applied sciences

Article

Transportation Mode Selection Using Reinforcement Learning in Simulation of Urban Mobility

Mehmet Bilge Han Taş [1,2,*], Kemal Özkan [1,3], İnci Sarıçiçek [3,4] and Ahmet Yazici [1,3]

[1] Department of Computer Engineering, Eskisehir Osmangazi University, Eskisehir 26040, Türkiye; kozkan@ogu.edu.tr (K.Ö.); ayazici@ogu.edu.tr (A.Y.)
[2] Department of Computer Engineering, Erzincan Binali Yildirim University, Erzincan 24100, Türkiye
[3] The Center for Intelligent Systems Applications Research (CISAR), Eskisehir 26040, Türkiye; incid@ogu.edu.tr
[4] Department of Industrial Engineering, Eskisehir Osmangazi University, Eskisehir 26040, Türkiye
* Correspondence: bilgehantas@erzincan.edu.tr

Abstract: Transportation mode selection is pivotal for navigating through cities plagued by heavy traffic congestion. This plays a crucial role in ensuring the efficient utilization of time and resources to achieve the desired objectives. Given the complex dynamics of urban mobility, strategically selecting a transportation mode can significantly mitigate delays and enhance overall productivity in densely populated areas. The objective of this study is to find the most efficient result among various transportation modes to make deliveries from different points on a university campus. To solve this problem, reinforcement learning was used and tested on the simulation environment SUMO. Traffic density was increased by using an equal number of different transportation modes, such as driving, cycling, motorbiking, and walking. Various traffic densities were generated, and different reward models were applied to select the best means of transportation. Various probability distributions were used as reward models to avoid the unfair distribution caused by how near or how far the road is when moving from random points to the destination region. As a result of the models created using the applied reward–penalty functions, it was determined that the best means of transportation in areas with a low traffic density is cycling, and in areas with high traffic density, the optimal mode of transportation is motorbiking.

Keywords: transportation mode selection; reinforcement learning; Gaussian distribution; Poisson distribution; simulation of urban mobility (SUMO)

check for updates

Received: 4 December 2024
Revised: 27 December 2024
Accepted: 9 January 2025
Published: 15 January 2025

Citation: Taş, M.B.T.; Özkan, K.; Sarıçiçek, İ.; Yazici, A. Transportation Mode Selection Using Reinforcement Learning in Simulation of Urban Mobility. *Appl. Sci.* **2025**, *15*, 806. https://doi.org/10.3390/app15020806

1. Introduction

Due to widespread urbanization and an increase in the number of vehicles, traffic congestion has become a severe problem. An increase in transportation time causes increased fuel consumption and air pollution [1,2]. Since eliminating routing and traffic problems will increase delivery efficiency, it is necessary to develop different solutions, especially in logistics [3]. Eto et al. chose off-the-shelf software or a delivery route at the discretion of the couriers. While going from one place to another, solutions are generally not produced within a systematic problem network [4], and on a very large map and in scenarios with different situations, it is necessary to select the best route. This wide and complex network is best resolved when various parameters are considered, such as traffic jams, traffic lights, and pedestrian factors. In this way, it is possible to select an advantageous transportation mode.

Transportation mode selection is an essential area regarding transportation systems. Different modes of transportation (car, motorcycle, bicycle, walking, etc.) are essential for mode selection and transportation characteristics, and mode choice analysis can solve traffic congestion or eliminate related transportation problems through considering different travel behaviors [5]. By using mode selection, it is possible to offer solutions that are suitable for different city conditions. By handling transportation mode selection differently, the best transportation method can be chosen, allowing for different alternatives to be used in situations where there is a lot of traffic.

Mode selection methods are frequently used to solve problems such as travel and route planning. Mode selection aims to maximize the selection of vehicles available on a route [6]. Transportation modes such as walking, cycling, and driving each have their own parameters. These parameters include many factors, such as the specifics of the vehicle to be used (maximum speed, fuel consumption, etc.), the condition of the roads, and the effect of traffic lights on it [5]. The aim is to investigate how the relationships between travel time, costs, and land use patterns influence choice and travel patterns relevant to where people live and work. The urban form of housing, workplaces, and travel time and cost are crucial factors in travel choice [7], and transportation mode selection is bound to reduce many costs in urban life. Transportation cost, time, and shortest route can be considered as costs here. It is inevitable that these factors will be faced much more in bigger, real-world problems. The methods used in the study aim to address transportation time regarding the use of alternative modes of transportation during traffic congestion that may occur in an area during an event or mass organization.

In addition to intuitive approaches, methods such as reinforcement learning (RL) and deep reinforcement learning (DRL) are also used to solve the transportation mode selection problem (TMSP). With these methods, it becomes possible to reach an instant solution by taking into account changing environmental activities. It is possible that different routes can provide more effective solutions due to instantaneous changes in the environment. Even in the conditions of roads or traffic lights used by vehicles or pedestrians, it is possible to obtain a better result using such approaches [8,9].

In this study, the following areas are addressed:

- A unique reward mechanism model is presented using the simple model, Gaussian distribution, and Poisson distribution.
- Deliveries were made to a collection area on a university campus using different starting points and various transportation modes. Within the framework of this scenario, it was determined how the most efficient vehicle and the most appropriate route would be selected.
- To solve this problem, RL was used, and the best reward mechanism was determined with the help of MDP. How each vehicle traveled to its destination, the environmental restrictions, different variations, and the effect of these on the result were examined.
- SUMO, which is a simulation environment, was used to ensure the realization of the study and to control the steps.

Background and Motivation

The main focus of this study is the increasing importance of traffic density and difficulties in choosing transportation modes in modern cities. Rapidly increasing urbanization and vehicle numbers have seriously increased urban traffic congestion, making it necessary to develop innovative solutions to optimize the transportation infrastructure of cities. Traffic congestion not only increases travel times but also leads to wider economic and environmental problems such as fuel consumption, air pollution, and general environmental impacts. For this reason, choosing transportation modes efficiently in limited areas such as urban areas and university campuses is of great importance both in terms of saving time and using resources efficiently [10–12].

In order to make traffic congestion and transportation systems more efficient, optimization methods should be used to select different transportation modes (car, bicycle, motorcycle, walking, etc.). These choices depend on many factors, such as environmental conditions, traffic density, road conditions, and public transportation facilities. In this context, in addition to traditional solution methods, modern artificial intelligence approaches such as reinforcement learning (RL) and deep reinforcement learning (DRL) offer powerful tools to solve such complex problem sets. These methods allow for dynamic solutions to be produced according to instantaneous changes in traffic conditions. Reinforcement learning (RL) is an algorithm that allows agents to learn by interacting with their environment and allows them to determine optimal actions over time. The use of RL algorithms in areas that require decisions based on time and environmental factors, such as traffic management, provides more effective and flexible transportation solutions. In this study, RL algorithms were used to select various transportation modes in scenarios where traffic density is increased on university campuses. Simulation environments such as SUMO (Simulation of Urban MObility) are frequently preferred tools for such traffic simulations, and this study aims to simulate real-world scenarios by providing integration with SUMO.

Providing a completely randomly distributed traffic environment allows for a comparison with real-world problems. The operation of a completely randomly distributed traffic environment allows for real-time problem matching. In this study, Gaussian and Poisson distributions were used to take into account the different distributions that occur in real-world scenarios. In addition, the effectiveness of these distributions was compared with the created simple model. As a result, this study aims to increase efficiency by using reinforcement learning and simulation techniques to select transportation modes on university campuses. It aims to test the applicability of these methods not only in limited areas such as university campuses but also in wider urban transportation systems.

2. Related Work

In traffic optimization studies, transportation mode selection using reinforcement learning is often used to find the best transportation mode. Various studies have been carried out to predict and control a traffic situation using deep learning. In these studies, traffic jam and situation studies were generally carried out [13–15]. It is also a frequently studied topic for solving complex networks, RL-DRL solutions, and transportation mode selection [16–19]. The best mode is usually found in such studies, and a reward policy is determined; this process is carried out using MDP. Wei et al. [20] used deep reinforcement learning for a system with traffic lights. In this work, real-world problems were not studied but were discussed theoretically, and the studies focused on reward policies and comparisons with real data, with the results supported by the SUMO simulation environment. Similarly, Liang et al. [21] carried out work to increase efficiency in relation to the inefficient operation of traffic lights and the energy problem caused by long waiting times. They emphasized the need to dynamically adjust real-time traffic information. The problem was solved by using DRL in the model they created. In addition, the reward mechanism was determined using MDP. In order to increase the performance of the proposed model, it was optimized using methods such as Q-learning, which was implemented in the SUMO environment. According to Isel et al. [22], by using Deep Q-Network, the intersections on roads could be autonomously handled. In these studies, a road network determined on SUMO was used and each step was graphed. In general, routing optimization is mentioned, and how the network can be best utilized with learning and reward mechanisms using reinforcement learning is revealed.

In order to reduce network delay, Stampa et al. [23] created a deep reinforcement learning agent that can automatically adapt to the state of traffic. This agent then suggests customized configurations; however, instead of considering the entire urban transportation network, it just focuses on one vehicle. Order dispatching is a continuous operation, and the state changes of the couriers at each time step resulting from the order assignment action form a sizable but finite MDP that may be resolved by using conventional RL techniques [24]. By choosing the appropriate action when interacting with the environment, the RL aims to maximize the returned reward. A popular model-free RL technique built on the Bellman equation is Q-learning [25]. The typical Q-learning method aims to produce an over-estimation of the Q-value; in contrast, the Double Q-learning algorithm uses two different action–value functions, Q and Q′, where the Q′ function compensates for computing the expected Q-value in the Bellman equation. Additionally, the Double Deep Q Network (DDQN) [26] was created to replace the tabular method of the Q-value with neural networks in order to overcome the disadvantage of discrete states. To increase the learning effectiveness of DQNs, a progressive transfer knowledge learning strategy with related features was proposed to handle the vast quantities of real-world spatiotemporal trip data [27]. In large-scale ride-sharing scenarios, a multi-agent reinforcement learning (MARL) approach was proposed to analyze ride-sharing data in a decentralized manner [28].

Li et al. [29] unraveled the complex relationships between transportation mode choices and the factors that influence them, which is valuable for transportation planning and development. However, a better understanding of the underlying influencing factors that shape passengers' comprehensive mode choices for travel, as well as the interactive and nonlinear effects between the two, is needed. In this study, a field survey was conducted in Xi'an, a tourist city in China, collecting travel mode data (airplane, high-speed rail (HSR), conventional trains, and express buses) to obtain a comprehensive range of passenger characteristics, including relatively under-researched factors such as passengers' online ticketing methods. Research on individual decision-making processes is fundamental to discovering macroscopic behavioral rules in travel mode selection. Qin et al. [30] designed a behavioral experiment with a process-tracing method to obtain data on repeated travel mode choices under different contexts. A stochastic and dynamic model based on Decision Space Theory has been proven to be reliable and is used to reproduce and analyze the repeated decision-making processes. Passengers can use this system to facilitate their choice of travel mode selection. Traffic policies can increase the use of Park and Ride by making passengers re-evaluate, weigh, and compare relevant information.

Based on the list of studies covered in this section, one can see the active area of study regarding transportation mode selection. Most studies attempt to explain this using utility theories, which state that the mode of transportation with maximum utility in a given situation will be chosen based on travel costs and duration, among other variables. There are many alternative transportation methods for urban transportation. These transportation modes change according to the priorities of the passengers. Travel demand may change for many reasons such as heavy or light traffic in cities and the use of different roads in traffic [30,31]. Arentze and Timmermans conducted a study on which is the most accurate travel mode choice in a transportation network. A solution was attempted using Deep Reinforcement Learning (DRL) by formulating a Markov Decision Process (MDP). It provides a suggestion with DRL algorithms by clustering the people who will travel. SUMO was used for verification purposes and the studies were confirmed in this simulation environment. Finding the ideal transportation mode was attempted by using various transportation vehicles such as cars, subways, buses, and time windows. Arentze and Timmermans [32] aimed to find the best means of transportation dynamically using RL.

Based on the interaction of passengers with the environment, continuous recommendations were provided by using reward functions. It was used for a large-scale system and the best means of transportation between the destination and the starting points were selected. It also showed effective performance in different traffic conditions.

