

Review

Investigating Routing in the VANET Network: Review and Classification of Approaches

Arun Kumar Sangaiah ^{1,2}, Amir Javadpour ^{3,4,*}, Chung-Chian Hsu ^{5,*}, Anandakumar Haldorai ⁶
and Ahmad Zeynivand ⁷

- ¹ International Graduate Institute of Artificial Intelligence, National Yunlin University of Science and Technology, Douliu 64002, Taiwan; aksangaiah@ieee.org
 - ² Department of Electrical and Computer Engineering, Lebanese American University, Byblos 13-5053, Lebanon
 - ³ Department of Computer Science and Technology (Cyberspace Security), Harbin Institute of Technology, Shenzhen 518057, China
 - ⁴ ADiT-Lab, Electrical and Telecommunications Department, Instituto Politécnico de Viana do Castelo, 4900-367 Viana do Castelo, Portugal
 - ⁵ Department of Information Management, International Graduate Institute of Artificial Intelligence, National Yunlin University of Science and Technology, Douliu 64002, Taiwan
 - ⁶ Department of Computer Science and Engineering, Sri Eshwar College of Engineering, Coimbatore 642109, India; anandakumar.h@sece.ac.in
 - ⁷ Department of Electrical & Computer Engineering, Tarbiat Modares University, Tehran 14115-111, Iran; ahmad.zeynivand@modares.ac.ir
- * Correspondence: a_javadpour@e.gzhu.edu.cn (A.J.); hsucc@yuntech.edu.tw (C.-C.H.)

Abstract: Vehicular Ad Hoc Network (VANETs) need methods to control traffic caused by a high volume of traffic during day and night, the interaction of vehicles, and pedestrians, vehicle collisions, increasing travel delays, and energy issues. Routing is one of the most critical problems in VANET. One of the machine learning categories is reinforcement learning (RL), which uses RL algorithms to find a more optimal path. According to the feedback they get from the environment, these methods can affect the system through learning from previous actions and reactions. This paper provides a comprehensive review of various methods such as reinforcement learning, deep reinforcement learning, and fuzzy learning in the traffic network, to obtain the best method for finding optimal routing in the VANET network. In fact, this paper deals with the advantages, disadvantages and performance of the methods introduced. Finally, we categorize the investigated methods and suggest the proper performance of each of them.

Keywords: machine learning; reinforcement learning; routing protocols; quality of service (QoS); Vehicular Ad Hoc Network (VANET); intelligent algorithm



Citation: Sangaiah, A.K.; Javadpour, A.; Hsu, C.-C.; Haldorai, A.; Zeynivand, A. Investigating Routing in the VANET Network: Review and Classification of Approaches. *Algorithms* **2023**, *16*, 381. <https://doi.org/10.3390/a16080381>

Academic Editor: Charalampos Konstantopoulos

Received: 11 May 2023
Revised: 25 July 2023
Accepted: 27 July 2023
Published: 7 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In this study, we review the methods presented in the field of routing in VANET network. The need to control the traffic network has led to many researches in this field. By categorizing the methods introduced in this paper, we can offer a suitable perspective and direction to those interested in this field. A lot of papers use fuzzy methods, regression based methods, linear and nonlinear methods, and also machine learning methods [1]. Recently, unsupervised and supervised methods as well as different RL methods have been used in this field.

Several studies have used reinforcement learning methods to optimize network throughput and improve latency, packet delivery rate, and quality of service in VANETs [2]. Fuzzy logic algorithms have been used to find the most optimal path in the VANET network and improve network performance [3]. Also, machine learning methods such as Q learning have been used to increase the efficiency of the vehicle network in terms of reducing delay and improving the packet delivery ratio [4]. Furthermore, deep reinforcement learning

algorithms have been used to improve VANET performance by predicting vehicle speed and position and identifying the most appropriate route [5]. The use of a centralized SDN controller as a learning agent for VANET routing has also been investigated, and recent methods including the use of satellites and drones to achieve better routing and convergence in VANET networks have been investigated [6].

Various researches have been conducted to evaluate and classify routing problems in this field, focusing on further analyzing the potential of reinforcement learning methods and artificial intelligence [7]. In some researches, the combination of fuzzy logic and reinforcement learning approaches has been effective to improve vehicle routing in VANET network [8]. Further analysis has been done on the effectiveness of fuzzy logic and reinforcement learning methods in network [9].

In addition, the use of multi-agent reinforcement learning techniques for traffic flow optimization has been promising [10]. Another study has investigated and classified routing and scheduling methods for emergency vehicles, including ambulance, fire trucks, and police, in VANET networks [11]. These studies provide valuable insights into the challenges and potential solutions for achieving optimal routing in VANETs.

The goal of this research is to propose the most suitable method for achieving optimal routing in VANETs. The advancement of technology in the automotive industry, including the development of connected and autonomous vehicles and computing, has created a need for optimal use of resources. One way to achieve this is through automotive cloud computing, which can perform computing tasks from the edge or remote cloud.

VANET is a subclass of mobile systems in which nodes are constantly moving and does not depend on a specific infrastructure. Routing protocols in VANET networks are divided into vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) routing. The research addresses challenges such as security and privacy issues, energy management, scheduling between centers at the origin and destination, network congestion control, and multiple routing and scheduling in VANET. In addition, the routing protocols in VANET networks are reviewed in Figure 1.

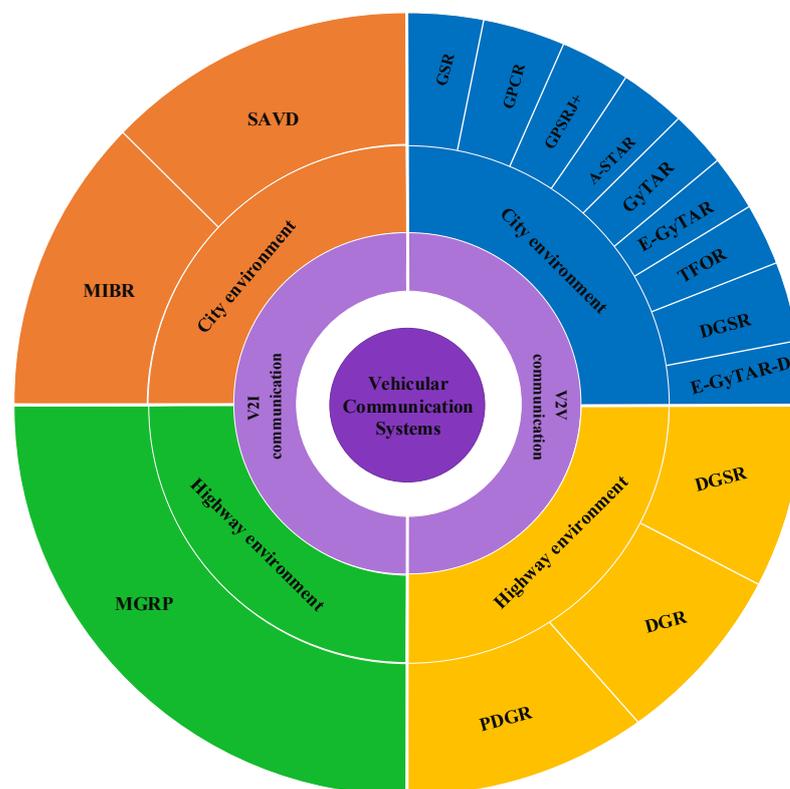


Figure 1. Classification of Routing Protocols.

The classification of position-based routing protocols is as follows:

- GSR: Geographic Resource Routing
- GPCR: Greedy Peripheral Coordinator Routing
- GPSRJ+: The Geographic Perimeter Stateless Routing Junction+
- A-STAR: Anchor-based Street and Traffic Aware Routing
- GyTAR: Greedy Traffic Aware Routing
- E-GyTAR: Enhanced Greedy Traffic Aware Routing
- TFOR: Traffic Flow Oriented Routing
- DGSR: Directional Greedy Source Routing
- E-GyTAR-D: Enhanced Greedy Traffic Aware Routing Directional
- GPSR: Greedy Perimeter Stateless Routing
- DGR: Directional Greedy Routing
- PDGR: Predictive Directional Greedy Routing
- SADV: Static-Node-Assisted Adaptive Data Dissemination in Vehicular Networks
- MIBR: Mobile Infrastructure-Based VANET Routing Protocol
- MGRP: Mobile Gateway Routing Protocol

There are many challenges such as security and privacy, energy, scheduling among centers at the origin and destination, network congestion control, and numerous routing and scheduling problems in the VANET network.

This research investigates traffic congestion control, optimal routing and network load reduction. The purpose of this study is to review various references in order to propose a suitable method for routing and load reduction in VANET network. The proposed method can be analyzed in several layers. In the first layer, according to the criterion of increasing network performance, our goals include increasing the quality of service, delay reduction and packet delivery ratio. In the second layer, a route is created at the origin and destination of the network, in which the cost and interaction of the routes between intersections are determined using RSUs. Through this algorithm, the best routes are selected based on reducing delay and increasing quality of service. In this layer, the speed of the most optimal routes is shown to the vehicles. In the third layer, the communication between servers, RSUs, traffic lights and vehicles in the environment is defined. Finally, an overview of routing in VANET network is drawn in

Significance of the Study

Cloud computing environments have limitations in extensive data processing systems. For example, nodes in a cloud environment encounter many clusters for calculations. Efficient task scheduling and resource allocation plans are required for fast data processing. These plans should distribute tasks on nodes in such a way that resource usage is maximized. For this purpose, we need task scheduling for data management. This research investigates scheduling and task scheduling with the goal of assigning tasks to appropriate resources and checking QoS in the network. A review paper is necessary for future research in various fields such as VANET, 5G, and 6G to choose the most appropriate methods and route.

Goals (Objectives)

This research aims to collect and classify related work in the fields of RL, FUZZY, and DRL in the VANET network to achieve the most optimal routing in the network. We classify the related work into four categories: reinforcement learning, fuzzy logic, overview, and deep reinforcement learning. The papers investigate criteria such as packet delivery ratio and delay, and one of the main goals is to study optimal routing and network overhead control. The areas of use of different protocols such as the type of routing, discovery of optimal routes, and quality of service (QoS) are evaluated. The study also investigates how to design and model different algorithms for routing in the network, and network efficiency for routing, network overhead control, and convergence speed.

