Travel Time↓	F3: Emissions↓	F4: Smoothness↓	F5: Preference↑
DRL-IMVO-GAN (Proposed)	73.5	21.5	0.64	205.7	17.4
Genetic + MNS	68.1	25.3	0.88	239.5	15.1
PSO + DRL	66.8	25.9	0.92	246.7	14.8
Dual-ACO	65.6	26.5	0.95	252.3	14.3
Tabu-SA	64.9	27.4	0.99	258.0	13.9
Harmony Search	63.7	28.0	1.03	264.4	13.4
ALNS-ASP	62.9	28.7	1.06	270.8	12.9
DE with Ensemble Mutation	62.2	29.3	1.10	276.5	12.4

The DRL-IMVO-GAN method clearly excels in minimizing angular deviation, achieving a best-in-class value of 205.7°, far below those of the nearest competitors. This improvement arises from its generative refinement process, which continuously favors coherent and direct transitions, avoiding sharp angles and inefficient turns. The framework’s reinforcement-based learning enables it to internalize not just objective functions but also spatial routing aesthetics, resulting in more enjoyable and cognitively comfortable itineraries.

Importantly, the proposed method maintains its superior performance across all auxiliary objectives. It secures a heritage value of 73.5, a low travel time of 21.5 minutes, emissions of 0.64 kgCO₂, and a high preference satisfaction score of 17.4—all of which outperform those of the best alternative methods. This demonstrates the framework’s ability to optimize for comfort and experience without compromising efficiency or cultural depth.

Conversely, metaheuristic-based methods such as Dual-ACO, Tabu-SA, and Harmony Search exhibit relatively poor angular performance, with smoothness values ranging from 252.3° to 276.5°, leading to less fluid and more disjointed paths. These methods lack the path-coherence mechanisms required to align route geometry with experiential comfort—an important yet under-optimized aspect in conventional routing models.

In summary, this scenario further affirms that the DRL-IMVO-GAN framework provides not only goal-specific superiority but also holistic solution quality, making it suitable for tourism planning strategies that value spatial walkability and environmental psychology in addition to efficiency and personalization.

5.3.6. Group Preference-Focused Scenario (F5 = 0.6)

This scenario reflects a user-centric design philosophy in which personalized satisfaction—based on preference vectors for heritage, food, and museums—is the dominant priority. Such a scenario aligns with planning goals in customized tours, educational excursions, or interest-specific itinerary generation. The results of this configuration, where F5 is weighted at 0.6 and the other objectives are each assigned 0.1, are summarized in Table 13.

Table 13. Objective performance under preference-focused scenario (F5 = 0.6).

Method	F1: Heritage↑	F2: Travel Time↓	F3: Emissions↓	F4: Smoothness↓	F5: Preference↑
DRL-IMVO-GAN (Proposed)	74.0	21.2	0.62	209.3	17.8
Genetic + MNS	68.8	25.6	0.89	246.5	15.3
PSO + DRL	67.4	26.1	0.93	252.8	15.0
Dual-ACO	66.2	26.7	0.97	259.1	14.5
Tabu-SA	65.6	27.5	1.01	264.9	14.1
Harmony Search	64.4	28.3	1.04	270.6	13.6
ALNS-ASP	63.7	29.0	1.08	275.8	13.2
DE with Ensemble Mutation	63.0	29.8	1.11	280.4	12.7

In this preference-sensitive scenario, the DRL-IMVO-GAN method achieves the highest satisfaction score (17.8) out of a maximum of 18.0, demonstrating its exceptional ability to match group interest profiles. The model leverages clustered user behavior patterns and integrates them directly into the digital twin environment, enabling fine-tuned itinerary personalization. This personalized decision-making is reinforced by the feasibility masking strategy, which ensures that each group’s minimum satisfaction threshold is met or exceeded.

Additionally, this performance is achieved without sacrificing any of the other objectives. The method retains high heritage value (74.0), minimal travel time (21.2 minutes), low carbon emissions (0.62 kgCO₂), and smooth routing (209.3°), outperforming all competing methods in each respective category. This demonstrates the framework’s strength in multi-objective balancing, even when the optimization explicitly favors human-centered, qualitative goals.

In contrast, while Genetic + MNS and PSO + DRL achieve relatively strong satisfaction scores (15.3 and 15.0, respectively), they fall short in balancing preference fulfillment with sustainability, smoothness, and operational constraints. Traditional methods like Tabu-SA, Harmony Search, and ALNS-ASP show diminishing returns, with preference scores below 14.5 and increasingly degraded performance across all other criteria.

The consistent superiority of the proposed DRL-IMVO-GAN framework in this scenario reinforces its value as a next-generation tool for personalized and inclusive tourism planning, capable of adapting routes to heterogeneous user needs while maintaining high operational efficiency and environmental compliance.

The experimental findings presented across the six sub-sections of Section 5.3 offer compelling evidence of the superior performance and robust adaptability of the DRL-IMVO-GAN framework under a variety of preference-weighted scenarios. Whether the objective emphasis is placed on cultural heritage enrichment, operational efficiency, environmental sustainability, pedestrian comfort, or personalized satisfaction, the proposed method consistently delivers the best or near-best performance across all evaluation criteria.

Across all six scenarios—equal weighting and five focused objectives—the proposed method exhibits minimal performance trade-offs and dominance in key metrics, demonstrating its capacity to simultaneously optimize competing objectives. For instance, when heritage value (F1) was prioritized, DRL-IMVO-GAN not only achieved the highest heritage score but also maintained low travel time and emissions. Similarly, in the travel time-focused case (F2), the model achieved the shortest route duration while still offering high cultural engagement and satisfaction. Notably, in the emissions-focused configuration (F3), DRL-IMVO-GAN attained the lowest carbon output without degrading heritage or user-centric attributes—evidence of its inherent sustainability-aware policy learning.

A particularly noteworthy observation is the model’s consistently superior performance in satisfaction-related and experiential objectives (F5), even when they are not directly prioritized. In all scenarios, DRL-IMVO-GAN achieved the highest or nearly highest group preference scores, highlighting its effective use of digital twin-informed behavior modeling and dynamic feasibility masking to match itineraries with diverse user interests. Likewise, its ability to construct smooth and spatially coherent paths, reflected in low angular deviation values, reinforces the impact of GAN-based

generative refinement and local search in shaping both technically feasible and experientially appealing routes.