3. Materials and Methods

In this study, trip generation, trip distribution, mode choice, and trip assignment, which are the steps of travel demand modeling, are partially discussed. For trip generation and trip distribution, the starting points on the simulation network were randomly assigned, but the destination points were fixed because they are an activity area [33–38]. While this approach provides a simplified model at the campus scale, modeling based on more detailed demographic or economic data was not performed. Mode selection was performed by a reinforcement learning algorithm based on traffic density, travel time, and defined reward mechanisms (Gaussian and Poisson distributions). This method ensured that the choice of transport modes adapted to changing traffic conditions and optimized the efficiency of the overall system as well as individual users. Trip assignment was performed in the SUMO simulation environment by means of internal routing algorithms. These algorithms assign routes by optimizing the trips according to the shortest travel time and network conditions on the existing traffic network. These steps effectively enable the evaluation and comparison of transport modes under different traffic scenarios. However, it should be noted that the framework does not fully incorporate real-world travel demand patterns (e.g., user preferences, land use, or socioeconomic factors).

3.1. Problem Definition

Practical and feasible solutions are proposed to address transportation mode selection problems on university campuses. The developed optimization method serves as an important tool to enhance the daily life of the campus community and to optimize campus transportation overall. By employing reinforcement learning (RL) for transportation mode selection, it seeks to identify the best mode through a reward-and-penalty system, incorporating numerous parameters and varying approaches. Furthermore, the findings have the potential to inspire other institutions and organizations facing similar challenges and contribute to the resolution of comparable issues. Ultimately, the work represents a valuable step toward improving transportation management on university campuses.

The primary objective is to provide practical and feasible solutions to transportation mode selection problems on university campuses. The developed optimization method serves as a crucial tool to facilitate the daily life of the campus community and to optimize campus transportation in general. By employing reinforcement learning (RL) for transportation mode selection, the study aims to identify the best mode through a reward-and-punishment system that integrates numerous parameters and innovative approaches. Additionally, this work has the potential to inspire other institutions and organizations facing similar challenges, contributing to the broader field of transportation management.

A scenario is considered where various starting points and transportation methods are randomly determined to reach the assembly area or destination on the university campus. The destination area shown in Figure 1 represents a significant point within the campus. Starting points, as well as the means used to reach this point, are selected randomly to simulate real-world conditions. This randomized approach provides a general simulation-based framework, enabling a better approximation of real-life scenarios involving diverse starting points and transportation options. The inclusion of random starting points reflects the arrival of students, staff, or visitors from different campus locations, while the use of various transportation modes represents alternatives such as walking, bicycling, driving, or using public transport.

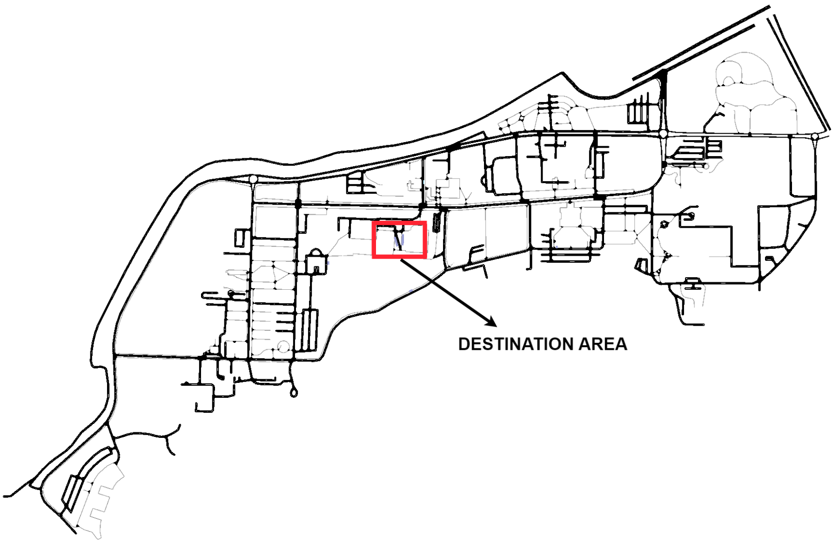

Figure 1. The map of the ESOGU campus used in the study.

According to the case scenario, a material needs to be transported to the destination area given in Figure 2 from many different points and via different means of transportation. The red square shown in the figure is determined as the destination arena. All modes of transportation are used to set off and try to arrive at this area. The blue rectangles are the roadside points required to ensure arrival, and they are set to be connected to both roads. In order to reach this area, one must comply with the traffic rules and, at the same time, decide what the best means of transportation is, taking into account the disadvantages of traffic congestion. In order to achieve these results, various reward models should be created and the results should be compared.

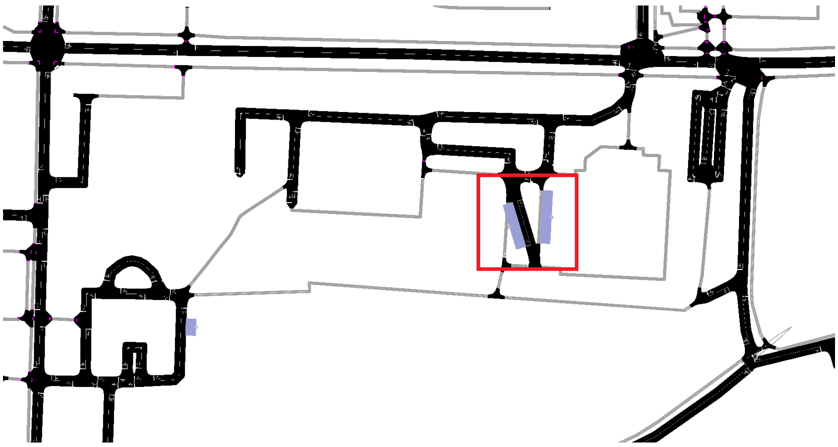

Figure 2. Destination area. The red box highlights the activity area within the destination zone.

The evaluation of the routes to be carried out in this scenario and the awards collected is an important step in terms of efficiency and the optimization of on-campus transportation. The aim of this study is to evaluate the efficiency and optimization of transportation on the university campus by analyzing randomly generated routes and collected rewards. This evaluation was carried out to measure and improve the performance of the proposed method and algorithm for solving the routing problem. In this way, practical and applicable solutions can be offered to optimize on-campus transportation and determine the best routes to the assembly area.

3.2. Reinforcement Learning

Reinforcement learning (RL) refers to a set of algorithms that address sequential decision-making and are capable of making intelligent decisions based on their local environment. A model that instructs an agent on the course of action to follow in a confined environment to maximize a predetermined total reward is known as an RL algorithm. The agent experiments with various strategies, calculating the overall payoff. Following numerous trials, the algorithm establishes a pattern of behavior by learning which behaviors yield more rewards. The algorithm is able to instruct the agent on what to do in any scenario. Figure 3 illustrates this interaction [39]. In Figure 3, r_t refers to the reward the agent receives from the environment, a_t refers to the actions the agent sends to the environment, and s_t refers to the states the agent is in the environment. At any time t, while the agent is in state s_t, it performs the action, receives the reward $r_t + 1$, and accordingly moves to state $s_t + 1$. The value-based implementation of RL is shown in the formula below [40].

The environment (E) is the world in which the agent operates and makes decisions. The agent is the entity applying the RL algorithm to learn and make decisions. The state (S) is a representation of the current situation of the agent in the environment. The action (A) is the set of all possible moves the agent can make in a given state. The reward (R) is the feedback received by the agent from the environment after performing an action, indicating the immediate benefit of that action. The Q-value (Q(s, a)) is the expected total reward for taking action a in state s.

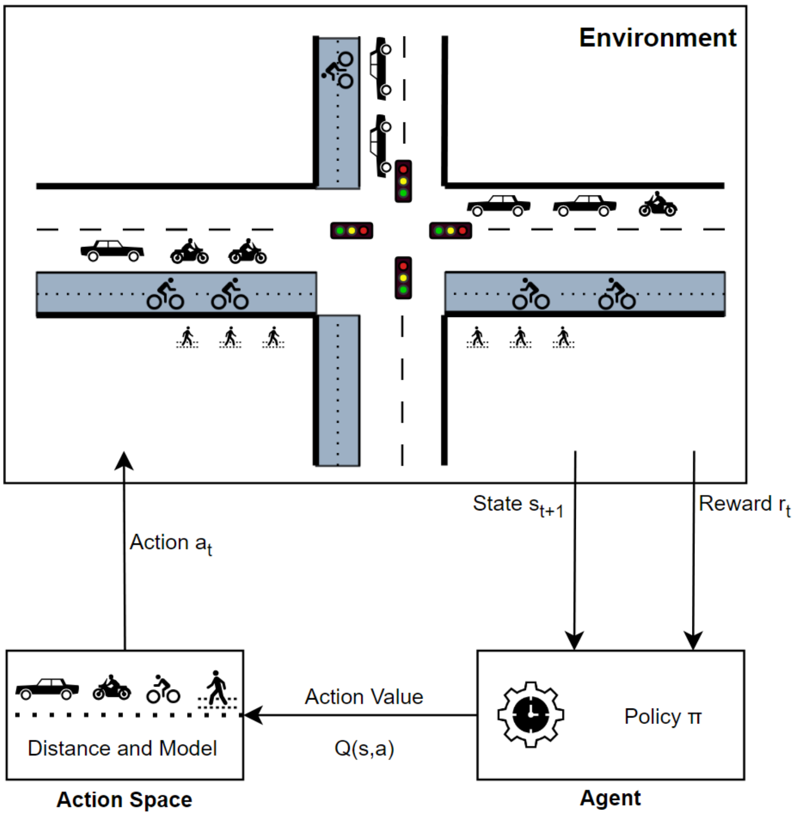

Figure 3. Types of vehicles used in urban traffic and the desired activity area.

The environment represents the setting or scenario where the agent operates, serving as the backdrop for the agent's decision-making processes. Actions refer to the choices made by the agent in response to the information it gathers from the environment. These actions can be of two types, continuous-valued or discrete-valued, depending on the specific problem and environment at hand. The state embodies the critical information that characterizes the agent's position and condition within the environment.

The main reason we used reinforcement learning (RL) is the ability of this method to provide effective solutions in complex, dynamic, and uncertain environments. The problems encountered in choosing transportation modes are not only static but also involve a number of factors that change over time and interact with each other. Such problems create difficulties in producing optimal solutions with traditional methods because existing solutions are usually based on fixed rules and cannot adapt quickly to environmental changes. Reinforcement learning allows agents to continuously learn to make better decisions at each step by interacting with their environment. The agent receives rewards according to the results of each action and optimizes future actions in light of these rewards. This process helps to find the best solution over time, especially in complex and dynamic systems. In choosing transportation modes, RL algorithms can evaluate the advantages and disadvantages of each vehicle (car, bicycle, motorcycle, or walking) in real time. For example, when road density increases, a motorcycle may be preferred over a bicycle, but these changes must be learned quickly based only on instantaneous data. Traditional optimization and modeling techniques usually work with fixed parameters and focus on providing the best solution valid at a given moment. However, this approach cannot quickly account for instantaneous changes (for example, fluctuations in traffic density). RL has a structure that can dynamically adapt to changes and learn from previously encountered situations. This capability facilitates continuous improvement in the selection of transportation modes and enables more accurate and effective decisions with each new experience. By employing RL, real-time and flexible solutions can be developed for transportation mode selection. This approach not only identifies the best solution under current conditions but also establishes a system that can quickly adapt to future changes and continuously improve over time.

A snapshot of the agent's current state can be represented in either continuous-time or discrete-time formats, depending on the temporal characteristics of the problem and environment. Rewards act as the feedback mechanism provided by the environment in response to the agent's actions. These rewards can take on either positive or negative values, reflecting the desirability or undesirability of the agent's decisions and behavior. Lastly, the policy plays a pivotal role in guiding the agent's decision-making process. It outlines how an agent, situated in state s_t at time t selects an action from the available set of actions. The policy can be designed as either deterministic or stochastic, representing the strategy or approach that the agent follows in navigating its environment.

Table 1 shows the key hyperparameters of the Reinforcement Learning (RL) model used in the study. The hyperparameters represent critical settings that affect the learning process of the model and were carefully selected to optimize the performance of the model:

Table 1. Hyperparameters used in the reinforcement learning model.