The Overall Structure of the Paper

Section 2 presents related work, and reinforcement learning, fuzzy logic, and deep reinforcement learning methods are classified. Section 3 describes the proposals, and Section 4 is devoted to discussion and conclusion.

Figure 2 for a better understanding of this issue.

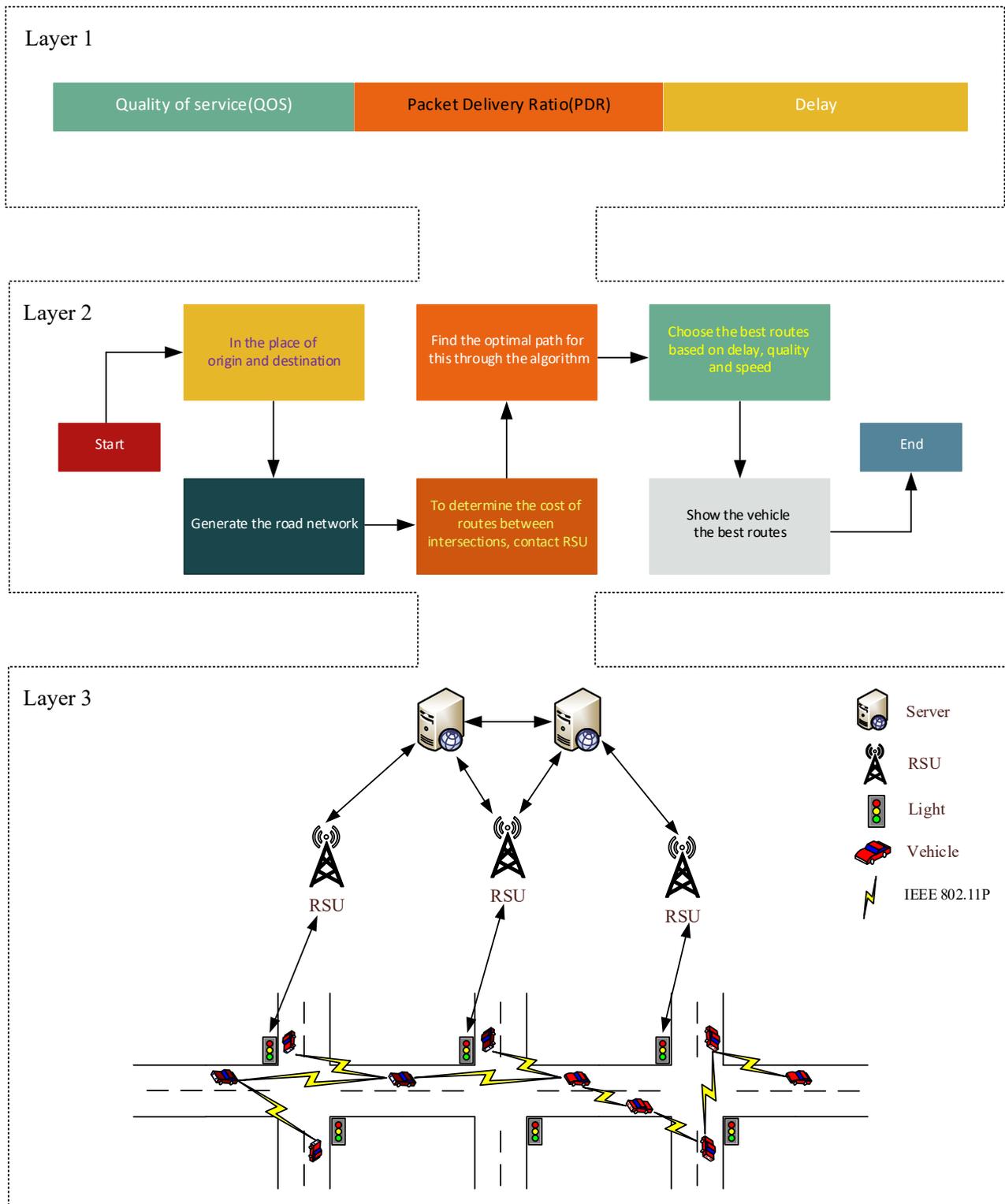


Figure 2. Routing in VANET network.

Machine Learning in VANET Networks

Machine learning includes reinforcement learning and algorithms in which a learning agent tries to maximize rewards in its environment to achieve a specific goal. By interacting with the environment, the learner selects appropriate actions to apply to the environment. In general, the RL agent’s objective is to maximize warehouses’ reward. An overview of machine learning in VANET networks is drawn in Figure 3 for better understanding.

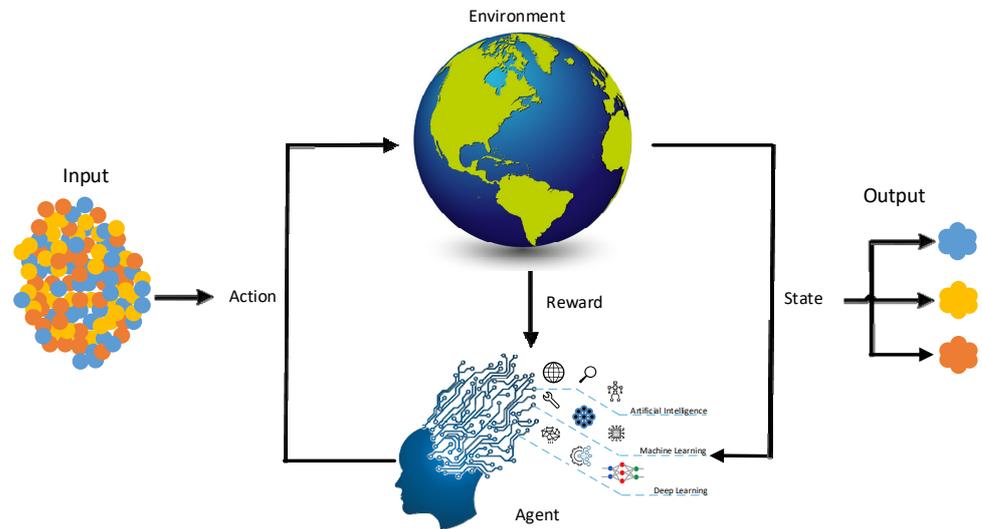


Figure 3. Machine learning.

In general, there are three approaches to reinforcement learning:

Value-oriented approach

In the value-oriented approach, the aim is to optimize the value function $V_{\pi}(s)$. The value function is a function that determines the maximum future reward that the agent receives in each state. The value of each state is equal to the total value of the reward that the agent can expect to gain in the future starting from that state.

$$V_{\pi}(s) = \mathbb{E}_{\pi} [R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | S_t = s]$$

Policy-oriented approach

Policy-oriented reinforcement learning aims to optimize the policy function $\pi(s)$ without using the value function. Policy is what determines the behavior of an agent at a given time. The agent learns a policy function. This helps him to map each situation to the best possible action.

$$a = \pi(s)$$

There are two types of policies:

Deterministic: The policy always returns the same action for a given state.

Stochastic (random): a probability distribution is considered for each action.

A stochastic policy, in which performing a specific action is conditional on a specific state, is defined as follows and Figure 4 shows finding random policies:

$$\pi(a, s) = \mathbb{P}[A_t = a | S_t = s]$$

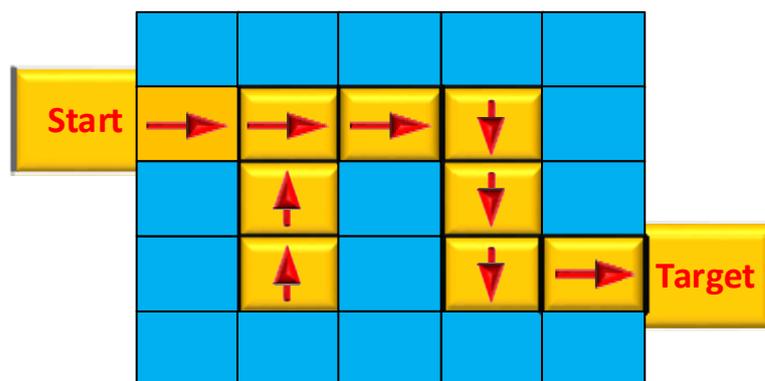


Figure 4. Finding stochastic policies.

Model-Oriented Approach

In model-oriented reinforcement learning, the environment is modeled, meaning that a model of the behavior of the environment is created. An important issue in this approach is providing a different model for each environment.

Supervised Learning Method

Supervised machine learning looks for a relationship among a series of functions that optimize the data cost function. For example, in the regression problem, the cost function can be the square of the difference between the forecast and the actual output values, or in the classification problem, the loss is equal to the negative logarithm of the output probability. The problem with learning neural networks is that this optimization problem is no longer convex, and we face local minima.

One of the common methods of solving the optimization problem in neural networks is backpropagation. This method calculates the gradient of the cost function for all weights of the neural network and then uses gradient descent methods to find a set of optimal weights. Gradient decreasing methods try to alternately move against the direction of the gradient and thereby minimize the cost function. Finding the gradient of the last layer is simple and can be obtained using partial differentiation. However, the gradient of the middle layers cannot be obtained directly, and methods such as the chain rule must be used in differentiation. The backpropagation method uses the chain rule to calculate the gradients, starting from the highest layer and distributing them in the lower layers.

Unsupervised Learning Methods

Unsupervised learning has much more difficult algorithms than supervised learning because there is little information about the data or the results. In unsupervised learning, we look for items with which we can form groups or make clusters, estimate density, and reduce dimensions. Compared to these two types of learning, unsupervised learning has fewer tests and models used to ensure the model's accuracy. In supervised learning, the data is "labeled" and classified according to these labels. In unsupervised learning, the model is allowed to discover information, and this action is hidden from the human eye.

Artificial Intelligence-Based Methods

Artificial intelligence is the simulation and modeling of human intelligence processes by machines, including computer systems. By examining the environment, artificial intelligence takes actions that increase its chances of success. By planning artificial intelligence, we can achieve the desired goals by getting environmental rewards.