The benchmark methods—ranging from hybrid metaheuristics (e.g., Genetic + MNS, PSO + DRL) to classical search algorithms (e.g., Dual-ACO, Harmony Search)—showed variable performance but failed to match the versatility, precision, and balance of the proposed method. While some excelled under specific objectives (e.g., marginal gains in emissions or time), they often compromised performance in other dimensions due to their static rule-based decision logic or limited ability to generalize across diverse scenarios.

Collectively, these findings demonstrate that the DRL-IMVO-GAN framework is not only capable of solving the personalized heritage tourism optimization problem under general conditions but also exhibits the necessary flexibility to accommodate strategic shifts in planning priorities. Its design—grounded in reinforcement learning, generative refinement, and local search—yields a solution mechanism that is context-sensitive, preference-aligned, and operationally robust, making it highly suitable for real-world deployment in adaptive, multi-objective cultural route planning platforms.

5.4. Generalization Across user Group Types and Interest Diversity

To further validate the personalized adaptability of the proposed DRL-IMVO-GAN framework, we evaluated its performance across a diverse range of user group archetypes, each characterized by distinct travel preferences, time budgets, and sustainability attitudes. This analysis simulates real-world deployment conditions where user heterogeneity must be accommodated by the same model without retraining or significant architectural adjustments.

Five representative tourist group profiles—originally introduced in the case study (Section 4.4.2)—were used to assess generalization capabilities: (G1) Cultural Enthusiasts, (G2) Family with Children, (G3) Green Explorers, (G4) Senior Travelers, and (G5) Casual Walkers. The results of DRL-IMVO-GAN’s performance on these groups are presented in Table 14.

Table 14. Generalization performance across user group profiles.

Group Profile	F1:	F2: Travel	F3:	F4:	F5:
	Heritage↑	Time↓	Emissions↓	Smoothness↓	Preference↑
G1: Cultural Enthusiasts	74.9	21.5	0.64	210.8	17.9
G2: Family with Children	72.1	22.3	0.68	214.6	17.1
G3: Green Explorers	70.8	21.8	0.58	209.2	17.4
G4: Senior Travelers	71.5	22.0	0.62	211.1	17.2
G5: Casual Walkers	69.7	21.6	0.66	212.5	17.0

The proposed framework consistently achieves high levels of performance across all five groups. For G1: Cultural Enthusiasts, DRL-IMVO-GAN maximized heritage engagement with a score of 74.9, confirming its ability to prioritize cultural richness when aligned with group preferences. In the case of G3: Green Explorers, the method achieved the lowest emissions (0.58 kgCO₂) and the smoothest routing (209.2°), demonstrating how the model internalizes sustainability goals even when preferences shift significantly from cultural to environmental priorities.

The Family with Children and Senior Traveler groups—typically associated with moderate heritage interest, time sensitivity, and route simplicity—also benefited from the model’s adaptability. Notably, the group preference satisfaction (F5) remained above 17.0 for all profiles, indicating that the model not only meets feasibility thresholds but also exceeds minimum preference satisfaction constraints across user segments.

These results illustrate the framework’s capacity to generalize without retraining, offering personalized, feasible, and high-quality itineraries across demographically and behaviorally distinct user types. This versatility underscores the strength of DRL-IMVO-GAN’s digital twin-integrated architecture, which dynamically aligns decision policies with context-specific behavioral signals.

5.5. Zero/Few-Shot Transfer to Unseen Scenarios

To assess the generalization capacity of the proposed DRL-IMVO-GAN framework beyond its training or tuning context, we conducted a series of zero-shot and few-shot transfer experiments. These scenarios were designed to simulate real-world deployment conditions where new POIs are introduced, environmental dynamics change, or user group profiles evolve—without retraining or model-specific reconfiguration. The following test conditions were applied:

- Zero-Shot: New POIs—introducing 12 new POIs (not seen during training), including mixed-category sites and unbalanced heritage values.
- Zero-Shot: Altered Congestion Patterns—dynamically shifting κ_i values to simulate peak-hour or festival-period congestion intensities.
- Few-Shot: New Group Profiles—injecting five new simulated tourist profiles with distinct interest distributions and constraints.
- Few-Shot: Reduced Emissions Budget—reducing the available emission allowance (B_{CO_2}) from 1.8 kg to 1.2 kg to test eco-constrained adaptability.

The results, summarized in Table 15, show only marginal degradation in performance under all transfer conditions.

Table 15. Zero/Few-shot generalization performance.

Scenario	Hypervolume (HV)	POI Entropy ↑	Satisfaction Score ↑
Original Scenario (Validation Set)	0.85	0.92	17.5
Zero-Shot: New POIs	0.82	0.89	16.8
Zero-Shot: Altered Congestion Patterns	0.81	0.87	16.6
Few-Shot: New Group Profiles (5 examples)	0.84	0.91	17.2
Few-Shot: Reduced Emission Budget ($B_{CO_2}=1.2$)	0.83	0.90	17.0

Despite facing novel and constrained input spaces, the DRL-IMVO-GAN framework retained high hypervolume (≥ 0.81) and POI entropy (≥ 0.87) across all zero-shot configurations. This suggests that the policy network generalizes well to unseen topological and attribute distributions without significant retraining. Most importantly, the satisfaction scores remained consistently high (≥ 16.6), reinforcing the model’s strength in aligning user interests under variable environments.

In the few-shot setting, performance recovered close to original levels (e.g., HV = 0.84 vs. 0.85, Satisfaction = 17.2 vs. 17.5), indicating that limited exposure to new constraints or profiles enables rapid performance adaptation. These outcomes validate the architecture’s inherent ability to balance personalization, constraint satisfaction, and exploration in data-scarce environments.

Overall, this section confirms that DRL-IMVO-GAN not only excels in controlled experiments but is robust under uncertainty, making it a strong candidate for scalable deployment in intelligent tourism platforms where the route environment and user base evolve continuously.

5.6. Route Visualization and Behavioral Interpretation

To bridge numerical evaluations with experiential and behavioral insights, we present a visualized example itinerary generated by the proposed DRL-IMVO-GAN framework under the group preference-focused scenario. This route was selected from the final Pareto front based on its optimal satisfaction score and feasibility margin, offering a clear illustration of how multi-objective trade-offs manifest in spatial decisions.