Parameter	Symbol	Value
Learning Rate	α	0.001
Discount Factor	γ	0.95
Epsilon Initial Value	ϵ_0	1.0
Epsilon Decay Rate	ϵ_{decay}	0.9
Maximum Steps per Episode	$N_{\max}$	1000

The learning Rate (α) determines how much learning the model performs at each step. A small learning rate (0.001) allows the model to learn slowly and steadily, thus avoiding sudden fluctuations. The discount Factor (γ) determines the impact of future rewards on the current decision. A value of 0.95 provides a balance between short-term and long-term rewards. The epsilon initial value (ϵ_0) defines the initial probability of discovery. A high initial value (1.0) allows the model to make a random discovery in the initial stages.

The epsilon decay rate (ϵ_{decay}) controls how the probability of discovery decreases with each iteration. A value of 0.9 ensures that the epsilon decays gradually and the model tends towards more stable strategies. The maximum steps per episode (N_{max}) specifies the maximum number of steps that can be performed in each episode. A limit of 1000 steps provides a suitable framework for the model to perform enough learning within an episode.

Initially, all Q-values are set to zero or some arbitrary value:

$$Q(s,a) = 0 \quad \text{for all states } s \text{ and actions } a \tag{1}$$

Agent Interaction and Action Selection

At each time step, the following occurs:

1. Observe the current state, s.
2. Select an action, a. Typically, the ϵ-greedy strategy is used:
   - With probability ϵ, choose a random action;
   - With probability $1-\epsilon$, choose the action with the highest Q-value:

$$a = \arg\max_{a'} Q(s,a') \tag{2}$$

Executing the Action and Receiving the Reward

The agent performs action a, observes the new state s', and receives reward r.

Updating values

The Q-values are updated using the following rule:

$$Q(s,a) \leftarrow Q(s,a) + \alpha\left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right] \tag{3}$$

where

- α is the learning rate (a value between 0 and 1);
- γ is the discount factor (a value between 0 and 1 that determines the importance of future rewards).

Updating the State

Update the current state to the new state as follows:

$$s \leftarrow s' \tag{4}$$

Iteration

Summary of Formulation

1. $Q(s,a) = 0$.
2. Repeat for each episode:
   (a) $s = s_0$;
   (b) Repeat for each step of the episode:
      i. $a = \text{action selection}(s, Q, \epsilon)$;
      ii. Take action a, and observe r and s';
      iii. $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$
      iv. $s \leftarrow s'$.
   (c) until s is terminal.

Unlike many other techniques of machine learning, RL requires the system to figure out which actions yield the highest reward through experimentation. Actions can affect not only the current reward but also the next state and all subsequent rewards in the most fascinating and difficult circumstances. The three primary characteristics of RL are its closed-loop nature, lack of explicit instructions on action selection, and the gradual emergence of action consequences, such as reward signals. The fundamental idea behind RL is to distill the salient features of the actual challenge that an agent faces when interacting with its environment in order to accomplish a goal. Evidently, such an agent should be able to sense the state of the environment to some extent and take actions that affect it. The goal of RL is to capture more complex structures and use more adaptive algorithms than classical machine learning, and RL algorithms are more dynamic in their behavior compared to classical machine learning algorithms. RL algorithms are based on the MDP. MDP is a specialized stochastic time control process for decision-making. The main actor of the RL algorithm is the following: an agent, which is a structural component designed to perceive and understand its surrounding environment, enabling it to make informed decisions and take appropriate actions within that context.

SUMO, also known as Simulation of Urban MObility, is a potent simulator with a route-planning and car-following model that can manage a huge load network and defined traffic demand. Additionally, it offers a wealth of relevant data, such as vehicle speed, model, and location. The Traffic Control Interface (TraCI), a Python API that considers the SUMO simulation as a server and enables users to view or alter traffic simulations, is one of the main components of SUMO. The integration of external systems (or libraries) with SUMO traffic simulation is made possible by TraCI. Urban traffic management is a prime illustration of how reinforcement learning is used in the real world. As the number of cars in our urban transportation system rises, there is a growing need for a smart traffic management system that can provide intelligent vehicle routing. Nonetheless, the intricacy of the urban transportation network poses several challenges in traffic management because of the rapid alterations in traffic patterns and the extensive dispersion of automobiles on the road. Given that reinforcement learning has already demonstrated its ability to tackle difficult optimization problems, it would be beneficial to apply this method to address these challenges [18].

3.3. Simulation of Urban Mobility (SUMO)

Here, a standardized simulation platform would be helpful. The desire to win the support of other universities was the second justification for making the simulation open source, and Figure 4 shows an example [41]. The Traffic Control Interface, or TraCI for short, is a Python API that treats the SUMO simulation like a server and lets users obtain data from it or alter a traffic simulation. This is one of the most significant components of SUMO. SUMO traffic simulations can be integrated with other systems or libraries thanks to the interface TraCI offers. This webinar explains how urban traffic management can benefit from the integration of reinforcement learning with SUMO via TraCl [18].

SUMO is a free, open-source traffic simulation and transportation modeling program widely used for urban planning, traffic management, and autonomous vehicle research. It offers tools like NETEDIT for creating network models, SUMO-GUI for interactive simulation control, and TRACI for integration with external applications. With its comprehensive database, SUMO enables the development of custom scenarios and realistic traffic simulations. Its key features include high accuracy, flexibility, scalability for large-scale simulations, real-time data analysis, and support for diverse traffic scenarios. Being open-source and highly extensible, SUMO is a preferred choice for traffic simulation and transportation modeling.

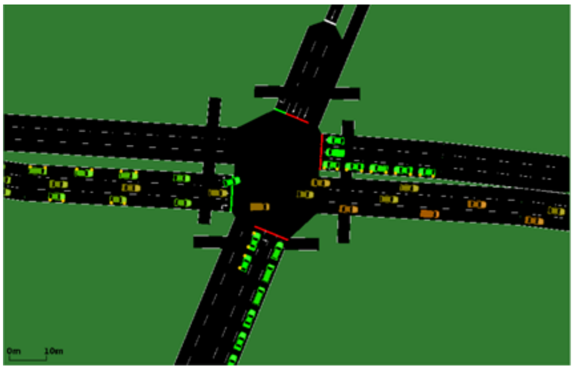

Figure 4. Example traffic scenario in the SUMO simulation environment. Different colors represent various types of vehicles or traffic states: green indicates moving vehicles, yellow represents vehicles slowing down, and red highlights vehicles that are stopped or waiting.

4. Experimental Setup

The reward mechanisms to which these rules will be linked are detailed in the following sub-headings. Thanks to this system, it will be seen what the results will be in a university campus in a normal traffic density and in a high traffic density. In this way, it is shown more clearly which is the best means of transportation.

4.1. Baseline Models

The baseline models used in this study were chosen to evaluate the effectiveness of the proposed Gaussian and Poisson reward mechanisms and to compare them with other methods. These models represent different types of approaches to travel mode choice and are used as reference points to more clearly demonstrate the performance advantages of the proposed mechanisms. The results of the baseline models were analyzed in comparison with the proposed methods in different traffic density scenarios. This analysis provides an important basis to demonstrate the accuracy and generalizability of the model.

B-coefficient Logistic regression is linked to utilization theory, where the mode chosen is based on users' preferences for certain variables. In other words, utilization theory uses individuals' beliefs based on their preferences and is widely used by economists to explain user choice behavior based on their preference rankings. The equation below shows the logit model derived based on the sigmoid function of logistic regression and is widely used in modeling travel mode choice [42].

$$P(Y_i) = \frac{1}{1 + e^{-(B_0 + B_1 X_{1i} + B_2 X_{2i} + \cdots + B_n X_{ni})}} \tag{5}$$

where

- $P(Y_i)$ is the predicted probability that Y is true for case i;
- e is a mathematical constant;
- B_0 is a constant value estimated from the data;
- B_1 is predicted from data B-coefficient;
- X_i is the observed score on variable X for case i.

The Random Utility Model [43], with two subgroupings (nests), determines a choice via the combination of the characteristics of the mode and the mode available. The utility function of the nested logit model has common unobserved characteristics (errors). Equation (6) shows the utility function for nested logit (NL) and cross-nested logit (CNL) [44].

$$U_{em}^n = V_e^n + V_m^n + V_{em}^n + \epsilon_e^n + \epsilon_{em}^n \tag{6}$$

where

- e is the existing mode;
- m is the travel mode;
- V_e^n is the deterministic component of the utility of the current mode;
- V_m^n is the deterministic component of the utility of the mode;
- V_{em}^n is the deterministic component of the co-benefit between the current mode and the mode;
- ϵ_e^n is the error term for present or absent modes (Gumbel distribution);
- ϵ_{em}^n is the error term for the combination of two subgroups (Gumbel distribution).

The Reinforcement Learning Reward Model [45] is a reward system with time included in the reward used. Observed travel time, pre-event waiting time, and late arrival time are discretized and added to the reward as a penalty. The discretization is configurable and is set to 5 min in the experiments. The reward function for each agent is defined as follows:

$$\text{Reward} = \begin{cases} 1 - ott & \text{if on time,} \\ 0 - (ott + t_{\text{wait}}) \cdot w_{\text{wait}} & \text{if too early,} \\ 0 - (ott + t_{\text{late}}) \cdot w_{\text{late}} & \text{if too late,} \\ 0 - (ett_{\max} \cdot 2) \cdot w_{\text{late}} & \text{if never arrived.} \end{cases} \tag{7}$$

If the agent shows up on time (that is, within 15 min prior to the event), the sole penalty is the observed travel time (ott), which measures how long it takes to get there. The waiting time is enhanced by the waiting time multiplier after being added to ott if the agent comes too early. w_{wait} = 1.0 is set for all experiments. The time from the event time to the destination is used to calculate the lateness (t_{late}) if the agent arrives too late. Similar to early arrival, t_{late} is increased by the lateness multiplier (w_{late}) after being added to ott if the agent arrives too late and chooses the worst kind of transportation at the event time. In this instance, figuring out the estimated travel times for each mode from the origin to the destination and choosing the highest number yields the maximum estimated trip time (ett_{max}). After multiplying (ett_{max}) by two, the penalty is computed by multiplying it by the delay multiplier (w_{late}).

4.2. Environment and Ground Rules

All initial conditions and traffic configurations are described in detail to ensure the reproducibility of the experimental setup used in this study. The experiments were conducted using SUMO 1.16 simulation software, and a campus-scale traffic network of 1 km $\times$ 1 km was configured. The simulation network includes basic elements such as roads, pedestrian crossings, and traffic lights. In addition, the same area can be recreated using the Open Street Map (OSM) provided by SUMO to replicate the environment. To test different traffic densities, 100, 200, 300, and 400 vehicle/pedestrian configurations were used. The origin and destination points were randomly assigned among all valid nodes in the network. The traffic lights were configured according to the default algorithms of SUMO, and each experiment lasted until the last vehicle completed its journey. The traffic density scenarios were low-density traffic (scenarios where the starting points were far apart), medium-density traffic (scenarios where the starting points partially overlapped), and high-density traffic (scenarios where the traffic lights caused congestion). Vehicle speed limits and congestion measurements were provided by the default settings of SUMO and the TraCI interface. In the future, it is planned to visualize the parameters used, such as starting points, traffic density distributions and traffic light durations, and to document the simulation scenarios in detail.

When designing the infrastructure of the learning environment, various vehicles, traffic lights, and the traffic congestion should be taken into consideration. For this reason, it is crucial to find out which vehicles travel faster to a gathering area, taking into account that each traffic environment has its own variables. Although vehicles are not homogeneously distributed in real-world scenarios, we deemed it important to use the same number of each type of vehicle in this study for a fairer evaluation of the experimental results. Reward mechanisms are not based on randomness but on some mathematical distribution functions to provide a certain mathematical basis. In this way, it is ensured that it can be used and generalized in different environments. Four different transportation groups were used in the simulation environment; these are the following: cars, motorcycles, bicycles, and pedestrians. As soon as the simulation environment starts, 400, 800, 1200, and 1600 vehicles, each in equal numbers, appear in the environment to travel from a completely random point to the destination point. In order to test the system performance, a large number of generated vehicles were tested. One of the main objectives here is to evaluate the performance of the SUMO environment and to observe the negative impact of waiting vehicles in heavy traffic. Each transportation vehicle acts in accordance with the traffic rules on the way to the destination point. In the experimental environment, there are various roads that can be found on a university campus, such as pedestrian-only roads, pedestrian-bicycle-only roads, or roads open to all vehicles. This leads to different advantages for the randomly distributed traffic population. A calculation is performed based on the arrival time for the vehicles to find the best route. The optimum arrival time is the average arrival time of the vehicles in that traffic. In this way, late arrivals and early departures are penalized with this reward mechanism. The reason for penalizing those who wait too long is to prevent unnecessary waiting for a vehicle that arrives at a point early. This will cause an overall delay in the system as the vehicles in this situation cannot be re-routed. The main goal is to reach the event area within the best time by traveling closest to the average arrival time.