Most artificial intelligence algorithms have the ability to learn from data. These algorithms can reinforce themselves by learning from past achievements. Through managing infrastructure, artificial intelligence can control the health of servers, storage and network equipment, and check the health of systems and predict the time of equipment failure. In addition, workload management helps direct data automatically toward appropriate infrastructures at a specific time. Moreover, it can guarantee security and regularly check

network traffic, alerting experts when problems occur. Finally, Figure 5 shows an overview of the artificial intelligence for better understanding.

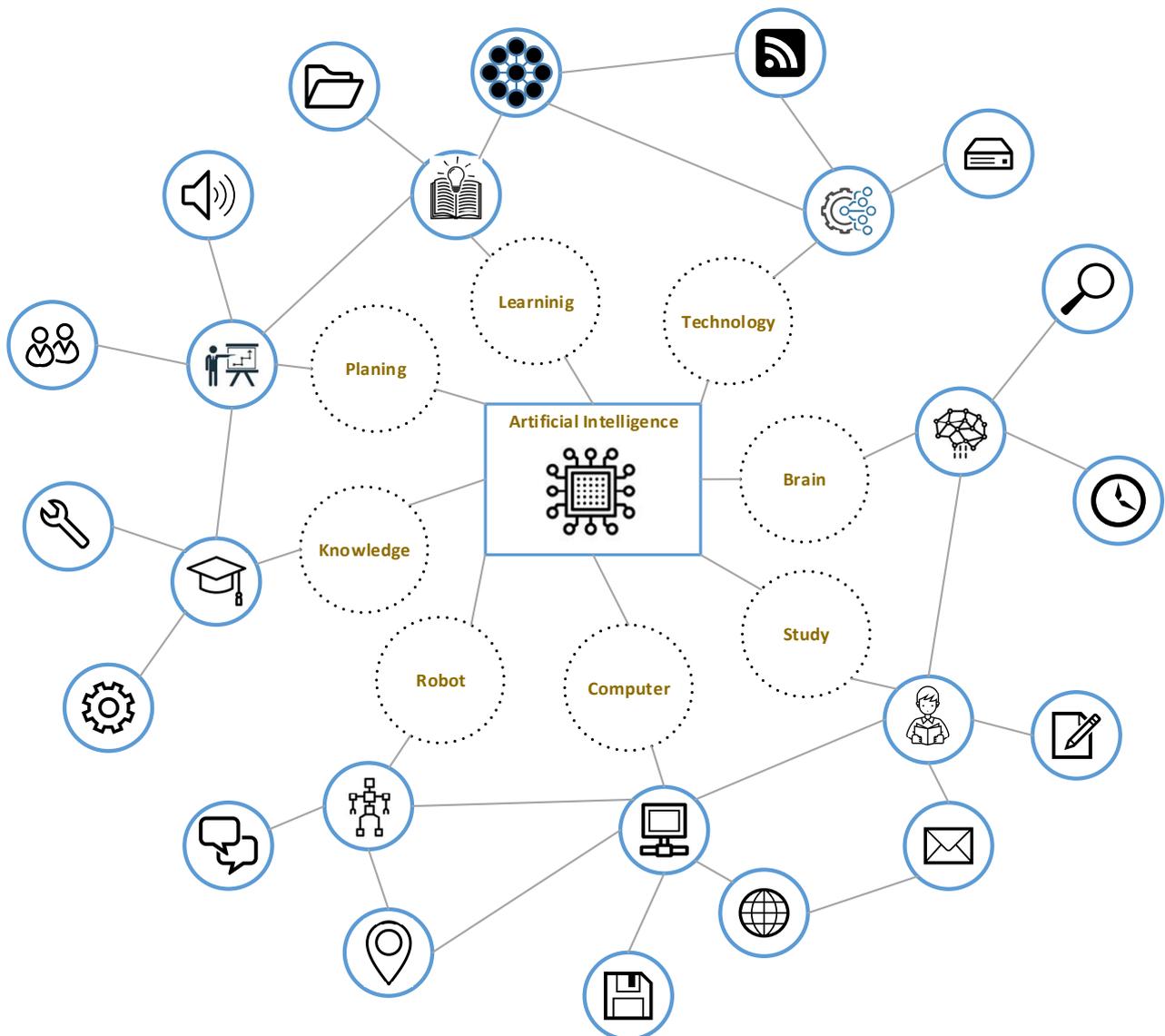


Figure 5. Artificial intelligence.

Neural Network Learning Methods

The neural network is trained through its inputs, including the input, hidden, and output layers. Neurons include a threshold value and an activation function. Our desired output is compared with the output of the neural network. The closer their values are, the lower the error and the more accurate the output. In one node, the input data is multiplied by a weight. The higher the weight, the greater the impact of the data. Then, the sum of the data multiplied by their weight is calculated, and the total value obtained passes through an activation function to produce the output.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks, or CNNs, are a special type of neural network used for image recognition and classification. CNNs perform a mathematical operation called convolution, which is a linear operation. Convolutional networks are similar to neural networks, but they use convolution instead of general matrix multiplication.

Recurrent Neural Networks (RNN)

Recurrent Neural Networks have a recurrent neuron, and the output of this neuron returns to itself t times. They are a type of artificial neural network used for speech recognition, sequential data processing, and natural language processing. They can remember their previous input due to their internal memory and use this memory to process a sequence of inputs. In other words, recurrent neural networks have a recurrent loop that prevents the loss of previously acquired information, allowing this information to remain in the network.

Random Learning Methods (Random Forest)

Random Forest is considered a supervised learning algorithm. As its name suggests, this algorithm creates a random forest, which is actually a group of decision trees. The forest is usually created using the bagging method, where a combination of learning models increases the overall results of the model. In other words, Random Forest constructs multiple decision trees and merges them together to produce more accurate and stable predictions.

One advantage of Random Forest is that it can be used for both classification and regression problems, which make up the majority of current machine learning systems. Here, the performance of Random Forest for classification will be explained, as classification is sometimes considered the building block of machine learning. In the picture below, you can see two Random Forests made up of two trees.

2. A Review of Past Research (Related Work)

The studies reviewed in this research are divided into several tables: reinforcement learning, fuzzy logic, review articles, and deep reinforcement learning, respectively. First, a short summary of the studies is provided in the tables, and the simulation tools used to simulate them are also mentioned. In addition, the advantages and disadvantages of each article are discussed. Finally, the evaluation criteria are presented in Table 1.

Reinforcement Learning

Machine learning includes Reinforcement Learning and algorithms in which a learning agent tries to maximize rewards in its environment to achieve a specific goal. By interacting with the environment, the learner selects appropriate actions to apply to the environment. In general, the objective of the RL agent is to maximize the stock reward.

In 2022, Ankita Singh et al. [1] presented an optimization algorithm using reinforcement learning for wireless mesh networks (WMN). This algorithm uses reinforcement learning to find the shortest path with minimum delay and deliver the packets in the WMN. In fact, this algorithm aims to check and reduce the criteria of packet transmission ratio, packet delivery ratio (PDR), delay and network congestion. This algorithm was able to improve network throughput.

In 2022, Jingjing Guo et al. [2] provided a routing method using intelligent clustering. This algorithm includes clustering, regulating clustering policies and routing modeling. This method examined ultrasonic unmanned networks and improved delay, packet delivery ratio and quality of service in this network.

In 2022, Ming Zhao et al. [4] presented a data transmission method based on the Manhattan model and used Temporal Convolutional Network (TCN) and Reinforcement Learning based Genetic Algorithm (RLGA) to improve the network load. This method solved the routing problem and reduced data transmission time and delay in the network.

In 2022, Arbelo Lolai et al. [5] proposed a reinforcement learning-based routing method I for VANETs. They also used Machine learning methods such as Q-learning to improve the structure of reinforcement learning based routing. Results showed that this method improved the efficiency of the vehicle network in terms of delay reduction and packet delivery ratio.

In 2022, Kazim Ergun et al. [6] investigated solving routing problems using the IoT systems reinforcement learning.

In 2018, Liang Xiao et al. [7] defined a VANETs network using aerial vehicles such as drones and then analyzed and controlled it using artificial intelligence and reinforcement

learning methods. The actions and reactions between drones are formulated as a game theory. According to this formulation, the interference of the Nash balance of the drone game is calculated in order to determine the cost of the optimal transmission of Drones.

In 2020, Jinqiao Wu et al. [8] designed a RSU strategy based on Q-learning as a routing mechanism for traffic network. The method had the required efficiency for routing in VANETs and also reduced the density and delay in this network.

In 2018, Fan Li et al. [9] focused on hierarchical routing in VANETs using reinforcement learning. This routing was done using Q-Grid algorithm and improved the packet delivery ratio. Q-Grid consists of Q-learning algorithms and network-based routing. In this method, Q-values are obtained according to the traffic flow and density of the network, and then, using Markov's law, optimal policies are chosen. This mechanism reduced the delay and network load.

In 2010, Celimuge Wu et al. [10] proposed a distributed reinforcement learning algorithm named Q-Learning AODV for routing in VANETs, in which a Q-Learning pattern is used to obtain VANETs network state data. This method reduces the length of the route and route discovery to choose the best route to deliver data. According to the research results, this mechanism significantly efficiently routed in VANETs network.

In 2018, Jinqiao Wu et al. [12] proposed a routing mechanism named ARPRL for VANETs using reinforcement learning methods. They used the Q-learning pattern to achieve optimal routing in the VANETs. In addition, MAC layer has been used to develop Q coordination with VANETs network. Finally, this method had the necessary efficiency in terms of packet delivery ratio, delay reduction, and selection of suitable routes for VANETs network.

In 2016, Celimuge Wu et al. [13] presented a method that stores the data in VANETs by transferring data to vehicles before moving, and then uses a fuzzy algorithm to investigate reward recognition for vehicle speed and Bandwidth efficiency in order for selecting the carrier nod. In addition, reinforcement learning and clustering have been used to investigate future reward and data collection to improve the efficiency of VANETs.

In 2020, Xiaohan Bi et al. [14] proposed a routing mechanism using reinforcement learning and clustering methods for VANETs to improve network performance and reduce the number of routes.

In 2021, Long Luo et al. [15] proposed a V2X routing mechanism for intersections in VANETs using artificial intelligence and reinforcement learning methods. This mechanism was intended for real-time networks in which Q-learning was used to calculate routing table. this routing mechanism could control network overhead, packet delivery ratio, network density and delay.

In 2021, Benjamin Sliwa et al. [16] presented an algorithm for routing prediction in the traffic network using reinforcement learning methods.