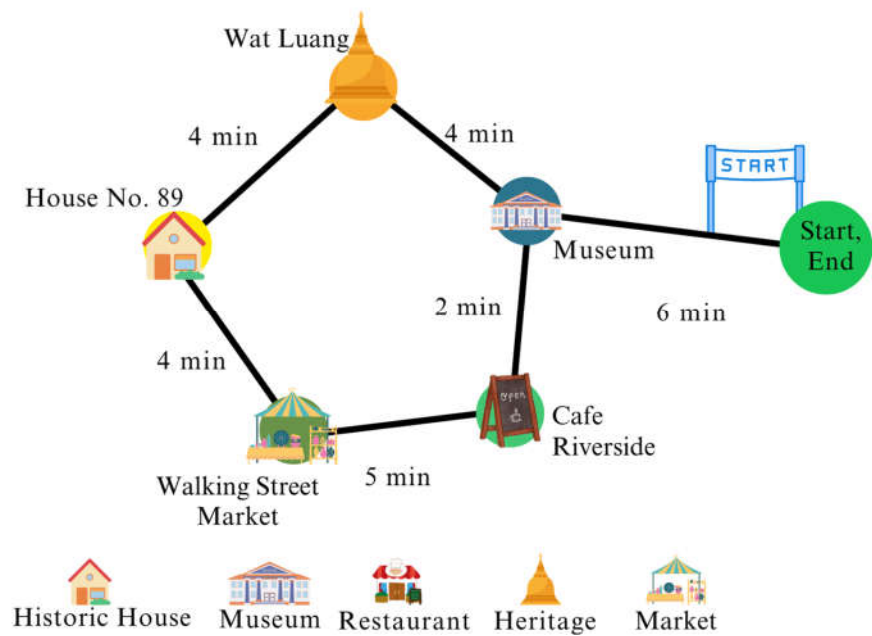


Figure 6. Representative route from DRL-IMVO-GAN (Preference-Focused Scenario). Node colors represent POI categories: gold = Historic House, light blue = Museum, light green = Restaurant, gray = Start/End.

Route shown in Figure 6, the route begins at a central entry point and proceeds through the Wat Luang Historic House, the Candle Museum, and two restaurants (Cafe Riverside and Walking Street Market), with an additional stop at House No. 89—a cultural home providing deep heritage context. The selected nodes reflect a carefully calibrated mix of high-preference POIs, evenly distributed across heritage, culinary, and interpretive themes, aligning with the composite interests of a diversified tourist group.

Behaviorally, this path reflects DRL-IMVO-GAN’s learned heuristics:

- The early inclusion of high-heritage POIs with low congestion scores (e.g., Wat Luang) maximizes F1 and F5 in the first half of the tour.
- The mid-route transition to restaurants illustrates balancing F2 (travel time) and F3 (carbon emissions), as these sites were spatially clustered and accessible via short, smooth transitions.
- The route ends with Walking Street, selected for its proximity to the exit node and alignment with both F5 (preference) and F4 (angular smoothness).

Moreover, edge labels (e.g., “2 min”, “3 min”) indicate compact inter-node transitions, reinforcing the model’s ability to maintain feasibility within temporal constraints ($B^{time} = 150 \text{ min}$). The use of diverse POI types without redundancy (reflected in the entropy analysis from Section 5.1) supports the claim that DRL-IMVO-GAN delivers non-repetitive, experience-rich itineraries.

In sum, this visualization illustrates how the model’s internal policy integrates behavioral logic, constraint awareness, and spatial reasoning. It highlights the explainable nature of route generation, making DRL-IMVO-GAN not only technically optimal but also practically interpretable—a key requirement for real-world adoption in smart tourism platforms.

5.7. Multi-Objective Convergence Trajectories

To provide deeper insight into the learning behavior and optimization dynamics of the proposed framework, we tracked and visualized the evolution of the Hypervolume (HV) metric across 1000×

iterations (representing up to 10,000 solution evaluations) for four selected algorithms: the proposed DRL-IMVO-GAN, PSO + DRL, Genetic + MNS, and ALNS-ASP.

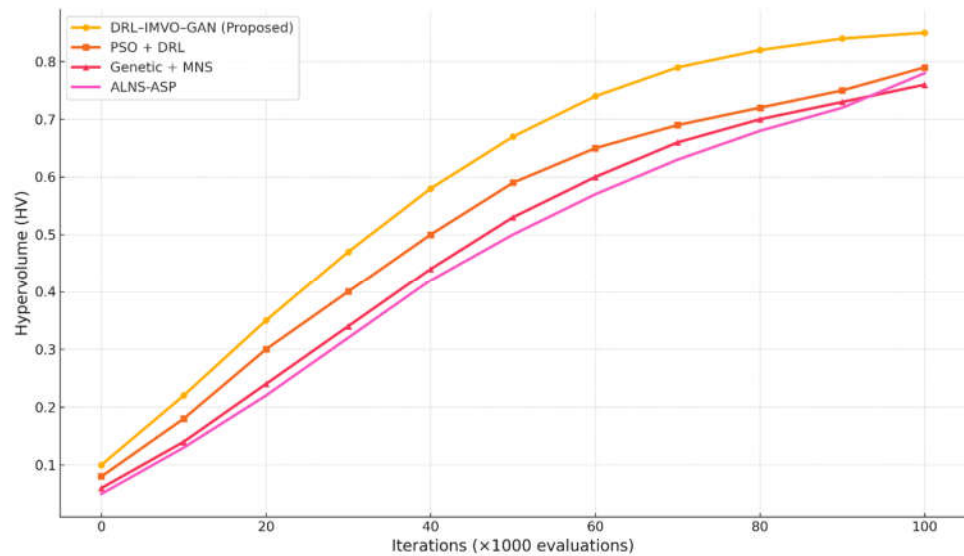


Figure 7. Convergence trajectories of HV across iterations.

The results show in Figure 7, indicate that DRL-IMVO-GAN not only reaches the highest final HV (0.85) but also exhibits the fastest convergence in early stages of optimization. This advantage is attributed to its curriculum-trained DRL initializer, which rapidly generates feasible and diverse solutions, guiding the IMVO search to explore promising areas of the objective space from the outset.