As given in Algorithm 1, a value-based algorithm was used to perform transportation mode selection. The main purpose of this algorithm is to apply a reward/penalty system to find which transportation mode is best in terms of reaching the destination. In the initial state, s for all states and a for actions are set to zero. Certain states are observed, and an action is performed using the epsilon-greedy strategy. Here, the main criterion is the distance covered by the method to be used for transportation and, in addition, the time to reach the event area. Depending on the arrival time, a penalty is applied in cases of early or late arrival. Thanks to this process, learning is provided to see the most suitable transportation mode. Since transportation modes do not have equal initial distances, this randomness will be provided in a fair manner with the given reward mechanism. At the beginning of the algorithm, the Q-values for all states and actions are reset. This ensures that each state and action pair is initially evaluated equally. This step specifies that the algorithm will repeat until a certain stopping criterion is reached. After observing the current state, the agent chooses an action using an ϵ-greedy strategy. If a randomly chosen probability value is less than ϵ, the agent chooses an action at random. If a randomly chosen probability value is greater than ϵ, the agent chooses the action with the highest Q-value given the current state. After performing the chosen action, the agent observes the new state and reward.

A penalty is applied when the vehicle's arrival time deviates from the scheduled time, either arriving earlier or later. The penalty is calculated based on the arrival time; for example, the penalty for arriving 2 min early or late is multiplied by 2, while the penalty for arriving 5 min early or late is multiplied by 5. These penalties are added to the vehicle's arrival time. The calculated penalty is applied by subtracting the current reward. The Q-values are updated using the observed reward and the expected future reward. The update formula is obtained by adding the difference between the reward multiplied by the learning rate and the expected future reward to the current Q-value. Here, the learning rate (α) is a value between 0 and 1, and the discount factor (γ) determines the importance of future rewards over current rewards. This is also a value between 0 and 1. The current state is updated as the new state, and the algorithm repeats these steps until a certain stopping criterion is reached. This criterion can usually be the completion of a certain number of episodes or when the change in Q-values is very small. In this way, the algorithm performs the learning process by taking into account the penalties according to the arrival time and updating the Q-values in order to make better decisions in the future.

Algorithm 1 Unified model for transportation mode selection.

Initialize Q-values: $Q(s,a) = 0$ for all states s and actions a.
repeat
  Observe the current state s.
  Select an action a using ϵ-greedy strategy:
  if random probability $< \epsilon$ **then**
    Choose a random action.
  else
    Choose the action with the highest Q-value: $a = \arg\max_{a'} Q(s,a')$.
  end if
  Execute the action a and observe the new state s' and reward r.
  if using penalty model **then**
    if arrival time is early or late **then**
      Apply penalty based on how early or late:

$$\text{Penalty} = \begin{cases} 2 \times \text{arrival time} & \text{if 2 min early/late} \\ 5 \times \text{arrival time} & \text{if 5 min early/late} \\ 10 \times \text{arrival time} & \text{if 10 min early/late} \\ 15 \times \text{arrival time} & \text{if 15 min early/late} \end{cases}$$

      Add penalty to the reward: $r \leftarrow r - \text{Penalty}$
    end if
  else if using Gaussian model **then**
    Apply Gaussian reward adjustment:

$$\text{gaussianReward} = \begin{cases} \text{arrivetime} - (\text{arrivetime} \times \text{reward}) & \text{if on time} \\ \text{arrivetime} - (\text{arrivetime} \times \text{reward}) & \text{if too early} \\ \text{arrivetime} - (\text{arrivetime} \times \text{reward}) & \text{if too late} \end{cases}$$

    Update reward $r \leftarrow \text{gaussianReward}$
  else if using Poisson model **then**
    Apply Poisson reward adjustment:

$$\text{poissonReward} = \begin{cases} \text{arrivetime} - (\text{arrivetime} \times \text{reward} \times v) & \text{if on time} \\ \text{arrivetime} - (\text{arrivetime} \times \text{reward} \times v) & \text{if too early} \\ \text{arrivetime} - (\text{arrivetime} \times \text{reward} \times v) & \text{if too late} \end{cases}$$

    Update reward $r \leftarrow \text{poissonReward}$
  end if
  Update Q-values:

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

  Update the current state: $s \leftarrow s'$.
until convergence

Early and late arrivals incur penalties, whereas arriving on time neither incurs penalties nor earns rewards. This neutral treatment of on-time arrivals can be considered a reward in itself, as it ensures no negative impact is associated with punctuality. Showing whether a vehicle is on time or not will be different for each traffic group. In this way, both a realistic system and a fixed value are created in every size scenario because the total time to reach the destination in a very crowded environment is different from the time to reach it in less crowded traffic. In order to overcome this situation, each group is evaluated within itself, and the average of the arrival times of all vehicles is taken.

All initial conditions and traffic configurations are described in detail to ensure the reproducibility of the experimental setup used in this study. The experiments were conducted using SUMO 1.16 simulation software, and a campus-scale traffic network of 1 km $\times$ 1 km was configured. The simulation network includes basic elements such as roads, pedestrian crossings, and traffic lights. In addition, the same area can be recreated using the Open Street Map (OSM) provided by SUMO to replicate the environment. To test different traffic densities, 100, 200, 300, and 400 vehicle/pedestrian configurations were used. The origin and destination points were randomly assigned among all valid nodes in the network. The traffic lights were configured according to SUMO's default algorithms, and each experiment lasted until the last vehicle completed its journey. The traffic density scenarios were low-density traffic (scenarios where the starting points were far apart), medium-density traffic (scenarios where the starting points partially overlapped), and high-density traffic (scenarios where the traffic lights caused congestion). Vehicle speed limits and congestion measurements were provided by the default settings of SUMO and the TraCI interface. In the future, it is planned to visualize the parameters used, such as starting points, traffic density distributions, and traffic light durations, and to document the simulation scenarios in detail.

4.3. Simple Model

The Simple model imposes an extra penalty for every 2, 5, 10, and 15 min these vehicles are late or early to the event site. These penalties are added back as a multiplier on top of the arrival time.

$$r = \begin{cases} \text{arrivetime} & \text{if on time} \\ \text{arrivetime} + (\text{arrivetime} \times \text{Penalty}) & \text{if too early or too late} \end{cases} \tag{5}$$

where

$$\text{Penalty} = \begin{cases} 2 & \text{if 2 min early/late} \\ 5 & \text{if 5 min early/late} \\ 10 & \text{if 10 min early/late} \\ 15 & \text{if 15 min early/late} \end{cases} \tag{6}$$

The Q-value update step incorporating the reward function is as follows:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[\begin{cases} \text{arrivetime} & \text{if on time} \\ \text{arrivetime} + (\text{arrivetime} \times \text{Penalty}) & \text{if too early or too late} \end{cases} + \gamma \max_{a'} Q(s',a') - Q(s,a) \right] \tag{7}$$

4.4. Gaussian Distribution Model

The Gaussian distribution is a form of probability distribution that is frequently utilized in probability theory and statistics. It also goes by the name of the normal distribution. The central limit theorem is closely related to the Gaussian distribution, which is used to represent a variety of natural events and datasets. Mean and standard deviation are the two fundamental characteristics of the Gaussian distribution. The formula is given below [32,46];

$$f(x) = \frac{1}{\sigma\sqrt{2\Pi}} exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}) \tag{8}$$

In the Gaussian distribution equation, we see the following: x is the value of the random variable, μ is the mean of the distribution, σ is the standard deviation of the distribution, and exp is Euler's number (approximately equal to 2.71) in the equation.

A bell curve with a symmetrical shape represents the Gaussian distribution. The mean establishes the distribution's center, and the standard deviation establishes how dispersed or tight the distribution is. Numerous applications of the Gaussian distribution exist, particularly in the domains of statistics, engineering, economics, natural sciences, and data science. Making predictions, testing hypotheses, and studying data distribution all benefit from using this method.

As shown in Figure 5, these distributions can be translated into a reward mechanism that is exactly fit for purpose. Vehicles that arrive at the destination on time will be rewarded as they approach the average arrival time. Late and early arrivals are penalized by deducting the arrival time penalties and distributing a reward. The reward model is as follows:

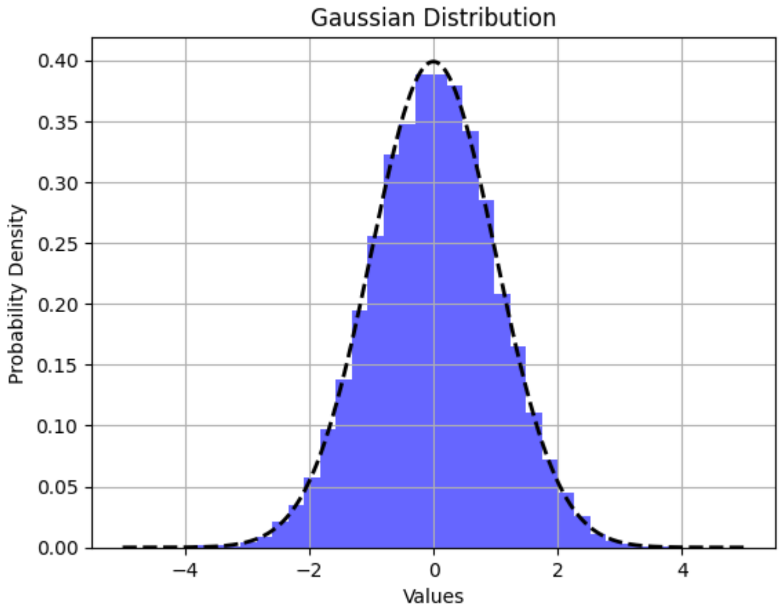

Figure 5. A sample Gaussian distribution generated in a large sample.

4.5. Poisson Distribution Model

The probability distribution used to model how frequently uncommon occurrences take place over time or in a certain location is called the Poisson distribution. The Poisson distribution, which is frequently used to explain the numbers or lengths of random events, is named after the French mathematician Siméon Denis Poisson. The Poisson distribution's probability function is as follows [47]:

$$P_x(k) = \frac{e^{-\lambda} * \lambda^k}{k!} \tag{9}$$

In the Poisson distribution equation, k is the number of successes (an integer); λ is the average number of successes (a positive real number); e is Euler's number; and $k!$ is a factorial of k, which is the product of all positive integers from 1 to k. This formula calculates the probability of a specific number of events (k) occurring within a fixed interval of time (λ). It is used to model situations where events occur independently and at a constant average rate. As shown in Figure 6, the Poisson distribution is more rigid and sharply distributed than the Gaussian distribution. Therefore, more rewarding values are entered for vehicles that reach the destination on time. The reward model of the distribution is as follows:

The Poisson distribution equation gives the number of successes k (an integer); the average number of successes λ (a positive real number); Euler's number e; $k!$, the factorial of k, which is the product of all positive integers from 1 to k. This formula calculates the probability of a specific number of events (k) occurring within a fixed interval of time (λ).

It is used to model situations where events occur independently and at a constant average rate. The Poisson distribution is more rigid and sharply distributed than the Gaussian distribution. Therefore, more rewarding values are entered for vehicles that reach the destination on time. The reward model of the distribution is as follows:

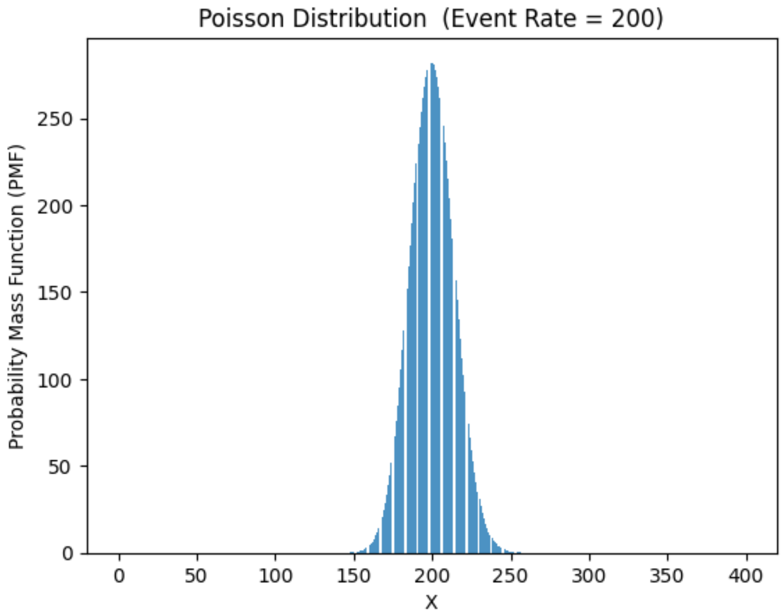

Figure 6. A sample Poisson distribution generated in a large sample space.

v is the coefficient applied when the scenario is realized. It is used as (v) = 1 in the 400-traffic scenario and (v) = 2 in the 800-traffic scenario. The main reason for this is that, in a Poisson distribution, the two opposite ends diminish very sharply. This gives more consistent results in a fair reward system.