In 2021, Chengyue Lu et al. [17] provided a reinforcement learning algorithm to control and reduce VANETs network congestion. This reinforcement learning algorithm has been investigated in a multi-agent way. In this method, an agent t receives a reward in real time. The reward depends on the previous interval and the interval after sending reward. If the sending fails, a penalty will be charged. This method has led to reducing the delay and controlling the congestion of vehicles in the traffic network.

In 2021, Zhang et al. [18] have investigated vehicle-to-vehicle (V2V) routing in VANETs. By combining fuzzy logic and reinforcement learning approaches, they have improved network performance, increased packet delivery ratio, and reduced delay. Also, dynamic management of VANET network and adaptability to network changes are also investigated.

In 2022, Hasanain Alabbas et al. [19] presented a new gateway selection mechanism to find the best mobile gateway for vehicles that require Internet access. In this method, there is a system with two cloud servers, the first collects the necessary information about CVs and MGs, collects the necessary information about CVs and MGs, and the second uses the data to train the agent.

In 2022, Nitika Phull et al. [20] combined game theory and reinforcement learning methods to improve routing in VANETs. The game theory was used to classify vehicles. and select a leader for the group. A reinforcement learning method (K-means) is also applied to expand the clusters. This proposed method causes better and more stable routing.

In 2022, J. Aznar-Poveda et al [21] dealt with traffic light control and traffic network congestion reduction using reinforcement learning techniques. The traffic network control problem is formulated using the Markov algorithm in this study. Problem-solving process and optimal actions are performed using reinforcement learning. Faster convergence, packet delivery, and optimal routing indicated the excellent performance of the algorithm.

Table 1. Review of articles that have used reinforcement learning methods to control the VANET network.

Refs	Short Description	Simulation	Advantages	Disadvantages	Evaluation Criteria				
					Waiting Time	Queue Length	Travel Time	Delay Time	PDR
[1]	An optimization algorithm using reinforcement learning for wireless mesh networks (WMN) is presented	NS-3	<ul style="list-style-type: none"> Improve system performance High flexibility in dealing with dynamic environment 	<ul style="list-style-type: none"> complexity of calculations 				✓	✓
[2]	A routing method using intelligent clustering is proposed	OPNET	<ul style="list-style-type: none"> By communicating with each other, each agent performs only high-level subtasks it learns the value of a common abstract action. Reduce calculations 	<ul style="list-style-type: none"> High cost 				✓	✓
[4]	Network control is done using reinforcement learning and genetic algorithm.	--	<ul style="list-style-type: none"> Improve system performance High flexibility in dealing with dynamic environment 	<ul style="list-style-type: none"> complexity of calculations 				✓	
[5]	A routing method in VANETs based on reinforcement learning algorithms is proposed.	MATLAB	<ul style="list-style-type: none"> Improve system performance High flexibility in dealing with dynamic environment 	<ul style="list-style-type: none"> complexity of calculations 				✓	✓
[6]	IoT system has been investigated in terms of network routing using reinforcement learning.	NS-3	<ul style="list-style-type: none"> IOT enhances decision making Automate tasks Help improve service quality 	<ul style="list-style-type: none"> Collecting and managing data from all devices is challenging In case of failure of one of the devices, there is a possibility of failure of other devices as well. 				✓	✓
[7]	Using aerial vehicles such as drones, a network of VANETs has been defined and then the proposed method has been analyzed using artificial intelligence and reinforcement learning methods.	--	<ul style="list-style-type: none"> Improve network quality Reduce interference 	<ul style="list-style-type: none"> High cost 					
[8]	Designing a routing mechanism for traffic network using RSU and reinforcement learning	VanetMobiSim	<ul style="list-style-type: none"> Reduce network congestion Reduce delay 	<ul style="list-style-type: none"> High volume of calculations 				✓	✓
[9]	Hierarchical routing in VANETs has been addressed using reinforcement learning	--	<ul style="list-style-type: none"> communicating with each other, each agent performs only high-level subtasks it learns the value of a common abstract action. The hierarchical method improves mobility in the traffic network 	<ul style="list-style-type: none"> High cost 				✓	
[10]	A distributed reinforcement learning algorithm is proposed for routing in VANETs.	NS-2	<ul style="list-style-type: none"> Improve system performance High flexibility in dealing with dynamic environment 	<ul style="list-style-type: none"> Complexity of calculations 					✓

Table 1. Cont.

Refs	Short Description	Simulation	Advantages	Disadvantages	Evaluation Criteria				
					Waiting Time	Queue Length	Travel Time	Delay Time	PDR
[11]	Using reinforcement learning methods, a routing mechanism for VANETs is proposed	QualNet	<ul style="list-style-type: none"> It is an easy learning method. You can rely on the experience and goodwill of the instructor. 	<ul style="list-style-type: none"> The quality of learning is influenced by the performance of learners. It is difficult to get the correct representation. Some things cannot be learned by imitation. 				✓	✓
[13]	fuzzy algorithms and reinforcement learning have been used to control VANETs network.	NS-2	<ul style="list-style-type: none"> Flexible implementation and simplicity of the algorithm The possibility of simulating human logic and thinking The possibility of creating two solutions for one problem 	<ul style="list-style-type: none"> Choosing the membership function and basic rules are the most difficult parts of creating fuzzy systems. 				✓	
[14]	A routing mechanism has been proposed for VANETs using reinforcement learning and clustering methods	Python	<ul style="list-style-type: none"> communicating with each other, each agent performs only high-level subtasks it learns the value of a common abstract action. 	<ul style="list-style-type: none"> High cost 				✓	✓
[15]	A V2X routing mechanism for intersections in VANETs is proposed using artificial intelligence and reinforcement learning methods, in which real-time networks were focused.	OMNeT++ and SUMO	<ul style="list-style-type: none"> Simultaneous control for fast processing High security and reliability in real-time systems that perform critical and sensitive tasks Predictability and guarantee of doing things 	<ul style="list-style-type: none"> Real-time systems are large and complex In these systems, it is not allowed to combine hardware and software of real-time systems. 				✓	✓
[16]	An algorithm for predicting routing in the traffic network using reinforcement learning methods is presented.	OMNeT++	<ul style="list-style-type: none"> It allows the optimization of the current time slot considering the future time slot the ability to predict future events and take appropriate actions. 	<ul style="list-style-type: none"> The prediction algorithm needs a precise model of the process, because in this controller, the future behavior of the system must be predicted in the first step. 				✓	✓
[17]	A reinforcement learning algorithm is presented to control and reduce VANETs network congestion.	3GPP TR	<ul style="list-style-type: none"> This review will play an important role in future studies for reinforcement learning based routing algorithms. 	<ul style="list-style-type: none"> The need for a large volume of data 				✓	✓
[19]	A new gateway selection mechanism is proposed to find the best mobile gateway for vehicles that require Internet access.	SUMO	<ul style="list-style-type: none"> Reduce network delay 	<ul style="list-style-type: none"> high cost 					
[20]	Combining game theory and reinforcement learning methods, an algorithm for improving routing in VANETs is presented.	--	<ul style="list-style-type: none"> Game theory provides the possibility of framing strategic decisions. This theory can predict the outcome of competitive conditions Identify optimal strategic decisions. 	<ul style="list-style-type: none"> Not considering the element of risk and the opponent's activities to defeat us 					
[21]	By using reinforcement learning techniques, traffic light control and traffic network congestion reduction have been done.	SUMO	<ul style="list-style-type: none"> Improve system performance High flexibility in dealing with dynamic environment 	<ul style="list-style-type: none"> Complexity of calculations 				✓	✓
[22]	Traffic flow control using multi-agent reinforcement learning (MARL) method and coordination between traffic lights at intersections	SUMO	<ul style="list-style-type: none"> Improve traffic flow efficiency Reduce travel time Reduce traffic congestion Reduce queue length 	<ul style="list-style-type: none"> The need for a large amount of data 		✓		✓	✓

Table 1. Cont.

Refs	Short Description	Simulation	Advantages	Disadvantages	Evaluation Criteria				
					Waiting Time	Queue Length	Travel Time	Delay Time	PDR
[23]	Presenting a Markov approach for Internet of Vehicles (IoV) transportation systems.		<ul style="list-style-type: none"> • Increase operational capacity • Reduce latency • Dynamic management of the IoV network 	<ul style="list-style-type: none"> • Coordination of vehicles is difficult • Complexity of Transition Markov Algorithm 				✓	

Fuzzy Logic

The optimal routing is obtained using a fuzzy constraint and Q-learning method in the VANET network. Fuzzy logic has four main parts which are introduced below. Then, Figure 6 shows the diagram of how to connect these parts.

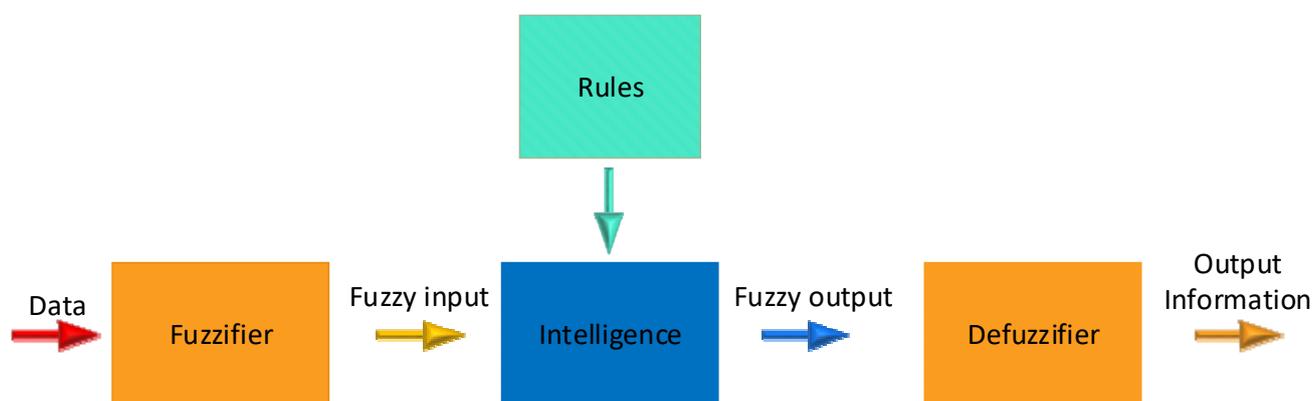


Figure 6. Fuzzy Logic.