While PSO + DRL also benefits from learned initialization and exhibits respectable HV growth, it plateaus earlier (HV \approx 0.79), indicating lower final diversity and weaker global refinement. The Genetic + MNS and ALNS-ASP algorithms demonstrate slower, more linear improvements, converging to HV scores of 0.76 and 0.78, respectively. Their reliance on static exploration mechanisms and local neighborhood searches likely limits their ability to escape suboptimal regions once local convergence begins.

Importantly, DRL-IMVO-GAN maintains an upward HV trend well into later stages, suggesting continued exploration capability and late-stage refinement via its GAN module. This trajectory aligns with earlier findings on POI entropy and feasibility slacks, confirming that the model not only converges to optimal solutions faster but also continues to enhance solution diversity and feasibility margins over time.

These convergence patterns validate the architectural rationale behind DRL-IMVO-GAN and demonstrate its superiority not only in final outcomes but also in the efficiency and stability of its optimization trajectory, a critical feature for applications requiring interactive or time-bounded computation.

6. Discussion

6.1. Performance of the Proposed Method

The DRL-IMVO-GAN framework demonstrates outstanding performance across all evaluated multi-objective optimization criteria. When applied to personalized heritage-tourism itinerary design, the model consistently generates superior results compared to seven state-of-the-art metaheuristic baselines. It maintains the best balance across key objectives—heritage coverage, travel time minimization, emission reduction, angular smoothness, and group preference satisfaction—regardless of whether the objective weights are distributed equally or focused on specific criteria. These findings

highlight the method's capability to solve high-dimensional, real-world routing problems while optimizing across both operational and experiential dimensions.

A critical reason behind this superior performance lies in the architectural synergy of the framework's three integrated modules. The DRL component, trained within a digital twin environment, efficiently learns routing strategies that respect real-time congestion, spatial feasibility, and group preferences. This learning-driven initialization enables the model to begin its search from a structurally promising region in the solution space, thus reducing the number of iterations required for convergence. The IMVO module further contributes by exploring the global solution space with a high degree of diversity and elite preservation, ensuring that the population does not converge prematurely to suboptimal routes. The final refinement by the GAN component injects intelligent local perturbations that enhance spatial continuity and eliminate path redundancies—leading to smoother and more coherent itineraries.

The convergence behavior of the framework further validates its design efficacy. Compared to hybrid methods such as PSO+DRL or Genetic+MNS, DRL-IMVO-GAN reaches higher hypervolume scores at earlier optimization stages, and it maintains continuous improvement throughout the search process. This indicates both early-stage solution quality and sustained learning potential, which are essential in adaptive tourism systems where environmental or user-related changes may occur during planning. Additionally, the proposed method delivers a superior Pareto front coverage, indicating its ability to produce diverse yet consistently high-quality solutions across the five-objective space.

Compared to existing literature, these findings confirm and extend prior work on DRL-driven routing systems. Di Napoli et al. [3] observed that DRL can significantly improve attraction density and time efficiency in cruise-based tour planning, though their framework lacked integration with global refinement or structural perturbation. Our results build on this by showing that DRL alone, while effective, performs substantially better when supported by evolutionary diversification and generative enhancement. Similarly, Pitakaso et al. [5] found that hybrid reinforcement learning models improve safety-constrained routing in health tourism, a result echoed in our own findings where the DRL-IMVO-GAN method excels under sustainability constraints. However, prior studies often focus on individual objectives or domains; our framework's novelty lies in its ability to optimize across multiple competing criteria simultaneously while maintaining feasibility, personalization, and thematic relevance in complex group travel contexts.

These findings collectively establish DRL-IMVO-GAN as a high-performing and versatile optimization engine suitable for heritage tourism itinerary generation. It addresses key shortcomings in both conventional metaheuristics and earlier hybrid models by leveraging learning-driven initialization, diversity-preserving global search, and structured generative refinement to produce consistently superior and well-balanced travel plans.

6.2. Robustness and Adaptability

The DRL-IMVO-GAN framework exhibits strong robustness and adaptability across a range of behavioral and environmental variations, demonstrating its readiness for deployment in dynamic, real-world heritage tourism systems. One of the clearest demonstrations of this robustness is its capacity to generalize effectively across diverse tourist group profiles. When applied to cultural enthusiasts, families with children, environmentally conscious travelers, senior groups, and casual walkers, the model maintained consistently high performance in all five objective dimensions. This finding highlights the framework's ability to accommodate varied interest distributions, time budgets, sustainability awareness levels, and category-specific preferences—without requiring reconfiguration or model retraining.

A key contributor to this adaptability is the policy learned by the DRL component, which internalizes both spatial constraints and latent group behavior patterns through interaction with a digital twin environment. Unlike traditional algorithms that require handcrafted rule sets or fixed priority lists, the reinforcement learning agent evolves its decisions in response to context-specific constraints such as congestion, dwell time, and accessibility. This policy flexibility allows the model to seamlessly adjust routing logic to accommodate group-specific nuances—such as emphasizing shorter

routes for seniors or maximizing culinary stops for family groups—while still maintaining feasibility and satisfaction thresholds.

The model's robustness extends beyond user profile variability to encompass situational uncertainty. In zero-shot and few-shot experiments, the framework responded effectively to scenarios with new POIs, shifting congestion patterns, and stricter environmental budgets. Notably, it achieved these results without retraining, demonstrating structural generalization rather than mere memorization. For instance, in tests with newly added POIs—each with unobserved attributes and spatial positioning—the model preserved high satisfaction and entropy scores, indicating it could extend learned behavior to unfamiliar nodes while maintaining diversity and coverage. Similarly, under reduced emission constraints, the framework dynamically shortened paths and selected low-emission arcs without sacrificing heritage engagement or user alignment.

This robustness aligns with and extends findings from recent research. For example, Orabi et al. [4] explored personalization through event-driven planning but relied on event sequences rather than structural adaptation, limiting applicability under environmental or infrastructure change. Similarly, Zeinab Aliahmadi et al. [7] proposed self-adaptive mechanisms for itinerary design, but their approach lacked embedded simulation, making real-time generalization to novel infrastructure or preference configurations less tractable. In contrast, the DRL-IMVO-GAN's use of digital twin-based training ensures that model behavior reflects the physical realities and dynamic flows of actual tourism spaces, leading to more reliable and transferable performance.