5. Results and Discussion

This study assessed the efficiency of individual transport modes (walking, cycling, motorbiking, and driving) on a university campus scale. Public transport and high-capacity vehicles (e.g., minibuses and buses) were excluded from the analysis. The primary reason for this exclusion was to simplify the simulation structure and focus on analyzing the performance of individual transport modes under traffic congestion. Additionally, the study centers on real-world problems, and the absence of public transport vehicles such as buses and trams on the campus further justifies their exclusion. While it is well established that high-capacity vehicles are effective under high-density traffic conditions, their inclusion would require detailed route planning and demand forecasting models, which are beyond the scope of this study. More comprehensive analyses integrating these modes of transport are planned for future research.

The arrival times given in the study are presented in Figure 7. There is no reward mechanism here. There are only total penalty times using the arrival time. It would be a useful way to find out which model is superior when no model is used, as well as which mode of transportation is more useful in various traffic scenarios. Figure 7 compares the arrival times of different transportation modes at different traffic densities. The horizontal line on the box in the graphs represents the median of the transportation times. These median values represent the middle value of the arrival times for each transportation mode when the starting points are randomly assigned. The figure shows that in low-density traffic scenarios (e.g., Figure 7A), the time it takes cyclists to reach their destination is generally shorter and the dispersion is narrower. However, as the traffic intensity increases (e.g., Figure 7D), the median values converge and the distribution widens across transport modes. This clearly demonstrates the impact of factors such as high traffic density and traffic lights on the efficiency of transport modes. For example, Figure 7D shows that in a very dense traffic environment with 1600 vehicles and pedestrians, the automobile is more advantageous and the median arrival time is lower than other modes. Overall, the median values show how the performance in transport times varies even when the origin and destination points are randomized, highlighting the importance of transport mode choice for different traffic scenarios.

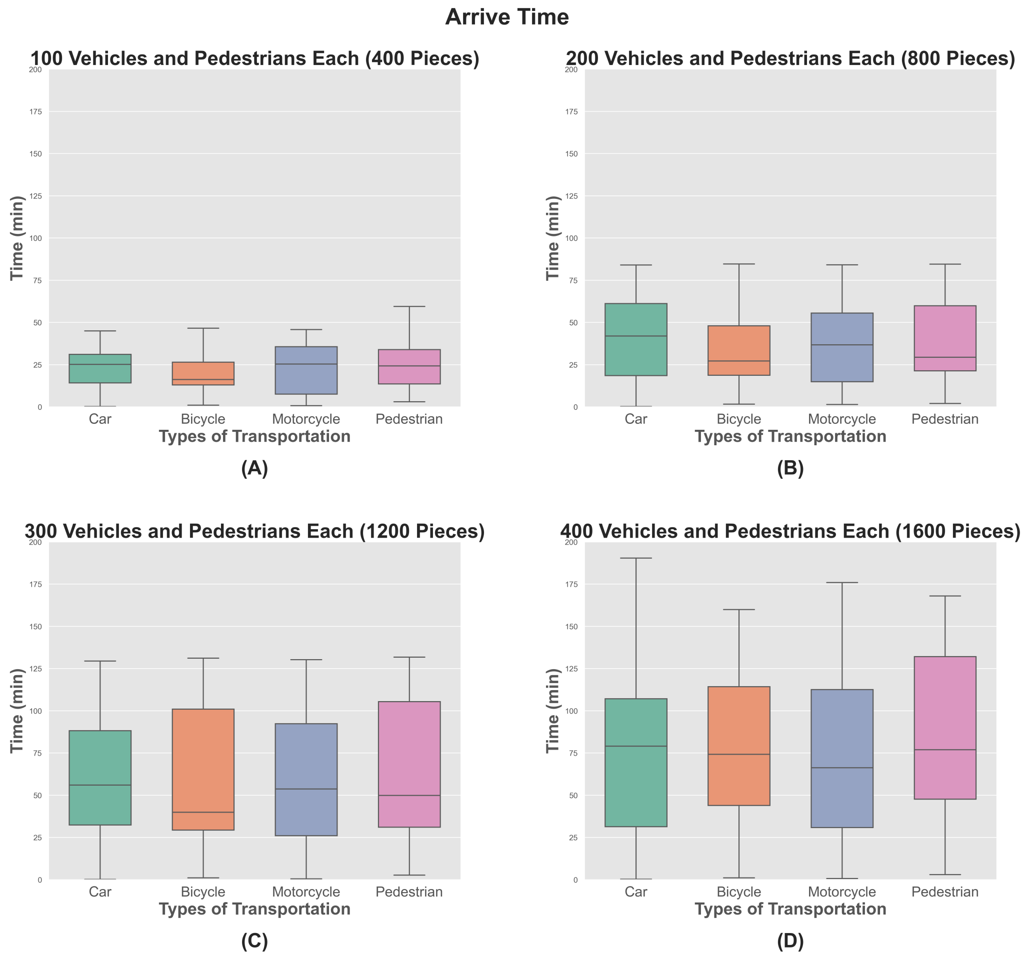

Figure 7. Arrival time distribution by transportation mode. Each subgraph shows the simulation results with different numbers of vehicles and pedestrians: (**A**) 100 vehicles and pedestrians (400 in total), (**B**) 200 vehicles and pedestrians (800 in total), (**C**) 300 vehicles and pedestrians (1200 in total), and (**D**) 400 vehicles and pedestrians (1600 in total). The median and range of arrival times for each transportation mode are shown.

The simulation ends when the transportation modes reach the target area after starting from random points. This evaluation is measured by the elapsed time. In order to make a better evaluation, an environment from low-density traffic to high-density traffic has been prepared. The average and general results for 100, 200, 300, and 400 vehicles are given in Figure 8. In low-density traffic environments, bicycle commuters' arrival times to their designated destination are significantly lower. However, as traffic density increases, the spread in target arrival times approaches. This is because in an environment with traffic lights and a high number of vehicles, transportation becomes extremely difficult. In the case of a very high density of vehicles, a total of 1600, it is seen that the best means of transportation is the car.

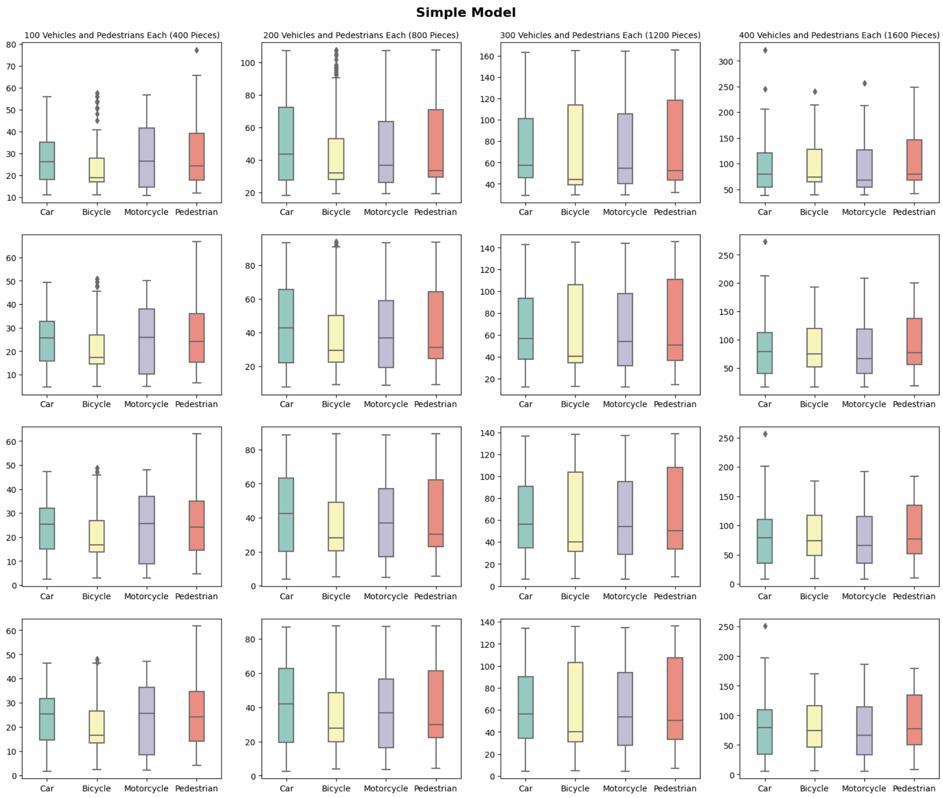

Figure 8. Arrival time for the simple method.

As shown in Table 2, the mean is the average value representing the general trend of the data. The maximum value is the highest value encountered in the dataset. The minimum value is the lowest value encountered in the dataset. Standard deviation expresses the distribution, spread, and variability of the data. The median indicates the middle value, i.e., the exact middle point, when the data are sorted. The mode indicates the most frequent value in the dataset. The table presents statistical summaries for vehicles, bicycles, motorcycles, and pedestrians for datasets of different sizes. The mean, maximum, minimum, standard deviation, median, and mode values/information are included for each dataset. The data generally show increasing variation and widening distribution as the cluster sizes increase. The increase in standard deviation indicates that the values have more variety as the datasets grow. This table provides a comprehensive perspective to compare and analyze the statistical properties of datasets of different sizes.

Table 2. Vehicle and pedestrian statistics across different dataset sizes.

	Mean	Maximum Value	Minimum Value	Standard Deviation	Median	Mode
Statistics of 100 vehicles and pedestrians (Total 400)						
Car	22.68	45.00	0.22	12.09	25.19	13.93
Bicycle	20.07	46.55	1.05	12.85	16.24	6.42
Motorcycle	23.49	45.77	0.67	13.99	25.38	6.15
Pedestrian	24.56	59.43	3.00	13.48	24.24	25.90
Statistics of 200 vehicles and pedestrians (Total 800)						
Car	40.76	84.01	0.20	23.91	41.97	77.45
Bicycle	33.75	84.62	1.60	22.28	27.22	12.43
Motorcycle	38.20	84.12	1.40	23.75	36.77	29.65
Pedestrian	38.39	84.52	2.05	23.60	29.39	27.27

Table 2. *Cont.*

	Mean	Maximum Value	Minimum Value	Standard Deviation	Median	Mode
Statistics of 300 vehicles and pedestrians (Total 1200)						
Car	61.04	129.48	0.18	35.55	56.05	48.32
Bicycle	55.93	131.18	1.03	38.70	39.89	10.70
Motorcycle	60.34	130.27	0.43	38.08	53.70	15.08
Pedestrian	63.89	131.80	2.63	39.19	49.87	130.12
Statistics of 400 vehicles and pedestrians (Total 1600)						
Car	75.96	240.51	0.27	45.69	79.08	37.67
Bicycle	76.34	159.93	1.03	43.51	74.20	17.38
Motorcycle	74.16	175.92	0.62	44.87	66.25	61.87
Pedestrian	87.42	168.05	2.97	47.80	76.95	34.42

5.1. Simple Reward

In Figure 8, the transportation times of 400, 800, 1200, and 1600 passengers are shown comparatively. Both average transportation times and distribution are given here. The differences between the results obtained using the simple reward mechanism and all transportation modes are visible. This is essential for comparing the transportation modes chosen for four different situations. In a simple reward model, we consider a mechanism in which late and early arrivals are penalized and on-time arrivals are rewarded. In this environment, at a low traffic density, the best means of transportation are cycling and, very close behind, cars. As the traffic density increases, the distribution in the system expands and randomness dominates. Moreover, as the traffic density increases, the automobile again becomes a better option. Likewise, in low-traffic environments, bicycle transportation leads to a better result. A graph of the system with the same number of vehicles (100, 200, 300, and 400) for each is given. In the study, 2, 5, 10, and 15 min late and early arrival cases are considered and their penalty points are given in the figure. Waiting for early arriving vehicles is as negative as being late because early arriving vehicles create a bad effect on the next plan (by waiting unnecessarily to go to the next transportation point).