Basic rules: This section contains all the rules and conditions specified as “if...then” by an expert to be able to control the “decision-making system”. According to the new methods in the fuzzy theory, it is possible to adjust and reduce the rules and regulations so that the best results can be obtained with the least number of rules.

Fuzzification: In this step, inputs are converted into fuzzy information. This means that the numbers and information to be processed will be converted into fuzzy sets and fuzzy numbers. The input data, for example, measured by sensors in a control system, are changed in this way and prepared for fuzzy logic-based processing.

Inference engine or intelligence: This section determines the degree of compliance of the inputs obtained from fuzzification with the basic rules. Based on the compliance percentage, different decisions are made as the results of the fuzzy inference engine.

Defuzzification: In the last step, the results of fuzzy inference, which are in the form of fuzzy sets, are converted into quantitative and numerical data and information. At this stage, you make the best decision according to the outputs, including different decisions with different compliance percentages. Usually, this choice will be based on the highest degree of compliance.

In 2022, Omid Jafarzadeh et al. [3] provided a routing algorithm based on multi-agent reinforcement learning (MARL) and fuzzy methods for VANETs. They also utilized Mamdani method to derive the fuzzy model. Metrics such as packet delivery ratio and delay have been improved.

Review articles

In 2019, Zoubir Mammeri et al. [11] reviewed and categorized articles in the field of routing in networks using reinforcement learning.

In 2021, Rezoan Ahmed Nazib et al. [24] reviewed studies on optimal routing in VANETs using artificial intelligence and reinforcement learning methods.

In 2021, Sifat Rezwan et al. [25] reviewed, categorized and compared studies in the field of reinforcement learning and artificial intelligence applications in flight networks. In addition, by examining the problems and ideas of various articles, they help researchers in conducting new research.

In 2022, Miri et al. [23] used Markov for Internet of Vehicles (IoV) systems, which finally proposed a resource interpretation scheme using Time Division Markov Multiple Access (TDMA) algorithm.

In 2022, Daniel Teixeira et al. [26] reviewed studies in the field of intelligent routing in VANETs. They also investigated the effectiveness of the algorithms used in the studies and evaluation of the routing methods proposed for the traffic network. According to this research, heuristic methods, fuzzy algorithms and reinforcement learning have the necessary efficiency to improve network performance. Also, Table 2 examines the advantages and disadvantages of the research that used the fuzzy method.

Table 2. Review of articles that have used fuzzy logic to control VANET network.

Refs	Short Description	Advantages	Disadvantages
[11]	Reviewing and categorizing studies in the field of routing in networks using reinforcement learning	The evaluation criteria have been reviewed.	The advantages and disadvantages of different articles are not discussed.
[23]	Presenting a Markov approach for Internet of Vehicles (IoV) transportation systems.	Dynamic management of the IoV network.	Complexity of Transition Markov Algorithm
[26]	Reviewing studies in the field of intelligent routing in VANETs	Different studies on machine learning are categorized, which provides researchers with an outlook for future work.	Deep learning methods are not reviewed and each paper are not reviewed in more detail
[27]	This method compares PSO with two routing approaches, namely AODV and DSR, and examines the effectiveness of the PSO algorithm in terms of performance metrics such as packet delivery ratio, end-to-end delay, and throughput.	The use of PSO algorithm, which has improved network performance, reduced delay and increased packet delivery ratio	The network is homogeneous and may not work in real-world scenarios. In addition, the PSO algorithm requires more iterations to reach convergence.
[28]	This article explores the use of reinforcement learning (RL) to detect impersonation attacks in device-to-device (D2D) communication. It describes the design and implementation of an RL-based detection system and presents experimental results demonstrating its effectiveness. The article provides insights into the potential benefits of using RL for improving cybersecurity in D2D communication.	The article provides experimental results that demonstrate the effectiveness of the proposed RL-based system compared to existing methods in terms of accuracy and efficiency	The article does not evaluate the proposed system on real-world data, instead relying on simulation experiments and allenges related to scalability and continuous learning, which the authors briefly discuss.

In Table 3, different routing approaches in the VANET network, such as fuzzy logic and reinforcement learning methods, have been compared, and this comparison has been made according to the advantages, disadvantages and evaluation criteria.

Table 3. Comparing different routing algorithms in VANET.

Refs	Short Description	Simulation	Advantages	Disadvantages	Evaluation Criteria				
					Waiting Time	Queue Length	Travel Time	Delay Time	PDR
[3]	A Routing Algorithm Based on Reinforcement Learning and Fuzzy Methods for VANETs	SUMO	<ul style="list-style-type: none"> Solving complex problems using more effective and simpler solutions The structure of fuzzy logic systems is simple and understandable. 	<ul style="list-style-type: none"> Implementing fuzzy logic in common hardware requires multiple and time-consuming tests. Choosing the membership function and basic rules is one of the most difficult parts of creating fuzzy systems 				✓	✓
[17]	Presenting a multi-factor augmentation approach for mitigation in VANET network	NS2	<ul style="list-style-type: none"> Delay reduction 	<ul style="list-style-type: none"> Dependence on the availability and reliability of communication channels 				✓	✓

Deep reinforcement learning

In the Q-learning algorithm, with the increase of the data volume, limitations occur that this algorithm can no longer solve the problem. So, it is possible that the algorithm does not reach convergence. Deep learning uses a memory that obtains output based on approximations of Q values. From this memory, the data are randomly selected for training.

DRL uses deep neural networks to solve reinforcement learning problems, hence the word ‘deep’ is used in its name. Q-learning is considered a classical reinforcement learning and is different from Deep Q-learning in some ways.

In the first approach, traditional algorithms are used to build a Q table to help the agent determine the action in each state. In the second approach, a neural network estimates the reward based on the state, q value. In general, deep reinforcement learning helps us make better decisions faster. Also, to better understand deep reinforcement learning, an overview is drawn in Figure 7.

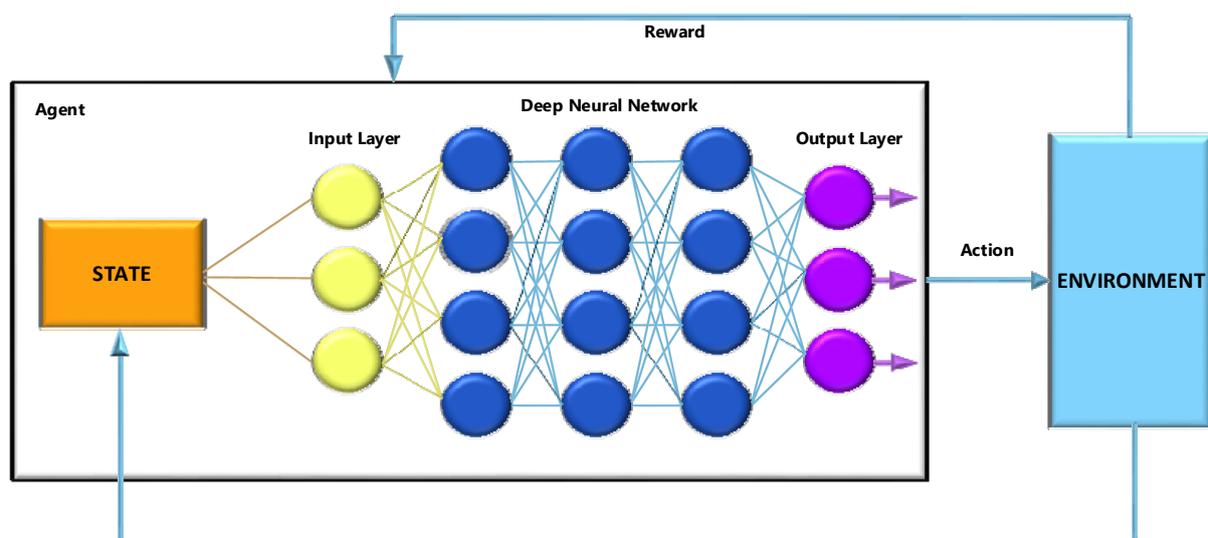


Figure 7. Schematic of reinforcement learning.

In deep reinforcement learning, a neural network is an Artificial Intelligence (AI) agent. The neural network interacts directly with its environment. The network observes the current state of the environment and decides which action to take according to its current state and past experiences (for example, move left, right, etc.) The AI agent may receive a reward or outcome based on the action the AI agent takes. The reward or outcome is

merely a scale. A negative outcome (for example -6) occurs if the agent performs poorly and a positive reward (for example $+6$) occurs if the agent performs well.

$$\mathbb{P}[S_{t+1}|S_t] = \mathbb{P}[S_{t+1}|S_1, S_2, S_3, \dots, S_t] \tag{1}$$

The quality of the action is directly related to the quantity of the reward and it predicts the probability of solving the problem (e.g., learning how to walk). In other words, an agent’s goal is to perform an action that maximizes the accumulated reward over time in any given situation.

Markov decision process

Markov decision process (MDP) is a time-discrete stochastic control process, the best approach we have so far for modeling the complex environment of an AI agent. Any problem that the artificial intelligence agent intends to solve can be considered as a sequence of states: $S_1, S_2, S_3, \dots, S_n$.

The agent performs actions and moves from one state to another. In the following, we show you the mathematics that determines which action the agent should perform in each situation.

Markov processes

Markov process (or Markov chain) is a stochastic (random) model that describes a sequence of possible states in which the current state depends only on the previous state. This feature is also called Markov. For reinforcement learning, it means that the next stage of an AI agent depends only on the last state and not on all previous states.