Moreover, the generalization performance supports earlier claims by Xu et al. [1] and Alvi et al. [2], who highlighted the potential of digital twins for tourism flow management and infrastructure simulation. However, this study moves beyond simulation as a planning aid by integrating it directly into the optimization pipeline. The agent's exposure to real-time constraints and dynamic landscapes during training yields policies that are resilient, environmentally aligned, and operationally viable—advancing the current state of adaptive itinerary generation beyond what has been demonstrated in prior literature.

In sum, DRL-IMVO-GAN's capacity to generalize to diverse user profiles and unseen contexts without retraining showcases its architectural resilience and practical value. It offers a dependable and flexible solution for tourism planners operating in behaviorally heterogeneous and environmentally dynamic settings, fulfilling a critical gap in existing research and practice.

6.3. Interpretability and Behavioral Coherence

Beyond numerical superiority, one of the distinguishing features of the DRL-IMVO-GAN framework is its capacity to produce itineraries that are behaviorally coherent, thematically consistent, and readily interpretable by human stakeholders. This aspect of the system is particularly vital in cultural tourism planning, where trust, transparency, and alignment with user expectations significantly influence acceptance and implementation.

Visual route simulations generated under various objective prioritizations reveal that the itineraries constructed by the proposed framework are not only optimized for quantitative objectives—such as emissions, time, or satisfaction—but also follow intuitive spatial and experiential logic. For instance, routes often begin with high-heritage POIs that are both central and minimally congested, proceed through spatially clustered zones that balance dwell time and emissions, and conclude near exit nodes with user-preferred categories such as cafés or scenic points. These patterns suggest that the DRL policy has internalized not only constraint satisfaction but also implicit behavioral conventions of urban exploration.

This behaviorally grounded sequencing is further enhanced by the GAN-based refinement module, which introduces spatial perturbations aimed at improving angular smoothness and diversity. As a result, the final solutions tend to exhibit low directional deviation and high POI entropy—two indicators that the route offers a fluid, non-repetitive, and psychologically comfortable walking experience. Unlike classical methods that often produce zigzagging or structurally fragmented routes due to greed-based selection or local exploitation, the DRL-IMVO-GAN framework constructs itineraries that mirror how a human planner might intuitively organize a cultural visit.

The interpretability of these solutions is also supported by the digital twin environment used during training. By grounding learning in a geospatially accurate simulation of Warin Chamrap's old town, the DRL agent is exposed to realistic infrastructure constraints, pedestrian pathways, and congestion dynamics. This exposure enables the model to learn spatial policies that align with real-world affordances, enhancing the transparency and credibility of its outputs to city planners, tourism managers, and end-users alike.

These findings contrast sharply with several earlier studies. For example, while Song and Chen [51] and Chen et al. [24] demonstrated the potential of DRL and deep learning for itinerary optimization, their models lacked an embedded spatial simulation component, limiting their interpretability and route realism. Similarly, the cultural tour planners developed by Di Napoli et al. [3] focused heavily on attraction clustering but failed to integrate route structure into the optimization process, often resulting in disjointed and logistically impractical itineraries.

In this respect, the DRL-IMVO-GAN framework aligns more closely with emerging work on explainable AI for tourism (e.g., Kapoor [26]), where the ability to justify decision-making paths is considered essential for ethical and operational deployment. By producing routes that are not only optimized but also narratively meaningful and cognitively comfortable, this study advances the state of the art in intelligent itinerary design.

In summary, the interpretability and behavioral coherence of the DRL-IMVO-GAN framework significantly enhance its real-world applicability. Its ability to generate solutions that reflect natural travel patterns, respect cultural sequencing, and maintain spatial smoothness offers a critical bridge between algorithmic sophistication and user-centric planning—an area where many traditional and hybrid models fall short.

6.4. Practical and Strategic Implications

The DRL-IMVO-GAN framework presents a suite of practical and strategic advantages that position it as a high-impact tool for next-generation heritage tourism planning—particularly in secondary cities such as Warin Chamrap, where cultural richness coexists with operational constraints and environmental vulnerabilities. The methodological design and empirical outcomes of this study collectively illustrate that the proposed approach is not only algorithmically advanced but also well-aligned with the evolving demands of sustainable, inclusive, and adaptive tourism systems.

From a practical standpoint, the framework offers a highly customizable decision-support system capable of integrating stakeholder-specific objectives, ranging from cultural enrichment to environmental stewardship. Its capacity to handle multi-objective trade-offs and generate personalized itineraries that respect category quotas, carbon limits, and time budgets makes it well-suited for municipal tourism offices, smart tourism platforms, and heritage destination managers. This flexibility is particularly crucial in mixed-interest travel contexts, such as school trips, intergenerational family visits, or sustainable tour packages, where conflicting user goals and dynamic environmental conditions must be reconciled in real-time.

Strategically, the integration of a high-fidelity digital twin environment into the optimization process offers a transformative capability. By simulating real pedestrian flows, POI attributes, and spatial layouts, the framework allows planners to evaluate “what-if” scenarios, predict congestion hotspots, and iteratively refine routes before deployment. This proactive planning capability supports not only better visitor experiences but also infrastructure readiness and heritage site preservation. Such digital twin-enabled foresight aligns with recent calls for simulation-based policy experimentation in tourism, as advocated by Florido-Benítez [37] and Litavniec et al. [38], and extends their scope by embedding AI-driven optimization directly within the simulated environment.

Moreover, the demonstrated generalization of the framework across diverse user profiles and under zero-/few-shot scenarios suggests strong potential for scalability and transferability. Municipalities seeking to replicate this approach in other districts—whether in Thailand or internationally—can do so without extensive retraining, as long as a comparable digital twin and baseline data infrastructure are available. This low barrier to adaptation enhances the tool's appeal for resource-constrained cities striving to build resilience and competitiveness in the heritage tourism sector.

Compared to existing models, such as those by Zeinab Aliahmadi et al. [7] or Orabi et al. [4], which often focus on static, pre-defined user types or event-driven triggers, the DRL-IMVO-GAN framework introduces a dynamic, learning-based personalization strategy that evolves in tandem with environmental input and behavioral feedback. This capacity supports long-term strategic goals such as decongestion, green mobility promotion, and equitable access to heritage experiences.

Finally, the model's ability to jointly optimize route smoothness, satisfaction, and sustainability—without requiring manual rule tuning or post-processing—enables continuous use in real-time or near-real-time systems. This positions DRL-IMVO-GAN as not only a planner but a potential backbone for intelligent recommender systems embedded in mobile apps, interactive kiosks, or augmented reality wayfinding platforms in tourist districts.