5.2. Gaussian Distribution Reward

Figure 9 shows the results according to the Gaussian distribution reward. The results of 100 (A), 200 (B), 300 (C), and 400 (D) transportation types are displayed. Transportation times vary depending on the traffic density. Map distribution based on the distribution function and containing randomness reveals the results more accurately.

In the Gaussian-distributed reward mechanism, when distribution multipliers are applied as rewards to eliminate the problems caused by randomness, the average rewards are very close in low-traffic environments and the bicycle is still the best transportation option. With increasing traffic density, although the average rewards are very close, the motorcycle shows a better spread and achieves a better result in average rewards. Figure 9 shows the results according to the Gaussian Distributed Reward mechanism. The results for the 100 (A), 200 (B), 300 (C), and 400 (D) transportation types are presented. Transportation times vary depending on low-traffic-density and high-traffic-density conditions.

Regarding the mechanism, when distribution factors are applied as rewards to eliminate the problems caused by randomness, the average rewards are quite close to each other in low-traffic environments and the bicycle still stands out as the best transportation option. However, with the increase in traffic density, it is seen that although the average rewards are close to each other, the motorcycle shows a better distribution and achieves a better result in average rewards. In this context, when the statistics shown in Table 2 are examined in detail, it is clearly seen how the statistical differences, such as mean value and standard deviation, between vehicle types change with traffic density.

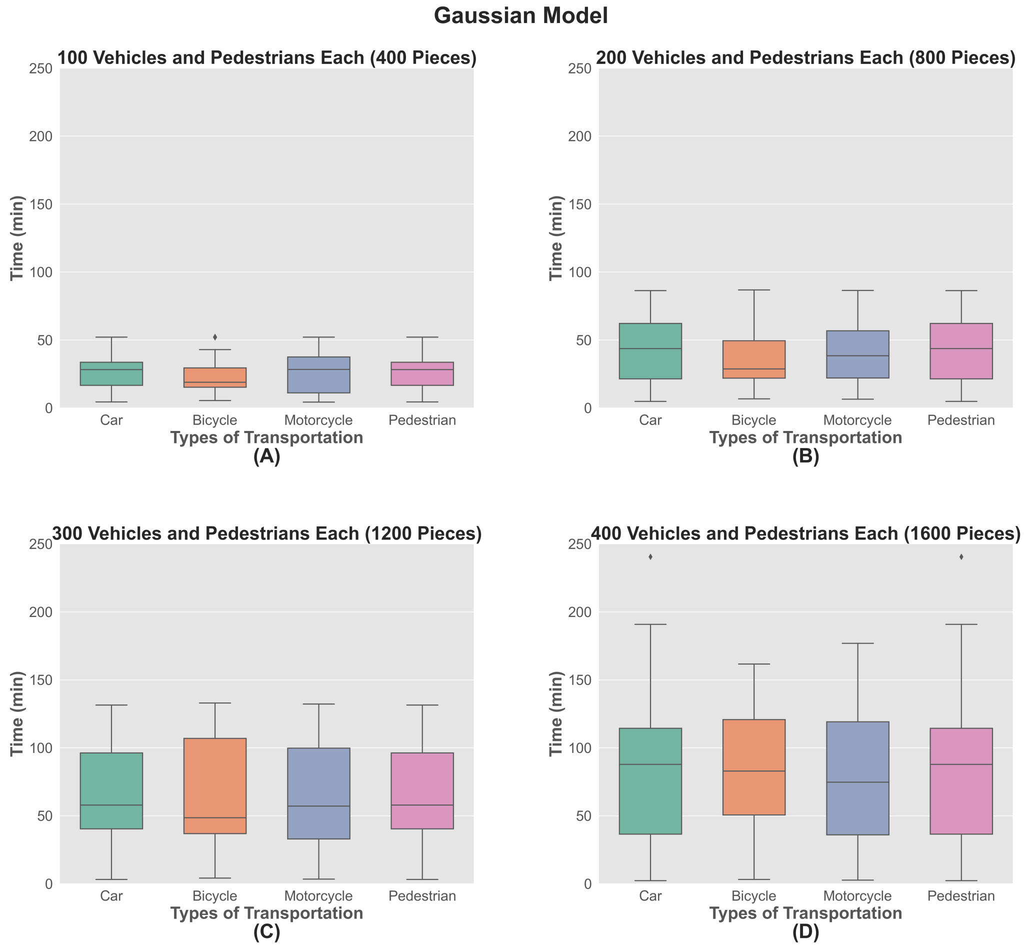

Figure 9. Arrival time distribution by transport mode according to a Gaussian distribution. Each subplot shows the simulation results with different numbers of vehicles and pedestrians: (**A**) 100 vehicles and pedestrians (400 in total), (**B**) 200 vehicles and pedestrians (800 in total), (**C**) 300 vehicles and pedestrians (1200 in total), and (**D**) 400 vehicles and pedestrians (1600 in total). The median and range of arrival times are shown for each transport mode.

For example, in the data of 100 vehicles and pedestrians (Total 400), bicycles and motorcycles have lower mean and standard deviation values. This indicates a more stable performance in a low-traffic density environment. In the data of 400 vehicles and pedestrians (Total 1600), it is seen that motorcycles show a better spread in average reward values and reach a higher standard deviation value. This supports the idea that motorcycles are a more advantageous option in heavy traffic environments. As a result, the performance differences of transportation types in both low- and high-traffic environments are supported by statistics, and it is seen that the Gaussian distribution reward mechanism reveals these differences fairly.

Table 3 summarizes the statistics based on the Gaussian Distribution Model for vehicles and pedestrians at different traffic densities. Statistical measures such as mean, maximum and minimum, standard deviation, median, and mode values are presented for four different transportation modes (car, bicycle, motorcycle, and pedestrian) in each dataset size (100, 200, 300, and 400 vehicles and pedestrians). The table shows the changes in these measures with the increase in traffic density. At a low traffic density, reward values are generally lower and standard deviations are smaller, indicating a more stable performance. It is observed that both mean and standard deviation values increase with the increase in traffic density, indicating that the performance variability increases in heavy traffic conditions. When comparing transportation modes, bicycles generally perform best at low traffic density, while motorcycles have an advantage in heavy traffic conditions. The Gaussian Distribution Model provides the opportunity to fairly evaluate the performance differences between different transportation modes by considering random reward distributions. This table provides a comprehensive perspective on understanding the effects of traffic density and transportation modes on performance.

Table 3. Vehicles and pedestrians statistics across different dataset sizes for a Gaussian distribution model.

	Mean	Maximum Value	Minimum Value	Standard Deviation	Median	Mode
Statistics of 100 vehicles and pedestrians (Total 400)						
Car	25.62	52.08	4.42	11.94	28.20	16.37
Bicycle	23.51	52.29	5.49	12.99	18.95	7.82
Motorcycle	26.37	52.16	4.31	14.32	28.38	7.52
Pedestrian	25.62	52.08	4.42	11.94	28.20	16.38
Statistics of 200 vehicles and pedestrians (Total 800)						
Car	44.37	86.41	4.87	23.67	43.65	81.47
Bicycle	36.80	86.89	6.73	22.12	28.76	21.91
Motorcycle	42.01	86.49	6.47	23.10	38.46	31.25
Pedestrian	44.37	86.41	4.87	23.67	43.65	81.48
Statistics of 300 vehicles and pedestrians (Total 1200)						
Car	66.39	131.46	3.17	35.61	57.89	49.32
Bicycle	61.53	133.00	4.12	37.84	48.58	15.16
Motorcycle	65.92	132.17	3.45	37.89	57.13	20.24
Pedestrian	66.39	131.46	3.17	35.60	57.89	49.32
Statistics of 400 vehicles and pedestrians (Total 1600)						
Car	81.87	240.53	2.26	45.89	87.86	43.53
Bicycle	82.41	161.69	3.08	43.35	82.95	20.94
Motorcycle	80.02	176.79	2.64	45.02	74.72	70.08
Pedestrian	81.87	240.53	2.26	45.89	87.86	43.53

5.3. Poisson Distribution Reward

Figure 10 shows the results according to the Poisson Distribution Reward. The results of 100 (A), 200 (B), 300 (C), and 400 (D) transportation types are displayed. Since the distribution formula gives sharper results, arriving at the destination on time is more rewarded, while arriving late is more penalized. For this reason, it has become preferable as a sharper reward mechanism.

Compared to the Gaussian method, the Poisson Distribution Reward Mechanism is stricter. Cycling in low-density traffic appears to be generally preferable. When distribution multipliers are used as incentives in the setting where they are used to mitigate the problems introduced by randomness, the average rewards are quite similar at low-traffic locations, and cycling is still considered the best mode of transportation. As traffic density increases, the average rewards are very similar, but the motorcycle has a better dispersion and produces a superior average reward result. Where traffic congestion increases, average arrival times are very similar. Heavy traffic congestion has a significant impact on the time it takes each vehicle to reach its destination. Therefore, when making transportation mode selection, this is the representation of the mode that should be selected according to traffic congestion. In high-density traffic, the bicycle is the best mode, while in lower-density situations, the motorcycle is a better transportation mode.

Figure 10 shows the results according to the Poisson distribution reward mechanism. While the bicycle stands out as the best option among the transportation modes at a low traffic density, it is seen that motorcycles performs better as traffic density increases. The sharp reward mechanism of the Poisson distribution rewards on-time arrivals more and penalizes late arrivals more. Therefore, a significant difference is observed between transportation times as traffic density increases and the bicycle is considered a more suitable option in high-density traffic situations. While the Poisson mechanism provides an effective method to reduce the problems caused by randomness, it clearly shows that the choice of transportation mode should change depending on the traffic density.

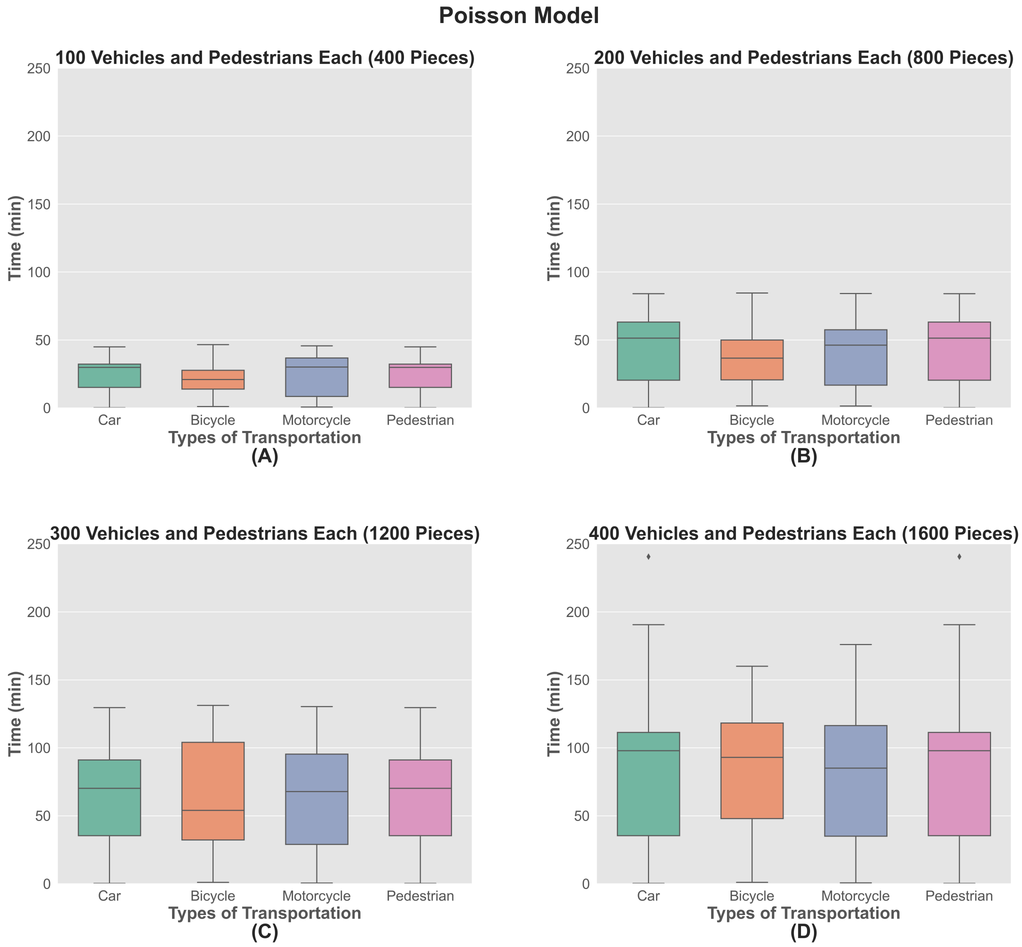

Figure 10. Arrival time distribution by transport mode according to Poisson distribution. Each subplot shows simulation results with different numbers of vehicles and pedestrians: (**A**) 100 vehicles and pedestrians (400 in total), (**B**) 200 vehicles and pedestrians (800 in total), (**C**) 300 vehicles and pedestrians (1200 in total), and (**D**) 400 vehicles and pedestrians (1600 in total). Median and range of arrival times are shown for each transport mode.

