

Research Article

A Clustering-Based Routing Protocol Using Path Pattern Discovery Method to Minimize Delay in VANET

Xiaoyun Xie ^{1,2}, Yahya Dorostkar Navaei ³, and Sajad Einy ⁴

¹School of Electronic Information Engineering, Gannan University of Science and Technology, Ganzhou, 341000 Jiangxi, China

²Ganzhou Key Laboratory of Cloud Computing and Big Data, Ganzhou, 341000 Jiangxi, China

³Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

⁴Istanbul Aydin University, Department of Application and Research Center for Advanced Studies, Istanbul, Turkey

Correspondence should be addressed to Yahya Dorostkar Navaei; y.dorostkar@qiau.ac.ir

Received 13 February 2022; Revised 22 January 2023; Accepted 28 April 2023; Published 14 June 2023

Academic Editor: Farhan Ullah

Copyright © 2023 Xiaoyun Xie et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In vehicular ad hoc networks (VANETs), vehicle-to-vehicle (V2V) communications can link vehicles to each other, and vehicle-to-infrastructure (V2I) messaging and communications can link roadside infrastructure such as routers. The vehicles in these networks act as relays that transmit critical messages in the network. Due to the high-speed movement of vehicles on the road, real-time messaging and minimizing the delay in sending messages is one of the most important objectives of VANET developers. On the other hand, the high mobility of vehicles causes communication interruptions and decreases the data delivery rate in VANET. To overcome this issue, predicting the path of vehicles can play an important role in sending data from the source to the destination. When an accident occurs on the road, the messages that are sensed by the imbedded sensors in the vehicles need to be sent, and if they are sent by the vehicles that change their route, these messages will not be sent to the destination and the performance of the network will be disturbed. Previous methods in the literature for data transmission in intervehicular networks have focused more on reliability and trust, and little attention has been paid to the prediction of vehicle movement paths in these types of networks. Therefore, for fast and reliable data transmission in VANET, accurate prediction of vehicle movement and creation of movement patterns can be effective in message transmission delay and data delivery rate. In this paper, we present an approach using a combination of cluster-based routing protocols and pattern discovery methods to minimize latency in VANETs. The outline of the proposed method has four modules: primary data collection and analysis, primary data preparation and analysis, pattern extraction and vehicle route discovery, and vehicle clustering and data/information transmission routing. The simulation results show that the proposed method with a delivery rate of 88.56% has significantly improved compared to the previous methods in terms of package delivery rate. Also, the proposed method with a total delay of 24.566 ms has a shorter delay than the previous methods in terms of message sending delay in the network.

1. Introduction

With the increasing usage of wireless communications, today we are witnessing the emergence of new types of wireless networks. The vehicular ad hoc network (VANET) is one of these new networks that enables wireless connections between vehicle nodes and roadside infrastructure. VANETs are a decentralized, self-organized communication network in which nodes consist of high-speed vehicles that are automatically routed between adjacent devices,

without the use of any infrastructure (such as routers and servers). Thus, vehicles can effectively transmit a message from one source node to other nodes, through their adjacent vehicle to vehicle (V2V) or even with communicate with some existing infrastructure (V2I), such as roadside units, in the same geographic area to create a better environment for safer driving [1, 2].

The use of VANETs has expanded rapidly in recent years due to its diverse and useful applications, but there are several challenges in using this technology. For example,

the high speed of vehicles in this network causes frequent changes in the network topology, and as a result, the communication link between vehicles will be unstable or may even be cut off. To establish route between source and destination in a VANET, we need to use routing protocols. Transferring packets from the node that is the source of the message to another node in the neighborhood of the source vehicle involves transferring packets across the network. In fact, in VANETs, unlike fixed networks, there is no clear connection between nodes, and this has made it difficult to route and transmit packets. Every vehicle in the network, when it receives data packets, must also send these packets to its neighboring vehicles, which is called VANET message retrieval. Routing in VANETs is done by sending data packets between the sender and receiver in the network until all the vehicles in the network receive the desired packets. Therefore, the use of machine learning in improving routing in VANET can provide an effective solution. The application of machine learning methods has been proven in various fields of science such as wireless networks [3, 4], social networks [5], network security [6–8], and pattern recognition [9–14]. One of the most important challenges in VANET is the delay in sending and receiving messages between vehicles.