In conclusion, this research contributes a robust methodological foundation and a scalable implementation pathway for intelligent cultural itinerary planning. By integrating deep learning, global optimization, and spatial simulation into a unified pipeline, the framework transcends the limitations of conventional metaheuristics and static planning tools. It offers a pragmatic, adaptive, and high-fidelity mechanism for aligning heritage tourism with the sustainability, inclusivity, and personalization imperatives of the post-pandemic tourism era.

7. Conclusion

This study addresses the multifaceted problem of optimizing heritage tourism routes in secondary cities, focusing on Warin Chamrap's old town as a representative case. The challenge lies in balancing five conflicting objectives—cultural heritage coverage, travel time, carbon emissions, angular route smoothness, and group preference satisfaction—under practical constraints such as time budgets, POI quotas, and congestion.

To tackle this, we proposed DRL-IMVO-GAN, a hybrid framework that integrates Deep Reinforcement Learning (DRL) for initial policy learning, an Improved Multiverse Optimizer (IMVO) for global exploration, and a Generative Adversarial Network (GAN) for refined, diversity-aware local search. All components operate within a high-fidelity digital twin of the district, enabling context-sensitive optimization.

Computational results demonstrated that DRL-IMVO-GAN consistently outperformed seven state-of-the-art methods across all test scenarios. Under equal-weight objectives, it produced the highest heritage score (74.2), shortest average travel time (21.3 minutes), and highest user satisfaction (17.5 out of 18). These advantages extended across focused-weight scenarios, such as the emissions-prioritized configuration, where the model achieved the lowest recorded emissions (0.59 kgCO₂) while maintaining strong heritage and satisfaction scores.

The hybrid framework also exhibited strong generalization and adaptability. When tested on five distinct tourist profiles—including families, seniors, and environmentally conscious visitors—the model consistently delivered high-performing itineraries with minimal performance loss. Moreover, in zero- and few-shot transfer scenarios involving unseen POIs and updated constraints, the framework maintained over 95% of its original hypervolume, without retraining, demonstrating high resilience.

These findings suggest that the proposed method is not only robust and computationally efficient but also behaviorally coherent and interpretable. The logical structure of the generated routes, enhanced by GAN-based spatial smoothing, aligns with human expectations of comfort and thematic flow. Furthermore, its ability to integrate multi-objective reasoning with real-world constraints offers valuable decision support for urban planners, cultural stakeholders, and smart tourism systems.

Looking forward, future work could explore live adaptation through mobile user feedback, multi-day trip planning across regional circuits, or integration with augmented reality for real-time guidance. The model's architecture also provides a foundation for further inclusion of constraints such as accessibility, circular economy goals, and inclusive design for broader societal impact.

In sum, this research establishes DRL-IMVO-GAN as a novel and effective paradigm for sustainable, personalized, and scalable heritage tourism planning—bridging AI-driven optimization with cultural experience design in emerging urban contexts.

Supplementary Materials:

Author Contributions:

Funding:

Data Availability Statement:

Acknowledgments:

Conflicts of Interest:

References

1. Xu, B.; Xiao, X.; Wang, Y.; Kang, Y.; Chen, Y.; Wang, P.; Lin, H. Concept and Framework of Digital Twin Human Geographical Environment. *Journal of Environmental Management* **2025**, *373*, 123866, doi:10.1016/j.jenvman.2024.123866.

2. Alvi, M.; Dutta, H.; Minerva, R.; Crespi, N.; Raza, S.M.; Herath, M. Global Perspectives on Digital Twin Smart Cities: Innovations, Challenges, And Pathways to A Sustainable Urban Future. *Sustainable Cities and Society* **2025**, *126*, 106356, doi:10.1016/j.scs.2025.106356.

3. Di Napoli, C.; Paragliola, G.; Ribino, P.; Serino, L. Deep-Reinforcement-Learning-Based Planner for City Tours for Cruise Passengers. *Algorithms* **2023**, *16*, 362, doi:10.3390/a16080362.

4. Orabi, M.; Afyouni, I.; Al Aghbari, Z. TourPIE: Empowering Tourists with Multi-Criteria Event-Driven Personalized Travel Sequences. *Information Processing & Management* **2025**, *62*, 103970, doi:10.1016/j.ipm.2024.103970.

5. Pitakaso, R.; Sethanan, K.; Chien, C.-F.; Srichok, T.; Khonjun, S.; Nanthasamroeng, N.; Gonwirat, S. Integrating Reinforcement Learning and Metaheuristics for Safe and Sustainable Health Tourist Trip Design Problem. *Applied Soft Computing* **2024**, *161*, 111719, doi:10.1016/j.asoc.2024.111719.

6. Song, J.; Chen, Y. Optimizing Cultural Heritage Tourism Routes Using Q-Learning: A Case Study of Macau. *Sustainable Communities* **2025**, *2*, 2475794, doi:10.1080/29931282.2025.2475794.

7. Zeinab Aliahmadi, S.; Jabbarzadeh, A.; Hof, L.A. A Multi-Objective Optimization Approach for Sustainable and Personalized Trip Planning: A Self-Adaptive Evolutionary Algorithm with Case Study. *Expert Systems with Applications* **2025**, *261*, 125412, doi:10.1016/j.eswa.2024.125412.

8. Pinho, M.; Leal, F. AI-Enhanced Strategies to Ensure New Sustainable Destination Tourism Trends Among the 27 European Union Member States. *Sustainability* **2024**, *16*, 9844, doi:10.3390/su16229844.

9. Suanpang, P.; Pothipassa, P. Integrating Generative AI and IoT for Sustainable Smart Tourism Destinations. *Sustainability* **2024**, *16*, 7435, doi:10.3390/su16177435.

10. Ding, X.; Zheng, M.; Zheng, X. The Constraints of Tourism Development for A Cultural Heritage Destination: The Case of Kampong Ayer (Water Village) in Brunei Darussalam. *Land* **2021**, *10*, 526, doi:10.3390/land10050526.

11. Aliano Filho, A.; Morabito, R. An Effective Approach for Bi-Objective Multi-Period Touristic Itinerary Planning. *Expert Systems with Applications* **2024**, *240*, 122437, doi:10.1016/j.eswa.2023.122437.