Table 4 shows the statistics obtained for vehicles and pedestrians under different traffic densities using the Poisson distribution reward model. The table includes statistical measurements such as mean, maximum, minimum, standard deviation, median, and mode values for four different transportation modes (car, bicycle, motorcycle, and pedestrian) in each dataset size (100, 200, 300, and 400 vehicles and pedestrians). The Poisson distribution reveals that the bicycle is the most efficient transportation mode at low traffic densities, while the motorcycle becomes a better option at high traffic densities. According to the table, while the reward values of both bicycle and motorcycle are distributed more steadily at low traffic densities, it is seen that motorcycle has a higher mean and a lower standard deviation at high traffic densities. This situation occurs because the Poisson reward mechanism rewards on-time arrivals more and penalizes late arrivals more severely. In addition, it is observed that the median and mode values for vehicles and pedestrians are generally low, indicating that the reward distributions are usually concentrated in a narrow range. As traffic density increases, a wider distribution is observed in other modes of transportation such as cars and pedestrians, which shows that the increase in traffic density has a significant effect on transportation time. The table shows that the Poisson distribution, with its sharp and fair reward mechanism, is a powerful tool for evaluating the performance of different modes of transportation depending on traffic density. In this context, the Poisson model clearly demonstrates that the choice of transportation mode should be based not only on average performance but also on the performance stability depending on traffic density.

Table 4. Vehicle and pedestrian statistics across different dataset sizes for the Poisson distribution model.

	Mean	Maximum Value	Minimum Value	Standard Deviation	Median	Mode
Statistics of 100 vehicles and pedestrians (Total 400)						
Car	24.35	45.02	0.22	12.34	29.88	0.22
Bicycle	21.74	46.57	1.06	12.63	20.94	1.06
Motorcycle	25.15	45.78	0.67	14.28	30.08	0.67
Pedestrian	24.35	45.01	0.22	12.34	29.88	0.22
Statistics of 200 vehicles and pedestrians (Total 800)						
Car	44.09	84.05	0.21	24.27	51.37	0.21
Bicycle	37.09	84.65	1.61	21.76	36.62	1.61
Motorcycle	41.53	84.15	1.41	23.82	46.17	1.41
Pedestrian	44.09	84.05	0.21	24.27	51.37	0.21
Statistics of 300 vehicles and pedestrians (Total 1200)						
Car	66.04	129.54	0.20	35.77	70.13	0.20
Bicycle	60.93	131.24	1.05	37.67	53.98	1.05
Motorcycle	65.34	130.33	0.45	38.14	67.78	0.45
Pedestrian	66.04	129.54	0.20	35.77	70.14	0.20
Statistics of 400 vehicles and pedestrians (Total 1600)						
Car	82.63	240.61	0.28	46.10	97.83	0.29
Bicycle	83.01	160.03	1.06	43.39	92.91	1.06
Motorcycle	80.83	176.00	0.64	44.93	85.00	0.64
Pedestrian	82.63	240.61	0.29	46.11	97.83	0.29

5.4. Results Comparison with Baseline Models

In this study, the performance of the proposed framework on different transportation modes is evaluated by comparing the simple model (SM), Gaussian distribution model (GDM), Poisson distribution model (PDM), and three different baseline models—B-coefficient (B-c), Random Utility Model (RUM), and Reinforcement Learning Reward Model (RLRM). Table 5 presents the average travel times of these models under different traffic density scenarios. According to the table, each model and transportation mode exhibited different performances under certain conditions, which revealed the importance of optimizing model selection according to traffic density and transportation mode. The results presented in Table 5 clearly show that the Gaussian (GDM) and Poisson (PDM) distribution models are superior to other models in terms of performance. GDM and PDM provided lower travel times than the simple model (SM), especially in low- and medium-density traffic scenarios, and their superiority was maintained at different traffic densities. At a high traffic density, it was observed that GDM and PDM produce optimized results in this mode thanks to the maneuverability and speed advantage of motorcycles. Bicycles are the most efficient transportation mode at low and medium traffic densities. They reduce the total travel time, with less dependence on traffic lights and flexible mobility. The GDM and PDM optimize the performance of bicycles under these conditions. In high-density traffic conditions, motorcycles have a superior performance compared to other modes. The GDM and PDM manage to keep the total travel time at the lowest level for this mode and flexibly adapt to the traffic density. Although the pedestrian mode is an effective option for short distances, it has a lower performance compared to other modes for long distances and at a high traffic density. However, the GDM and PDM are effective in optimizing pedestrian performance.

We have examined the performance of transportation modes under different traffic densities in detail and demonstrated the consistency of the GDM and PDM in different traffic conditions. In the future, we plan to expand the model further with real-time data integration, public transportation systems, and cost analyses. In addition, evaluation of environmental impacts and carbon emissions will enable the model to contribute to sustainable transportation policies. As a result, the proposed framework offers wide potential for both small-scale (e.g., campuses) and large-scale (urban areas) applications. The performance of Gaussian and Poisson distribution models under traffic density proves the adaptability and flexibility of the proposed system to different conditions. This framework can make significant contributions to the development of urban transportation policies in the future.

Table 5. Comparison of mean travel times across different models and categories.

Category	B-c	RUM	RLRM	SM	GDM	PDM
Car (100)	29.54	27.65	26.92	28.44	25.62	24.35
Car (200)	46.26	45.48	44.76	45.25	44.37	44.09
Car (300)	67.81	68.52	68.84	68.10	66.39	66.04
Car (400)	92.13	89.36	87.22	90.35	81.87	82.63
Bicycle (100)	25.35	24.37	23.64	24.28	23.51	21.74
Bicycle (200)	39.80	38.94	37.84	38.51	36.80	37.09
Bicycle (300)	67.16	66.81	68.93	66.85	61.53	60.93
Bicycle (400)	85.41	83.27	84.14	84.48	82.41	83.01
Motorcycle (100)	31.23	27.93	26.78	30.63	26.37	25.15
Motorcycle (200)	50.58	43.62	42.58	49.12	42.01	41.53
Motorcycle (300)	70.37	68.78	65.89	69.72	65.92	65.34
Motorcycle (400)	86.65	89.45	86.42	96.21	80.02	80.83
Pedestrian (100)	36.10	26.41	27.83	35.38	25.62	24.35
Pedestrian (200)	56.83	43.71	42.32	55.54	44.37	44.09
Pedestrian (300)	79.37	69.62	66.24	78.65	66.39	66.04
Pedestrian (400)	92.29	95.83	94.62	98.14	81.87	82.63

The proposed framework combines reinforcement learning algorithms and traffic simulations to provide a dynamic and adaptive approach to transportation mode selection. In terms of applicability, the framework is particularly suitable for traffic management and transport mode optimization in closed or semi-closed areas (e.g., university campuses or industrial zones). When tested at the campus scale, the model provided effective results by selecting the right transport modes under low- and medium-density traffic conditions. However, for the framework to be applied to large-scale urban areas, it needs to be extended to consider more complex traffic networks, user behavior, and environmental impacts. For example, the integration of real-time traffic data and the inclusion of public transportation systems in the framework will provide a wider potential for application. In terms of robustness, the proposed framework achieved a good performance in transportation mode selection in different traffic density scenarios. The Gaussian and Poisson reward mechanisms provided flexibility with regard to mode choice according to the traffic density and demonstrated the system's ability to adapt to unexpected traffic conditions. However, it was observed that the performance of the model may be limited in very high-density traffic environments. Therefore, in the future, additional enhancements are planned to improve the robustness of the model to various traffic anomalies (e.g., road closures and accidents) and changes in user preferences. In conclusion, the proposed framework is applicable to both small-scale and large-scale applications.

In this study, the performance of transport modes in different traffic density scenarios is analyzed. The results clearly show the advantages of each mode under different conditions:

- Low Traffic Density : Cycling stands out as the most efficient mode of transportation at a low traffic density. This is mainly because bicycle users can move quickly and flexibly in low-density scenarios, independently of other vehicles. As they are less dependent on traffic lights, bicycle users generally prefer direct routes, significantly reducing total travel time.
- Medium Traffic Density: In medium-traffic-density scenarios, the bicycle remains the most advantageous mode of transport. As traffic density increases, the congestion problems observed in other modes of transportation affect bicycle users less. The ability to move faster, especially at traffic lights, makes the bicycle stand out in this density range.
- High Traffic Density: In high-traffic-density scenarios, the motorcycle is the most efficient mode of transportation. This is because the motorcycle can move faster both in congested traffic conditions and at traffic lights. Thanks to its high maneuverability, the motorcycle can more effectively overcome obstacles in the flow of traffic, which reduces the total travel time. In peak traffic scenarios, the motorcycle's flexibility and mobility give it an advantage over other modes.
- Pedestrian Performance: Although the pedestrian mode is an effective option for short-distance travel, it is less effective than other modes in heavy traffic scenarios and over long distances. Traffic lights and distance significantly increase pedestrian travel times.

The performance of transport modes is affected by a combination of factors such as traffic density, traffic lights, and mobility. The proposed model is able to dynamically respond to these factors and select the most appropriate transport mode in specific traffic scenarios. In particular, the superior performance of the bicycles in low- and medium-density traffic and the motorcycle in high-density traffic shows that the model can adapt to different conditions. In the future, performing more comprehensive evaluations will be made possible by incorporating additional factors such as user preferences, environmental impacts, and cost into the model.

6. Conclusions

This paper focuses on transportation mode selection. Particularly, issues with traffic congestion and employing the right type of transportation considering the specific environment are examined. The study uses RL as a solution for transportation mode selection. In this context, various approaches were tested by creating different reward models. These reward models included approaches such as simple reward, Gaussian Distribution Reward, and Poisson Distribution Reward. The study results show that traffic density is vital in transportation mode selection. Research has been carried out by creating various reward models to determine which mode of transportation should be preferred, and the results show that traffic density, in particular, determines which mode of transportation is preferred. In addition, travel time was highlighted as important in the decision-making mechanism. A mathematical basis was established by stating that starting early does not make a difference in terms of unnecessary waiting but does in terms of different traffic densities. This basis includes a target time as the average of the arrival times of all vehicles to make the mode choice usable in different scenarios and environments. While a bicycle is a very good means of transportation in low-density traffic, a motorcycle is a better option in heavy traffic. In conclusion, this study shows that a reinforcement learning-based approach to transportation mode selection can provide an effective solution, considering traffic density and travel time factors. Such models are expected to positively impact future urban planning and transportation strategies.

The Gaussian and Poisson reward mechanisms used in the study were chosen depending on the traffic density and dispersion characteristics. The Gaussian model is used to ensure a balanced distribution of rewards at low and medium traffic volumes. This model allows users to be rewarded more consistently in terms of travel time and speed when traffic conditions are more stable. The Poisson model is more suitable for high-density traffic. This model provides a more sensitive solution to extreme traffic conditions, such as congestion, by creating sharper and denser reward/penalty structures.

This study proposes a methodology for transportation mode selection at different traffic densities using reinforcement learning algorithms and traffic simulation. The obtained results highlight the effectiveness of transport modes such as bicycles and motorcycles, especially in low- and high-traffic-density scenarios. However, this study was carried out on a limited campus scale and scalability to real-world scenarios has some limitations. Real-world applications should be extended to include more complex traffic networks, user preferences, and environmental factors. For example, different city structures, larger-scale traffic networks, and the integration of public transportation modes will require an extended version of the method. Furthermore, factors such as the compatibility of the proposed model with real-time data and computational cost in large-scale simulations need to be evaluated. In the future, real-world data will be used to improve the accuracy of the model and test its applicability for large-scale scenarios. This will make the proposed framework applicable not only at the campus scale but also to traffic management problems in large cities.