The Markov process is stochastic (random), which means that the transition from the current state (s) to the next state (s') occurs only with a certain probability ($P_{ss'}$)

$$P_{ss'}\mathbb{P}[S_{t+1} = s|S_t = s'] \tag{2}$$

$P_{ss'}$ is the entry in the state transition matrix P , which defines the transition probabilities from all states to all successor states (s').

$$p = from \begin{bmatrix} P_{11} & \dots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{n1} & \dots & P_{nn} \end{bmatrix} \tag{3}$$

Markov reward process

The reward Markov process is a tuple $\langle S, P, R \rangle$. Here, R is the reward that the agent expects to receive in state s . The AI agent is motivated by the fact that to achieve a particular goal (e.g., winning a game of chess), certain states (game configurations) are more promising than others in terms of strategy and potential for winning the game.

$$R_s = \mathbb{E}_\pi[R_{t+1}|S_t = s] \tag{4}$$

The main issue is the total reward (G_t) equation, which is the expected cumulative reward that the agent receives across the sequence of all states. The discount coefficient $\gamma \in [0, 1]$ measures each reward. We find discounting rewards mathematically convenient because doing so avoids infinite returns in cyclic Markov processes. In addition, the discount factor means that the further into the future we consider, the less important the rewards are to immediate satisfaction since the future is often uncertain. For example, if the reward is financial, immediate rewards may be more beneficial than delayed rewards.

$$G_t = R_{t+1}, R_{t+2}, R_{t+3} + \dots = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \tag{5}$$

Value function

Another important concept is the concept of the value function $v(s)$. This function draws a measure for each state 's'. The value of a state 's' is the total expected reward that the AI agent will receive if it starts progressing in state 's'.

$$v(s) = \mathbb{E}[G_t | S_t = s] \tag{6}$$

The value function can be divided into two parts:

- The agent receives the immediate reward R_{t+1} in state 's'.
- The discounted value $v(s)_{t+1}$ of the next state after the state 's'.

$$\begin{aligned} v(s) &= \mathbb{E}[G_t | S_t = s] \\ &= \mathbb{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | S_t = s] \\ &= \mathbb{E}[R_{t+1} + \gamma(R_{t+2} + \gamma R_{t+3} + \dots) | S_t = s] \\ &= \mathbb{E}[R_{t+1} + \gamma v(S_{t+1}) | S_t = s] \end{aligned} \tag{7}$$

Bellman equation for Markov reward processes

The decomposed value function is considered as the Bellman equation for Markov reward processes. This function is visualized using node graph. Starting at state 's' results in the value $v(s)$. By being in state 's', we have a certain probability (P_{ss}) to end up in the next state 's'. In this particular case, two next states may occur. To obtain the value of $v(s)$, the values of $v(s)$ of the possible next states weighted by the probabilities of P_{ss} should be added and then added to the immediate reward of being in state 's'. Also, for better visualization, a graph is drawn in Figure 8.

$$v_\pi(s) = \mathbb{E}[R_{t+1} + \gamma v_\pi(S_{t+1}) | S_t = s] \tag{8}$$

$$v(s) = R_s + \sum_{s' \in S} P_{ss'} v_\pi(s') \tag{9}$$

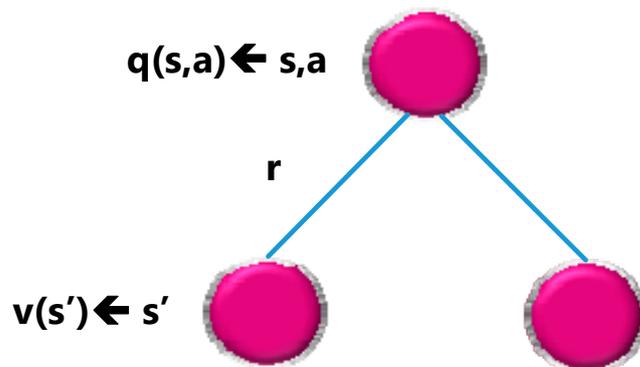


Figure 8. Visualization with node graph.

Markov decision process vs. Markov reward process

Markov decision process is a reward Markov process with a decision that is described by a set of tuples $\langle S, A, P, R \rangle$. In this example, A is the finite set of possible actions that the agent can take in state 's'. Therefore, the immediate reward of being in state 's' also depends on the agent's action in this state.

$$R_s^a = \mathbb{E}[R_{t+1} | S_t = s, A_t = a] \tag{10}$$

Policy

The agent decides what action to take in a particular situation. The policy determines the π of an action. Mathematically, a policy is a distribution over all actions given to a state, which determines the mapping of a state to an action that the agent should perform.

$$\pi(\mathbf{s}|\mathbf{a}) = \mathbb{P}[A_t = \mathbf{a} | S_t = \mathbf{s}] \tag{11}$$

In other words, we can describe the policy π as the agent’s strategy to choose specific actions depending on the current situation. This policy leads to a new definition of the value function -state $v(\mathbf{s})$.

$$v_\pi(\mathbf{s}) = \mathbb{E}_\pi[G_t | S_t = \mathbf{s}] \tag{12}$$

ACTION-VALUE Function

Another essential function is the action value function $q(\mathbf{s},\mathbf{a})$, which is the expected return obtained by starting in state ‘ s ’, performing action \mathbf{a} , and then following a policy π . Note that a state s , $q(\mathbf{s},\mathbf{a})$ can have multiple values because there are multiple actions that the agent can perform in state ‘ s ’. The neural network calculates $Q(\mathbf{s},\mathbf{a})$. Considering state ‘ s ’ as input, the network calculates the quality of every possible action in this state in a scalar form. Higher quality means better action according to the given goal. Finally, Figure 9 graph of the description of the action value function is drawn.

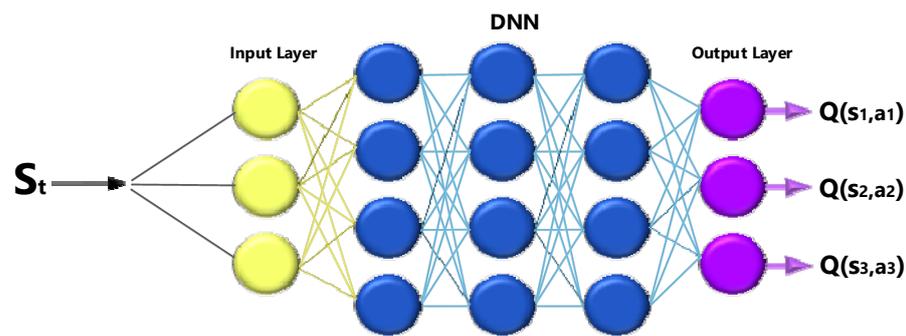


Figure 9. Action-value function description.

Remember that the Action-value function tells us how good it is to perform a specific action in a certain state.

$$q_\pi(\mathbf{s}, \mathbf{a}) = \mathbb{E}_\pi[G_t | S_t = \mathbf{s}, A_t = \mathbf{a}] \tag{13}$$

$$\downarrow$$

$$v_\pi(\mathbf{s}, \mathbf{a}) = \mathbb{E}[R_{t+1} + \gamma v_\pi(S_{t+1}) | S_t = \mathbf{s}] \tag{14}$$

$$\downarrow$$

$$q_\pi(\mathbf{s}, \mathbf{a}) = \mathbb{E}[R_{t+1} + \gamma q_\pi(s_{t+1}, a_{t+1}) | S_t = \mathbf{s}, A_t = \mathbf{a}] \tag{15}$$

By definition, performing a specific action in a specific state results in an action value $q(\mathbf{s},\mathbf{a})$. The value function $v(\mathbf{s})$ is the sum of possible $q(\mathbf{s},\mathbf{a})$ weighted by the probability (which is nothing but the policy π) of performing an action in state ‘ s ’. Figure 10 shows a visualization with a node graph.

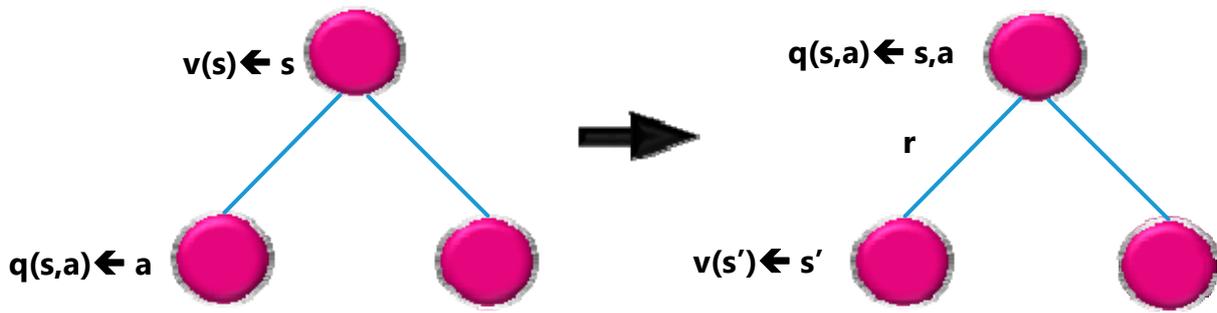


Figure 10. Visualization with node graph.

To get the action value, you need to take the discounted state values with the P_{ss} probabilities in all possible states (in this case just two states) and add the immediate reward:

$$q_{\pi}(s, a) = R_s^a + \sum_{s' \in S} P_{ss'} v_{\pi}(s') \tag{16}$$

$$q_{\pi}(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'} \sum_{a' \in A} \pi(a'|s') q_{\pi}(s', a') \tag{17}$$

This recursion relation can be visualized in another binary tree, with $q(s,a)$ ending in the next state 's' with a given probability P_{ss} , as an action 'a' is performed with the probability π and ends with the action value $q(s',a')$. Figure 11 shows the tree design with node graph.

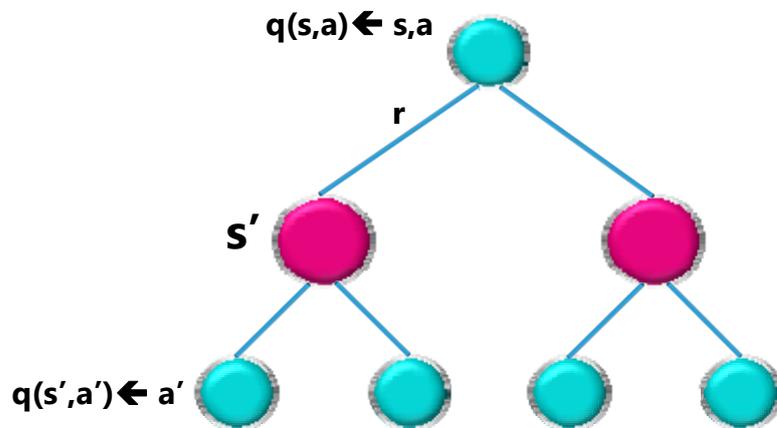


Figure 11. Drawing tree with node graph.