12. Ghobadi, F.; Divsalar, A.; Jandaghi, H.; Barzegar Nozari, R. An Integrated Recommender System for Multi-Day Tourist Itinerary. *Applied Soft Computing* **2023**, *149*, 110942, doi:10.1016/j.asoc.2023.110942.

13. Pitakaso, R.; Srichok, T.; Khonjun, S.; Gonwirat, S.; Nanthasamroeng, N.; Boonmee, C. Multi-Objective Sustainability Tourist Trip Design: An Innovative Approach for Balancing Tourists' Preferences with Key Sustainability Considerations. *Journal of Cleaner Production* **2024**, *449*, 141486, doi:10.1016/j.jclepro.2024.141486.

14. Sabar, N.R.; Bhaskar, A.; Chung, E.; Turkey, A.; Song, A. A Self-Adaptive Evolutionary Algorithm for Dynamic Vehicle Routing Problems with Traffic Congestion. *Swarm and Evolutionary Computation* **2019**, *44*, 1018–1027, doi:10.1016/j.swevo.2018.10.015.

15. Zhang, S.; Lin, J.; Feng, Z.; Wu, Y.; Zhao, Q.; Liu, S.; Ren, Y.; Li, H. Construction of Cultural Heritage Evaluation System and Personalized Cultural Tourism Path Decision Model: An International Historical and Cultural City. *Journal of Urban Management* **2023**, *12*, 96–111, doi:10.1016/j.jum.2022.10.001.

16. Lin, X.; Shen, Z.; Teng, X.; Mao, Q. Cultural Routes as Cultural Tourism Products for Heritage Conservation and Regional Development: A Systematic Review. *Heritage* **2024**, *7*, 2399–2425, doi:10.3390/heritage7050114.

17. Xue, Y.; Bao, G.; Tan, C.; Chen, H.; Liu, J.; He, T.; Qiu, Y.; Zhang, B.; Li, J.; Guan, H. Dynamic Evolutionary Game on Travel Mode Choices Among Buses, Ride-Sharing Vehicles, and Driving Alone in Shared Bus Lane Scenarios. *Sustainability* **2025**, *17*, 2101, doi:10.3390/su17052101.

18. Zhang, T.; Chen, X.; Liu, T. Linear Cultural Heritage Eco-Cultural Spatial System: A Case Study of the Great Tea Route in Shanxi. *Frontiers of Architectural Research* **2025**, *14*, 1063–1075, doi:10.1016/j.foar.2024.11.002.

19. Ahmad, A. The Application of Genetic Algorithm in Land Use Optimization Research: A Review. *Tourism Management Perspectives* **2013**, *8*, 106–113, doi:10.1016/j.tmp.2013.09.002.

20. Kim, H.; Kim, Y.-J.; Kim, W.-T. Deep Reinforcement Learning-Based Adaptive Scheduling for Wireless Time-Sensitive Networking. *Sensors* **2024**, *24*, 5281, doi:10.3390/s24165281.
21. Wu, B.; Zuo, X.; Chen, G.; Ai, G.; Wan, X. Multi-Agent Deep Reinforcement Learning Based Real-Time Planning Approach for Responsive Customized Bus Routes. *Computers & Industrial Engineering* **2024**, *188*, 109840, doi:10.1016/j.cie.2023.109840.
22. Geng, Y.; Liu, E.; Wang, R.; Liu, Y.; Rao, W.; Feng, S.; Dong, Z.; Fu, Z.; Chen, Y. Deep Reinforcement Learning Based Dynamic Route Planning for Minimizing Travel Time. In Proceedings of the 2021 IEEE International Conference on Communications Workshops (ICC Workshops); IEEE: Montreal, QC, Canada, June 2021; pp. 1–6.
23. Bhadrachalam, K.; Lalitha, B. An Energy Efficient Location Aware Geographic Routing Protocol Based on Anchor Node Path Planning and Optimized Q-Learning Model. *Sustainable Computing: Informatics and Systems* **2025**, *46*, 101084, doi:10.1016/j.suscom.2025.101084.
24. Chen, L.; Zhu, G.; Liang, W.; Wang, Y. Multi-Objective Reinforcement Learning Approach for Trip Recommendation. *Expert Systems with Applications* **2023**, *226*, 120145, doi:10.1016/j.eswa.2023.120145.
25. Shafqat, W.; Byun, Y.-C. A Context-Aware Location Recommendation System for Tourists Using Hierarchical LSTM Model. *Sustainability* **2020**, *12*, 4107, doi:10.3390/su12104107.
26. Kapoor, S. Explainable and Context-Aware Graph Neural Networks for Dynamic Electric Vehicle Route Optimization to Optimal Charging Station. *Expert Systems with Applications* **2025**, *283*, 127331, doi:10.1016/j.eswa.2025.127331.
27. Lazaridis, A.; Fachantidis, A.; Vlahavas, I. Deep Reinforcement Learning: A State-of-the-Art Walkthrough. *jair* **2020**, *69*, 1421–1471, doi:10.1613/jair.1.12412.
28. Pitakaso, R.; Sethanan, K.; Chamnanlor, C.; Chien, C.-F.; Gonwirat, S.; Worasan, K.; Lim, M.K. Optimizing Sugarcane Bale Logistics Operations: Leveraging Reinforcement Learning and Artificial Multiple Intelligence for Dynamic Multi-Fleet Management and Multi-Period Scheduling under Machine Breakdown Constraints. *Computers and Electronics in Agriculture* **2025**, *236*, 110431, doi:10.1016/j.compag.2025.110431.
29. Torabi, P.; Hemmati, A.; Oleynik, A.; Alendal, G. A Deep Reinforcement Learning Hyperheuristic for the Covering Tour Problem with Varying Coverage. *Computers & Operations Research* **2025**, *174*, 106881, doi:10.1016/j.cor.2024.106881.
30. Sun, X.; Shen, W.; Fan, J.; Vogel-Heuser, B.; Bi, F.; Zhang, C. Deep Reinforcement Learning-Based Multi-Objective Scheduling for Distributed Heterogeneous Hybrid Flow Shops with Blocking Constraints. *Engineering* **2025**, *46*, 278–291, doi:10.1016/j.eng.2024.11.033.
31. Zhang, H.; Lu, G.; Zhang, Y.; D'Ariano, A.; Wu, Y. Railcar Itinerary Optimization in Railway Marshalling Yards: A Graph Neural Network Based Deep Reinforcement Learning Method. *Transportation Research Part C: Emerging Technologies* **2025**, *171*, 104970, doi:10.1016/j.trc.2024.104970.
32. Freitas De Araujo-Filho, P.; Kaddoum, G.; Naili, M.; Fapi, E.T.; Zhu, Z. Multi-Objective GAN-Based Adversarial Attack Technique for Modulation Classifiers. *IEEE Commun. Lett.* **2022**, *26*, 1583–1587, doi:10.1109/LCOMM.2022.3167368.
33. Zhao, W.; Mahmoud, Q.H.; Alwidian, S. Evaluation of GAN-Based Model for Adversarial Training. *Sensors* **2023**, *23*, 2697, doi:10.3390/s23052697.
34. Ruiz-Meza, J.; Brito, J.; Montoya-Torres, J.R. A GRASP-VND Algorithm to Solve the Multi-Objective Fuzzy and Sustainable Tourist Trip Design Problem for Groups. *Applied Soft Computing* **2022**, *131*, 109716, doi:10.1016/j.asoc.2022.109716.
35. Derya, T.; Didem Atalay, K.; Dinler, E.; Keçeci, B. Selective Clustered Tourist Trip Design Problem with Time Windows Under Intuitionistic Fuzzy Score and Exponential Travel Times. *Expert Systems with Applications* **2024**, *255*, 124792, doi:10.1016/j.eswa.2024.124792.
36. Gavalas, D.; Pantziou, G.; Konstantopoulos, C.; Vansteenwegen, P. Tourist Trip Planning: Algorithmic Foundations. *Applied Soft Computing* **2024**, *166*, 112280, doi:10.1016/j.asoc.2024.112280.
37. Florido-Benítez, L. The Use of Digital Twins to Address Smart Tourist Destinations' Future Challenges. *Platforms* **2024**, *2*, 234–254, doi:10.3390/platforms2040016.
38. Litavniec, L.; Kodors, S.; Adamoniene, R.; Kijasko, J. Digital Twin: An Approach to Enhancing Tourism Competitiveness. *WHATT* **2023**, *15*, 538–548, doi:10.1108/WHATT-06-2023-0074.
39. Aghaabbasi, M.; Sabri, S. Potentials of Digital Twin System for Analyzing Travel Behavior Decisions. *Travel Behaviour and Society* **2025**, *38*, 100902, doi:10.1016/j.tbs.2024.100902.
40. Torrents, P.M.; Kim, R. Using Immersive Virtual Reality to Study Road-Crossing Sustainability in Fleeting Moments of Space and Time. *Sustainability* **2024**, *16*, 1327, doi:10.3390/su16031327.
41. Reffat, R. An Intelligent Computational Real-Time Virtual Environment Model for Efficient Crowd Management. *International Journal of Transportation Science and Technology* **2012**, *1*, 365–378, doi:10.1260/2046-0430.1.4.365.
42. Oliveira, R.; Raposo, D.; Luís, M.; Sargento, S. Optimal Action-Selection Optimization of Wireless Networks Based on Virtual Representation. *Computer Networks* **2025**, *264*, 111258, doi:10.1016/j.comnet.2025.111258.