The main purpose of VANETs is to effectively transmit a message from one source node to other nodes located in the same geographical area and to create a suitable environment for safer, more efficient, and easier driving [15, 16]. This platform has warning programs to warn other vehicles, weather and traffic conditions, the exchange of multimedia files between vehicles (for example, songs and movies) and road units (TV, radio, or news), and so on [3]. For example, information about traffic areas can be sent from one vehicle to another, or drivers can be notified of potential accidents or traffic jams. In addition, it is possible to introduce alternative routes in the VANET in order to avoid possible or delayed accidents in high-traffic areas. Similarly, a vehicle through its built-in sensors can detect potential accidents such as a frozen road and notify other vehicles [17].

The motivation of this paper is to find the vehicles that aligned with the vehicle carrying the message and extracting the movement pattern of these vehicles according to the movements of the previous similar vehicles. In the proposed method, vehicle information is received based on moment-to-moment location changes and their instantaneous and average speed as a time series. Then, according to the frequent patterns in the movement of vehicles in the time series, a fixed pattern for the movement of vehicles is found according to their speed and direction. Then, based on the change of location of the vehicles, it is determined whether these cars are aligned with the car carrying the message or not. On the other hand, according to the current location of the vehicles, the network is clustered and neighboring cars are identified. Then, according to the movement patterns of the vehicles, the next path is predicted for the neighboring cars. A vehicle that meets the following three conditions is determined as the cluster head vehicle that is able to receive the message:

- (1) Be in the middle of the cluster based on your distance with other cars
- (2) Be in line with the vehicle carrying the message
- (3) Match the predicted movement pattern

After determining the cluster head vehicles, the message can be transferred to the destination through the cluster head vehicles. The proposed method uses the extraction of the movement pattern of cars based on the information of the vehicles in the form of time series as an innovation, which has been given less attention in previous articles.

The main contributions of this paper are as follows:

- (i) Using the decision tree classification method to classify vehicles in aligned and not aligned vehicles
- (ii) Using sequential pattern mining to discover vehicle movement patterns
- (iii) Clustering vehicles in the network based on the current position to detect neighbor vehicles
- (iv) Finding best cluster head to transfer messages using combination of above three steps

This article is organized in 6 sections. In the second section, we will have an overview of the background of researches in the field of VANET. In the third part, we will explain the proposed method, and in the fourth part, we will bring the simulation results of the proposed method, and in the fifth part, we will explain the evaluation results of the proposed method. In the sixth section, we will have a conclusion.

2. Related Works

Kakkasageri and Manvi [18] used regression mechanism and proposed a method for collecting information and disseminating critical information based on cognitive factor in VANET. The regression-based cognitive approach effectively aggregates the critical information collected and minimizes the dissemination of transmitted data. The proposed scheme works on cluster vehicles using a set of static and moving agents. The scheme has steps including (1) validating and filtering the critical information collected, (2) generating knowledge based on important and critical information filtered, (3) gathering knowledge to motivate using regression techniques, (4) increasing motivation for better quality of information aggregation, and (5) disseminating the collected information to neighboring clusters.

Rehman et al. [19] proposed a scheme for selecting bidirectional stable communication (BDSC) for multistep broadcast protocols on a wide range of vehicles. The selection of relay nodes based on the quantitative representation of link characteristics for single-step neighboring nodes is proposed using a link quality estimation algorithm. The BDSC scheme is designed to improve packet delivery rates and minimize delays in communicating in a high-density network with nodes distributed over a large coverage area. To achieve this goal, the proposed design in this study attempts to balance

the quality of the estimated link and the distance between the source distributor and potential senders when selecting the next nodes to transmit the sent messages.

Louazani and Sekhri [20] introduced a clustering mechanism based on connection maintenance in VANET called AODV-CV. In this research, a formal model using net Petri time as a mathematical tool to prove the properties of the protocol is presented. Also, a mobile virtual clustering protocol for VANET is proposed to improve connectivity on a highway.