Optimal policy

The most important issue in deep reinforcement learning is to find the optimal action-value function q^* . Finding q^* means that the agent knows exactly the quality of an action in each given state. Additionally, the AI agent can decide which action to take based on the desired quality.

If we want to define q^* . The best possible action-value function is the one that follows the policy that maximizes the action values:

$$q_*(s, a) = \max_{\pi} q_{\pi}(s, a) \tag{18}$$

We need to maximize $q(s,a)$ to find the best possible policy. Maximization means to select only the action among all possible actions for which $q(s,a)$ has the highest value. This gives the following definition for the optimal policy π :

$$\pi_*(s, a) = \begin{cases} 1, & \text{if } a = \arg \max q_*(s, a) \\ 0, & \text{Otherwise} \end{cases} \tag{19}$$

Optimal Bellman equation

If the AI agent can solve this equation, it means that it has solved the problem in the given environment. The agent in any state or situation knows the quality of any possible action according to the goal and can act accordingly. Solving Bellman’s optimality equation will be the subject of future research in this field.

$$q^*(s, a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'} \max_{a'} q_*(s', a') \tag{20}$$

One of the DRL method’s limitations is that researchers must completely understand their environment. Not all systems can be made completely and without defects using deep reinforcement learning. In systems where the consequences of wrong decisions can be dire, deep reinforcement learning cannot work on its own. Table 4 examines the articles that are used to control the VANET network using deep reinforcement learning methods.

Table 4. Review of articles that have dealt with VANET network control using deep reinforcement learning methods.

Refs	Short Description	Simulation	Advantages	Disadvantages	Evaluation Criteria				
					Waiting Time	Queue Length	Travel Time	Delay Time	PDR
[29]	Using artificial intelligence methods and deep reinforcement learning algorithm, a routing mechanism for VANETs network is proposed.	NS-2	<ul style="list-style-type: none"> One of the advantages of deep learning is that features are automatically inferred 	<ul style="list-style-type: none"> High volume of calculations and data 				✓	✓
[30]	A deep reinforcement learning method for connected vehicle communication is proposed, in which a centralized SDN controller is used for VANETs routing.	OPNET	<ul style="list-style-type: none"> Deep learning is flexible to adapt to new problems in the future. 	<ul style="list-style-type: none"> Slowing down the learning process if the data volume increases 				✓	
[31]	Proposing intelligent congestion control for automotive communication systems with a deep reinforcement learning (DRL) approach.	SUMO	<ul style="list-style-type: none"> Reduce congestion Improving the allocation of communication resources Increase network efficiency 	<ul style="list-style-type: none"> Requires a significant amount of training data 				✓	

In 2019, M. Saravanan et al. [14] proposed a routing mechanism for VANETs using artificial intelligence methods and deep reinforcement learning algorithm. In this method, the vehicle’s speed and position is predicted and appropriate routing is chosen, which increases the performance of VANETs network. According to the results, this routing method has reduced the transmission delay and has been effective for routing in VANETs.

In 2018, Dajun Zhang et al. [18] proposed a deep reinforcement learning method for connected vehicle communication. Centralized SDN controller is also used for VANETs network routing. This controller has also been used as a learning agent to learn routing in VANETs network. In addition, a deep Q-learning algorithm has been used to determine the routing policy.

The comparison between routing protocols based on machine learning and fuzzy logic depends on the evaluation criteria used and the results obtained and the specific application in the VANET network. However, in general, machine learning-based methods are more popular due to their ability to learn from data and adapt to changing conditions

in a VANET network, while fuzzy logic methods are more widely used due to their ability to handle uncertainty and imprecision in the network.

In general, Methods based on machine learning, especially deep learning and reinforcement learning, can automatically learn complex algorithms and solve various problems in the VANET network, while methods based on fuzzy logic do not have this feature. On the other hand, machine learning-based methods may provide better performance than fuzzy methods due to the fact that they use previous data in the decision-making process. In addition, in some cases it is possible to use the combination of machine learning methods and fuzzy logic to cover the shortcomings of both methods.

Finally, choosing the appropriate method between machine learning and fuzzy logic depends on various factors such as the specific application, data, evaluation criteria, and specific constraints of the VANET application. When there are frequent changes in topology and traffic conditions in VANET, machine learning methods may be more suitable, while when VANET is exposed to uncertain or imprecise data, fuzzy logic methods may be more suitable. In Table 5, there is also a review of the articles that deal with VANET network control using fuzzy logic and machine learning methods.

Table 5. Review of articles that deal with VANET network control using fuzzy logic and machine learning methods.

Refs	Short Description	Simulation	Advantages	Disadvantages	Evaluation Criteria				
					Waiting Time	Queue Length	Travel Time	Delay Time	PDR
[32]	A machine learning-based routing scheme that makes decisions based on VANET network conditions such as vehicle speed, distance, and direction.	VEINS	<ul style="list-style-type: none"> Improved routing performance and throughput New approach and routing accuracy 	<ul style="list-style-type: none"> No comparison with other machine learning designs 				✓	✓
[33]	A review of the challenges in using machine learning in VANET network	-	<ul style="list-style-type: none"> Lack of accurate comparison of different ML models and techniques 	<ul style="list-style-type: none"> It examines different ML methods for traffic management in VANETs and discusses their advantages and disadvantages 					
[34]	Using machine learning for routing in VANET network	NS-3	<ul style="list-style-type: none"> A new approach uses decision tree and random forest algorithms to analyze the performance of communication protocols 	<ul style="list-style-type: none"> It assumes that the data collected from the simulations are representative of real-world scenarios 				✓	✓
[35]	A new approach to delay minimization using learning machine in VANET	NS-3	<ul style="list-style-type: none"> Using a short-term memory (LSTM) neural network to predict future locations of vehicles based on past trajectories 	<ul style="list-style-type: none"> The paper does not provide information on the computational complexity or resource requirements of the proposed approach 				✓	✓
[36]	Proposes a new approach to improve routing in VANET network using fuzzy logic	NS-3	<ul style="list-style-type: none"> Provide useful information about fuzzy logic controller architecture and evaluation criteria 	<ul style="list-style-type: none"> The paper does not provide information on the computational complexity for the proposed approach 				✓	✓
[37]	A fuzzy logic based routing in VANET network is presented	-	<ul style="list-style-type: none"> Considering the directional location information of vehicles 	<ul style="list-style-type: none"> Failure to examine the limitations and challenges in implementing the proposed protocol in real-world scenarios 			✓	✓	✓

Table 5. Cont.

Refs	Short Description	Simulation	Advantages	Disadvantages	Evaluation Criteria				
					Waiting Time	Queue Length	Travel Time	Delay Time	PDR
[38]	Investigating traffic in VANET using unsupervised machine learning such as K-means algorithms	-	<ul style="list-style-type: none"> Analysis and investigation of network traffic using machine learning methods 	<ul style="list-style-type: none"> Failure to examine the limitations and challenges in implementing the proposed protocol in real-world scenarios 				✓	✓
[39]	A review of fuzzy methods for routing and traffic management in VANET network	-	<ul style="list-style-type: none"> Advantages and limitations of VANETs based on fuzzy logic and their ability to control uncertainty and inaccuracy in the network 	<ul style="list-style-type: none"> Failure to present a new evaluation of the VANET network based on fuzzy logic 					
[40]	Using an intelligent hello propagation model that considers the mobility and position of nodes in the network	NS-2	<ul style="list-style-type: none"> Evaluation of packet delivery ratio, delay and network overhead 	<ul style="list-style-type: none"> Failure to examine the limitations and challenges in implementing the proposed protocol in real-world scenarios 				✓	✓
[41]	A combination of modified K-means clustering algorithm and continuous hopfield network for routing in VANET network	NS-2	<ul style="list-style-type: none"> Adequate performance of the proposed method in terms of packet delivery ratio and end-to-end delay 	<ul style="list-style-type: none"> Lack of detailed examination of modified K-means clustering algorithm and continuous Hopfield network 				✓	✓

3. Suggestions

According to the studies we reviewed in this research, machine learning methods have performed better than fuzzy methods due to their better analysis of evaluation criteria such as delay and PDR. In order to reduce delay and have better routing in VANET network, we can use machine learning methods. In addition, machine learning methods are less complicated than fuzzy methods. As a future plan, we suggest researchers work on routing and MAC in VANET network.

4. Discussion and Conclusions

Finding the most optimal route in the VANET network is one of the most important challenges in this field. Reinforcement learning maintains its effectiveness over time. RL algorithm can maintain its performance and also improve it. In general, three reinforcement learning methods, fuzzy logic and deep reinforcement learning have been described and mentioned in tables in detail. In each table, the description of the papers is given. Besides, the simulation and evaluation criteria of each paper are also given. A summary of the methods used in each paper is mentioned as well. By reviewing 30 paper, including reinforcement learning, fuzzy and deep reinforcement learning methods, a perspective is provided to researchers, which can lead their research in the future. Finally, the mentioned algorithms are analyzed in order to find the most optimal routing in the network.

Author Contributions: Conceptualization, A.K.S. and A.J.; methodology, A.K.S., A.J., C.-C.H., A.H. and A.Z.; investigation, resources, A.K.S., A.J., C.-C.H., A.H. and A.Z.; writing—original draft preparation, A.K.S., A.J., C.-C.H., A.H. and A.Z.; writing—review and editing, A.K.S., A.J., C.-C.H., A.H. and A.Z.; supervision, A.J.; project administration, A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, upon reasonable request.