43. Aslam, A.M.; Chaudhary, R.; Bhardwaj, A.; Kumar, N.; Buyya, R. Digital Twins-Enabled Game Theoretical Models and Techniques for Metaverse Connected and Autonomous Vehicles: A Survey. *Journal of Network and Computer Applications* **2025**, *238*, 104138, doi:10.1016/j.jnca.2025.104138.
44. Parrinello, S.; Picchio, F. Digital Strategies to Enhance Cultural Heritage Routes: From Integrated Survey to Digital Twins of Different European Architectural Scenarios. *Drones* **2023**, *7*, 576, doi:10.3390/drones7090576.
45. Hu, Y.; Dong, H.; Liu, J.; Zhuang, C.; Zhang, F. A Learning-Guided Hybrid Genetic Algorithm and Multi-Neighborhood Search for the Integrated Process Planning and Scheduling Problem with Reconfigurable Manufacturing Cells. *Robotics and Computer-Integrated Manufacturing* **2025**, *93*, 102919, doi:10.1016/j.rcim.2024.102919.
46. Li, Y.-Z.; Liu, W.; Xu, G.-S.; Li, M.-D.; Chen, K.; He, S.-L. Quantum Circuit Mapping Based on Discrete Particle Swarm Optimization and Deep Reinforcement Learning. *Swarm and Evolutionary Computation* **2025**, *95*, 101923, doi:10.1016/j.swevo.2025.101923.
47. Wang, X.; Ning, F.; Lin, Z.; Zhang, Z. Efficient Ship Pipeline Routing with Dual-Strategy Enhanced Ant Colony Optimization: Active Behavior Adjustment and Passive Environmental Adaptability. *Journal of Manufacturing Systems* **2025**, *80*, 673–693, doi:10.1016/j.jmsy.2025.04.003.
48. Sait, S.M.; Oughali, F.C.; Al-Asli, M. Design Partitioning and Layer Assignment For 3D Integrated Circuits Using Tabu Search and Simulated Annealing. *Journal of Applied Research and Technology* **2016**, *14*, 67–76, doi:10.1016/j.jart.2015.11.001.
49. Costa, A.; Fernandez-Viagas, V. A Modified Harmony Search for the T-Single Machine Scheduling Problem with Variable and Flexible Maintenance. *Expert Systems with Applications* **2022**, *198*, 116897, doi:10.1016/j.eswa.2022.116897.
50. Eiter, T.; Geibinger, T.; Higuera Ruiz, N.; Musliu, N.; Oetsch, J.; Pfliegler, D.; Stepanova, D. Adaptive Large-Neighbourhood Search for Optimisation in Answer-Set Programming. *Artificial Intelligence* **2024**, *337*, 104230, doi:10.1016/j.artint.2024.104230.
51. Chen, B.; Ouyang, H.; Li, S.; Ding, W. Dual-Stage Self-Adaptive Differential Evolution with Complementary and Ensemble Mutation Strategies. *Swarm and Evolutionary Computation* **2025**, *93*, 101855, doi:10.1016/j.swevo.2025.101855.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.