In future studies, parameters that negatively affect traffic, such as waiting times at traffic lights, will also be added to the model, as well as many other variables such as possible alternative routes (instead of entering an area with a red light). Traffic density can be generated by adding more diverse distribution functions and more vehicle types. In addition, there will be a focus on integrating public transport options (e.g., buses and minibuses) into the proposed framework. This expansion will require the development of specific reward mechanisms to assess the performance of these modes in different traffic densities and scenarios. Furthermore, the integration of public transport modes will be tested in larger urban areas using real-world datasets, increasing the applicability of the model. This approach will enable a more comprehensive evaluation of the proposed system at both the campus scale and the city scale. In this way, it will be possible to obtain more comprehensive and clearer results.

Author Contributions: Conceptualization, M.B.H.T., K.Ö., I.S. and A.Y.; methodology, M.B.H.T., K.Ö., I.S. and A.Y.; software, M.B.H.T., K.Ö., I.S. and A.Y.; validation, M.B.H.T., K.Ö., I.S. and A.Y.; formal analysis, M.B.H.T., K.Ö., I.S. and A.Y.; investigation, M.B.H.T., K.Ö., I.S. and A.Y.; resources, M.B.H.T., K.Ö., I.S. and A.Y.; data curation, M.B.H.T., K.Ö., I.S. and A.Y.; writing—original draft preparation, M.B.H.T., K.Ö., I.S. and A.Y.; writing—review and editing, M.B.H.T., K.Ö., I.S. and A.Y.; visualization, M.B.H.T., K.Ö., I.S. and A.Y.; supervision, M.B.H.T., K.Ö., I.S. and A.Y.; project administration, M.B.H.T., K.Ö., I.S. and A.Y.; funding acquisition, M.B.H.T., K.Ö., I.S. and A.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Grant Number 22AG040. The authors give thanks to TUBITAK for their support.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Schmidt, R.; Weisser, H.; Schulenberg, P.; Goellinger, H. Autonomous driving on vehicle test tracks: Overview, implementation and results. In Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511), Dearborn, MI, USA, 5 October 2000; pp. 152–155.
2. Hu, J.; Bhowmick, P.; Arvin, F.; Lanzon, A.; Lennox, B. Cooperative control of heterogeneous connected vehicle platoons: An adaptive leader-following approach. *IEEE Robot. Autom. Lett.* **2020**, *5*, 977–984. [CrossRef]
3. Wang, S.; Tao, F.; Shi, Y.; Wen, H.J.S. Optimization of vehicle routing problem with time windows for cold chain logistics based on carbon tax. *Sustainability* **2017**, *9*, 694. [CrossRef]
4. Eto, R.; Sakata, K.; Yamakawa, J. Driving force distribution based on tyre energy for independent wheel-drive vehicle on rough ground. *J. Terramech.* **2018**, *76*, 29–38. [CrossRef]
5. Quinlan, J.R. Improved use of continuous attributes in C4.5. *J. Artif. Intell. Res.* **1996**, *4*, 77–90. [CrossRef]
6. Zenina, N.; Borisov, A.J.C.S. Transportation mode choice analysis based on classification methods. *Comput. Sci.* **2011**, *50*, 49–53. [CrossRef]
7. Frank, L.; Bradley, M.; Kavage, S.; Chapman, J.; Lawton, T.K. Urban form, travel time, and cost relationships with tour complexity and mode choice. *Transportation* **2008**, *35*, 37–54. [CrossRef]
8. Haydari, A.Y.; Yasin, Y. Deep reinforcement learning for intelligent transportation systems: A survey. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 11–32. [CrossRef]
9. Khamidehi, B.; Sousa, E.S. Distributed deep reinforcement learning for intelligent traffic monitoring with a team of aerial robots. In Proceedings of the IEEE International Intelligent Transportation Systems Conference (ITSC), Indianapolis, IN, USA, 19–22 September 2021; pp. 341–347.
10. Akter, T.; Alam, B.M. Travel mode choice behavior analysis using multinomial logit models towards creating sustainable college campus: A case study of the University of Toledo, Ohio. *Front. Future Transp.* **2024**, *5*, 1389614. [CrossRef]
11. Turay, S.S.; Ababio-Donkor, A.; Adams, C.A.; Massaquoi, A.B. Statistical modelling of travel mode choice of public university students in Freetown, Sierra Leone: The case of three campuses. *Urban Plan. Transp. Res.* **2024**, *12*, 2304589. [CrossRef]
12. Asiyah, N.; Damanik, S.; Ramadhan, R.J.V.J. Transportation mode choice to campus study in the new normal era (case study: Al-Azhar University). *J. VORTEKS* **2024**, *5*, 376–386.
13. Lv, Y.; Duan, Y.; Kang, W.; Li, Z.; Wang, F.Y. Traffic flow prediction with big data: A deep learning approach. *IEEE Trans. Intell. Transp. Syst.* **2014**, *16*, 865–873. [CrossRef]
14. Polson, N.G.; Sokolov, V.O. Deep learning for short-term traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 1–17. [CrossRef]
15. Belletti, F.; Haziza, D.; Gomes, G.; Bayen, A.M. Expert level control of ramp metering based on multi-task deep reinforcement learning. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 1198–1207. [CrossRef]
16. Wu, C.P.; Kanaad, N.; Kheterpal, N.; Dickstein, L.; Mehta, A.; Vinitsky, E.; Bayen, A.M. Framework for control and deep reinforcement learning in traffic. In Proceedings of the IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–8.
17. Yang, X.; Xu, Y.; Kuang, L.; Wang, Z.; Gao, H.; Wang, X. An information fusion approach to intelligent traffic signal control using the joint methods of multiagent reinforcement learning and artificial intelligence of things. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 9335–9345. [CrossRef]
18. Koh, S.S.; Zhou, B.; Yang, P.; Yang, Z.; Fang, H.; Feng, J. Reinforcement learning for vehicle route optimization in SUMO. In Proceedings of the IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 28–30 June 2018; pp. 1468–1473.
19. Hoen, K.M.R.; Tan, T.; Fransoo, J.C.; van Houtum, G.J. Effect of carbon emission regulations on transport mode selection under stochastic demand. *Flex. Serv. Manuf. J.* **2014**, *26*, 170–195. [CrossRef]
20. Wei, H.; Zheng, G.; Yao, H.; Li, Z. Intellilight: A reinforcement learning approach for intelligent traffic light control. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 2496–2505.
21. Liang, X.; Du, X.; Wang, G.; Han, Z. A deep reinforcement learning network for traffic light cycle control. *IEEE Trans. Veh. Technol.* **2019**, *68*, 1243–1253. [CrossRef]

22. Isele, D.; Cosgun, A.; Fujimura, K. Analyzing knowledge transfer in deep q-networks for autonomously handling multiple intersections. *arXiv* **2017**, arXiv:1705.01197.
23. Stampa, G.; Arias, M.; Sánchez-Charles, D.; Muntés-Mulero, V.; Cabellos, A. A deep-reinforcement learning approach for software-defined networking routing optimization. *arXiv* **2017**, arXiv:1709.07080.
24. Li, C.; Li, Y.; Zhao, Y.; Peng, P.; Geng, X. SLER: Self-generated long-term experience replay for continual reinforcement learning. *Appl. Intell.* **2021**, *51*, 185–201. [CrossRef]
25. Ruiz, E.; Soto-Mendoza, V.; Barbosa, A.E.R.; Reyes, R. Solving the open vehicle routing problem with capacity and distance constraints with a biased random key genetic algorithm. *Comput. Ind. Eng.* **2019**, *133*, 207–219. [CrossRef]
26. Van Hasselt, H.; Guez, A.; Silver, D. Deep reinforcement learning with double q-learning. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30, p. 1.
27. Pan, W.; Liu, S.Q. Deep reinforcement learning for the dynamic and uncertain vehicle routing problem. *Appl. Intell.* **2023**, *53*, 405–422. [CrossRef]
28. Li, M.; Qin, Z.; Jiao, Y.; Yang, Y.; Wang, J.; Wang, C.; Wu, G.; Ye, J. Efficient ridesharing order dispatching with mean field multi-agent reinforcement learning. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 983–994.
29. Li, X.; Shi, L.; Shi, Y.; Tang, J.; Zhao, P.; Wang, Y.; Chen, J. Exploring interactive and nonlinear effects of key factors on intercity travel mode choice using XGBoost. *Appl. Geogr.* **2024**, *166*, 103264. [CrossRef]
30. Qin, H.; Gao, J.; Wu, Y.J.; Yan, H. Analysis on context change and repetitive travel mode choices based on a dynamic, computational model. *Transp. Policy* **2019**, *79*, 155–164. [CrossRef]
31. Gu, Z.; Wang, Y.; Ma, W.; Liu, Z. A joint travel mode and departure time choice model in dynamic multimodal transportation networks based on deep reinforcement learning. *Multimodal Transp.* **2024**, *3.3*, 100137. [CrossRef]
32. Arentze, T.; Timmermans, H. Modeling learning and adaptation processes in activity-travel choice: A framework and numerical experiment. *Transportation* **2003**, *30*, 37–62. [CrossRef]
33. Modi, K.B.; Zala, L.B.; Umrigar, F.S.; Desai, T.A. Transportation planning models: A review. In Proceedings of the National Conference on Recent Trends in Engineering and Technology, Gujarat, India, 8–10 December 2011.
34. Qawasmeh, B.; Qawasmeh, S.; Al Tawil, A.; Qawasmeh, D. Estimation of trip-based generation models and calibration of mode choice models for the American travel behavior. *Open Transp. J.* **2024**, *18*, e26671212348473. [CrossRef]
35. Cttu, Y.L. *Combined Trip Distribution and Assignment Model Incorporating Captive Travel Behavior*; Transportation Research Board: Washington, DC, USA, 1990.
36. Barff, R.; Mackay, D.; Olshavsky, R.W. A selective review of travel-mode choice models. *J. Consum. Res.* **1982**, *8*, 370–380. [CrossRef]
37. Varotto, S.; Glerum, A.; Stathopoulos, A.; Bierlaire, M. Modelling travel time perception in transport mode choices. In Proceedings of the 14th Swiss Transport Research Conference, Ascona, Switzerland, 14–16 May 2014.
38. Abou-Zeid, M.; Scott, D.M. Current issues in mode choice modeling. *Transportation* **2011**, *38*, 581–585. [CrossRef]
39. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
40. Bellman, R. On the theory of dynamic programming. *Proc. Natl. Acad. Sci. USA* **1952**, *38*, 716–719. [CrossRef]
41. Behrisch, M.; Bieker, L.; Erdmann, J.; Krajzewicz, D. SUMO–simulation of urban mobility: An overview. In Proceedings of the SIMUL 2011, the Third International Conference on Advances in System Simulation, Barcelona, Spain, 23–29 October 2011; ThinkMind: Coimbatore, India, 2011.
42. Ali, N.F.M.; Sadullah, A.F.M.; Majeed, A.P.; Razman, M.A.M.; Zakaria, M.A.; Nasir, A.F.A. Travel mode choice modeling: Predictive efficacy between machine learning models and discrete choice model. *Open Transp. J.* **2021**, *15*, 1. [CrossRef]
43. Bierlaire, M. BIOGEME: A free package for the estimation of discrete choice models. In Proceedings of the Swiss Transport Research Conference, Ascona, Switzerland, 19–21 March 2003.
44. Nam, D.; Kim, H.; Cho, J.; Jayakrishnan, R. A model based on deep learning for predicting travel mode choice. In Proceedings of the Transportation Research Board 96th Annual Meeting, Washington, DC, USA, 8–12 January 2017; Transportation Research Board: Washington, DC, USA, 2017; pp. 8–12.
45. Codeca, L.; Cahill, V. Using deep reinforcement learning to coordinate multi-modal journey planning with limited transportation capacity. *SUMO Conf. Proc.* **2021**, *2*, 13–32. [CrossRef]
46. Chhikara, R.S.; Folks, J.L. The inverse Gaussian distribution as a lifetime model. *Technometrics* **1977**,*19*, 461–468. [CrossRef]
47. Consul, P.C.; Jain, G.C. A generalization of the Poisson distribution. *Technometrics* **1973**, *15*, 791–799. [CrossRef]