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

References

1. Singh, A.; Prakash, S.; Singh, S. Optimization of reinforcement routing for wireless mesh network using machine learning and high-performance computing. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6960. [\[CrossRef\]](#)
2. Guo, J.; Gao, H.; Liu, Z.; Huang, F.; Zhang, J.; Li, X.; Ma, J. ICRA: An Intelligent Clustering Routing Approach for UAV Ad Hoc Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *24*, 2447–2460. [\[CrossRef\]](#)
3. Jafarzadeh, O.; Dehghan, M.; Sargolzaey, H.; Esnaashari, M.M. A Model-Based Reinforcement Learning Protocol for Routing in Vehicular Ad hoc Network. *Wirel. Pers. Commun.* **2022**, *123*, 975–1001. [\[CrossRef\]](#)
4. Zhao, M.; Li, J.; Tang, F.; Asif, S.; Zhu, Y. Learning Based Massive Data Offloading in the IoV: Routing Based on Pre-RLGA. *IEEE Trans. Netw. Sci. Eng.* **2022**, *4697*, 2330–2340. [\[CrossRef\]](#)
5. Lolai, A.; Wang, X.; Hawbani, A.; Dharejo, F.A.; Qureshi, T.; Farooq, M.U.; Mujahid, M.; Babar, A.H. Reinforcement learning based on routing with infrastructure nodes for data dissemination in vehicular networks (RRIN). *Wirel. Netw.* **2022**, *28*, 2169–2184. [\[CrossRef\]](#)
6. Ergun, K.; Ayoub, R.; Mercati, P.; Rosing, T. Reinforcement learning based reliability-aware routing in IoT networks. *Ad Hoc Netw.* **2022**, *132*, 102869. [\[CrossRef\]](#)
7. Xiao, L.; Lu, X.; Xu, D.; Tang, Y.; Wang, L.; Zhuang, W. UAV Relay in VANETs Against Smart Jamming With Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4087–4097. [\[CrossRef\]](#)
8. Wu, J.; Fang, M.; Li, H.; Li, X. RSU-Assisted Traffic-Aware Routing Based on Reinforcement Learning for Urban Vanets. *IEEE Access* **2020**, *8*, 5733–5748. [\[CrossRef\]](#)
9. Li, F.; Song, X.; Chen, H.; Li, X.; Wang, Y. Hierarchical Routing for Vehicular Ad Hoc Networks via Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2019**, *68*, 1852–1865. [\[CrossRef\]](#)
10. Wu, C.; Kumekawa, K.; Kato, T. Distributed Reinforcement Learning Approach for Vehicular Ad Hoc Networks. *IEICE Trans. Commun.* **2010**, *93*, 1431–1442. [\[CrossRef\]](#)
11. Mammeri, Z. Reinforcement Learning Based Routing in Networks: Review and Classification of Approaches. *IEEE Access* **2019**, *7*, 55916–55950. [\[CrossRef\]](#)
12. Wu, J.; Fang, M.; Li, X. Reinforcement Learning Based Mobility Adaptive Routing for Vehicular Ad-Hoc Networks. *Wirel. Pers. Commun.* **2018**, *101*, 2143–2171. [\[CrossRef\]](#)
13. Wu, C.; Yoshinaga, T.; Ji, Y.; Murase, T.; Zhang, Y. A Reinforcement Learning-Based Data Storage Scheme for Vehicular Ad Hoc Networks. *IEEE Trans. Veh. Technol.* **2016**, *66*, 6336–6348. [\[CrossRef\]](#)
14. Bi, X.; Gao, D.; Yang, M. A Reinforcement Learning-Based Routing Protocol for Clustered EV-VANET. In Proceedings of the 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 12–14 June 2020; pp. 1769–1773. [\[CrossRef\]](#)
15. Luo, L.; Sheng, L.; Yu, H.; Sun, G. Intersection-Based V2X Routing via Reinforcement Learning in Vehicular Ad Hoc Networks. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 5446–5459. [\[CrossRef\]](#)
16. Sliwa, B.; Schuler, C.; Patchou, M.; Wietfeld, C. PARRoT: Predictive Ad-hoc Routing Fueled by Reinforcement Learning and Trajectory Knowledge. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; pp. 1–7. [\[CrossRef\]](#)
17. Lu, C.; Wang, Z.; Ding, W.; Li, G.; Liu, S.; Cheng, L. MARVEL: Multi-agent reinforcement learning for VANET delay minimization. *China Commun.* **2021**, *18*, 1–11. [\[CrossRef\]](#)
18. Zhang, W.; Yang, X.; Song, Q.; Zhao, L. V2V Routing in VANET Based on Fuzzy Logic and Reinforcement Learning. *Int. J. Comput. Commun. Control* **2021**, *16*. [\[CrossRef\]](#)
19. Alabbas, H.; Huszák, Á. Reinforcement Learning based Gateway Selection in VANETs. *Int. J. Electr. Comput. Eng. Syst.* **2022**, *13*, 195–202. [\[CrossRef\]](#)
20. Phull, N.; Singh, P.; Shabaz, M.; Sammy, F. Enhancing Vehicular Ad Hoc Networks' Dynamic Behavior by Integrating Game Theory and Machine Learning Techniques for Reliable and Stable Routing. *Secur. Commun. Netw.* **2022**, *2022*, 4108231. [\[CrossRef\]](#)
21. Aznar-Poveda, J.; García-Sánchez, A.-J.; Egea-López, E.; García-Haro, J. Approximate reinforcement learning to control beaconing congestion in distributed networks. *Sci. Rep.* **2022**, *12*, 142. [\[CrossRef\]](#)
22. Zeynivand, A.; Javadpour, A.; Bolouki, S.; Sangaiah, A.; Ja'fari, F.; Pinto, P.; Zhang, W. Traffic flow control using multi-agent reinforcement learning. *J. Netw. Comput. Appl.* **2022**, *207*, 103497. [\[CrossRef\]](#)
23. Miri, F.; Javadpour, A.; Ja'fari, F.; Sangaiah, A.K.; Pazzi, R. Improving Resources in Internet of Vehicles Transportation Systems Using Markov Transition and TDMA Protocol. *IEEE Trans. Intell. Transp. Syst. (Early Access)* **2023**, 1–18. [\[CrossRef\]](#)
24. Nazib, R.A.; Moh, S. Reinforcement Learning-Based Routing Protocols for Vehicular Ad Hoc Networks: A Comparative Survey. *IEEE Access* **2021**, *9*, 27552–27587. [\[CrossRef\]](#)
25. Rezwan, S.; Choi, W. A Survey on Applications of Reinforcement Learning in Flying Ad-Hoc Networks. *Electronics* **2021**, *10*, 449. [\[CrossRef\]](#)
26. Teixeira, D.; Ferreira, J.; Macedo, J. Systematic Literature Review of AI/ML Techniques Applied to VANET Routing. In *Advances in Information and Communication. FICC 2022. Lecture Notes in Networks and Systems*; Springer: Cham, Switzerland, 2022; Volume 439, pp. 339–361. [\[CrossRef\]](#)

27. Javadpour, A.; Rezaei, S.; Sangaiah, A.K.; Slowik, A.; Khaniabadi, S.M. Enhancement in Quality of Routing Service Using Metaheuristic PSO Algorithm in VANET Networks. *Soft Comput.* **2021**, *27*, 2739–2750. [[CrossRef](#)]
28. Tu, S.; Waqas, M.; Rehman, S.U.; Mir, T.; Abbas, G.; Abbas, Z.H.; Halim, Z.; Ahmad, I. Reinforcement Learning Assisted Impersonation Attack Detection in Device-to-Device Communications. *IEEE Trans. Veh. Technol.* **2021**, *70*, 1474–1479. [[CrossRef](#)]
29. Saravanan, M.; Ganeshkumar, P. Routing using reinforcement learning in vehicular ad hoc networks. *Comput. Intell.* **2020**, *36*, 682–697. [[CrossRef](#)]
30. Zhang, D.; Yu, F.R.; Yang, R.; Tang, H. A deep reinforcement learning-based trust management scheme for software-defined vehicular networks. In Proceedings of the 8th ACM Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications-DIVANet'18, Miami Beach, FL, USA, 25–29 November 2018; Volume 8, pp. 1–7. [[CrossRef](#)]
31. Wang, H.; Li, H.; Zhao, Y. An Intelligent Congestion Control Strategy in Heterogeneous V2X Based on Deep Reinforcement Learning. *Symmetry* **2022**, *14*, 947. [[CrossRef](#)]
32. Kandali, K.; Bennis, L.; El Bannay, O.; Bennis, H. An Intelligent Machine Learning Based Routing Scheme for VANET. *IEEE Access* **2022**, *10*, 74318–74333. [[CrossRef](#)]
33. Khatri, S.; Vachhani, H.; Shah, S.; Bhatia, J.; Chaturvedi, M.; Tanwar, S.; Kumar, N. Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges. *Peer-to-Peer Netw. Appl.* **2021**, *14*, 1778–1805. [[CrossRef](#)]
34. Budholiya, A.; Manwar, A.B. Machine learning based analysis of VANET communication protocols in wireless sensor networks. In Proceedings of the 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 1–3 December 2022; IEEE: Coimbatore, India, 2022.
35. Tang, Y.; Cheng, N.; Wu, W.; Wang, M.; Dai, Y.; Shen, X. Delay-Minimization Routing for Heterogeneous VANETs With Machine Learning Based Mobility Prediction. *IEEE Trans. Veh. Technol.* **2019**, *68*, 3967–3979. [[CrossRef](#)]
36. Aljabry, I.A.; Al-Suhail, G.A. Improving the Route Selection for Geographic Routing Using Fuzzy-Logic in VANET. In *Intelligent Computing & Optimization: Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021 (ICO2021)*; Springer International Publishing: Berlin/Heidelberg, Germany, 2022.
37. Rana, K.K.; Tripathi, S.; Raw, R.S. Fuzzy Logic-Based Directional Location Routing in Vehicular Ad Hoc Network. *Proc. Natl. Acad. Sci. USA* **2021**, *91*, 135–146. [[CrossRef](#)]
38. Aouedi, O.; Piamrat, K.; Hamma, S.; Perera, J.K.M. Network traffic analysis using machine learning: An unsupervised approach to understand and slice your network. *Ann. Telecommun.* **2021**, *77*, 297–309. [[CrossRef](#)]
39. Maddiboyina, H.V.; Ponnappalli, V.S. Fuzzy logic based VANETS: A review on smart transportation system. In Proceedings of the 2019 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 23–25 January 2019; IEEE: Coimbatore, India, 2019.
40. Ayub, M.S.; Adasme, P.; Melgarejo, D.C.; Rosa, R.L.; Rodriguez, D.Z. Intelligent Hello Dissemination Model for FANET Routing Protocols. *IEEE Access* **2022**, *10*, 46513–46525. [[CrossRef](#)]
41. Kandali, K.; Bennis, L.; Bennis, H. A New Hybrid Routing Protocol Using a Modified K-Means Clustering Algorithm and Continuous Hopfield Network for VANET. *IEEE Access* **2021**, *9*, 47169–47183. [[CrossRef](#)]

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