Jalooli et al. [21] investigated the message propagation performance in the VANET environment and proposed a safety-based disconnected RSU replacement algorithm (S-BRP). VANETs are dramatically designed to increase road safety and traffic efficiency through vehicle-to-vehicle communications and vehicle-to-road infrastructure. Roadside units (RSUs) play an important role in terms of connectivity, routing, and data transmission latency in VANETs. However, using RSU cannot provide enough coverage for our target area. The S-BRP algorithm has been evaluated through extensive simulation. The results show that this algorithm has a good performance in terms of reducing the propagation delay and traffic flow.

Abuashour and Kadoch [22] have proposed three protocols (including Cluster-Based Life-Time Routing (CBLTR), Intersection Dynamic VANET Routing (IDVR), and Control Overhead Reduction Algorithm (CORA)). The CBLTR protocol is designed to increase path stability and average performance in a two-way road scenario. Cluster nodes (CHs) are selected based on the maximum lifespan of all vehicles within each cluster. The IDVR protocol is designed to increase path stability and average performance and reduce point-to-point latency in network topology. The selected node receives a Set of Candidate Shortest Routes (SCSRs) to the nearest intended destination from the defined software network. The IDVR protocol selects the optimal route from the SCSR based on the current location, destination location, and maximum average power. Finally, the CORA implements control overhead messages between cluster members and CH with the aim of reducing the amount of message control overhead in clusters by creating a new mechanism for calculating optimal numbers.

Shahidi and Ahmed [23] proposed an approach for the two-way multilane highway that can be efficient in both directions due to the movement of vehicles. In this approach, the possible end-to-end delay distribution is calculated and its dependence on system parameters (such as two-way speed distribution, communication range, and vehicle density) is investigated. Simulation of this approach has been used to investigate the analytical model. The good agreement between the simulation results and the analytical calculations shows the accuracy of the proposed analytical model.

Mohammed Nasr et al. [24] proposed a VANET cluster routing algorithm for desert areas and other rugged environments that provides a stable cluster structure and a reliable route between source and destination nodes. In addition, it uses vehicles equipped with mobile or satellite links to act as gateways to inaccessible destinations. A new method for

selecting eclipses has also been proposed in this study and has been theoretically analyzed. During this study, numerical simulations were used to evaluate the proposed method for selecting the cluster head. The simulation results were also performed to evaluate the proposed routing algorithm in terms of packet delivery ration (PDR), terminal-to-terminal latency, and cluster stability and compare it with other options.

Ardakani [25] presented the ACR algorithm, which is a cluster-based routing protocol for transmitting network traffic in VANET. Using this protocol, each node first selects an identifier, called a LOCO, based on its location and mobility. The network is then divided into a set of clusters based on the node mobility pattern. The Hamming distance criterion is used to measure the similarity of moving nodes using LOCO values. Nodes are categorized using a lightweight clustering algorithm. Each cluster is managed by a cluster (CH) whose functionality is to communicate with RSUs.

In [26], a blockchain technique has considered the best technique that provides secrecy and protection to the control system in real-time conditions. In [27], the trust-based framework with a novel mechanism to determine DDoS attacks in VANET has developed. In [28], a secure information management scheme has proposed named Third Eye, by satisfying all the performance parameters as well as satisfying the trust metrics among the vehicles and devices. In [29], a portable VANET routing protocol that learns the optimal route by employing a fuzzy constraint Q-learning algorithm has proposed. In [30], an efficient routing solution based on a flooding technique has conceived to make the data delivery more reliable and to guarantee robust paths. In [31], a flooding scheme that automatically reacts at each topology variation while overcoming the present obstacles while exchanging data in ad hoc mode with drones that are commonly called unmanned aerial vehicles (UAVs) has designed. In [32], a routing algorithm in software-defined vehicular network (SDVN) based on the hidden Markov model (HMM) and temporal graph has introduced that considers the vehicular network as a temporal graph, in which each data transmission as an edge has its specific temporal information. In [33], a social computing inspired predictive routing scheme (SPIDER) for SDVNs has proposed to low-latency reliable data exchange under dynamic vehicular networks. In Table 1, the important evaluation parameters in the previous methods are reviewed.

As shown in Table 1, previous related methods in the literature for data transmission in VANET have focused more on reliability and trust, and little attention has been paid to the prediction of vehicle movement and path pattern discovery in these types of networks. Therefore, in order to overcome this issue, the proposed method has presented an approach based on clustering and frequent pattern discovery in predicting the movement path of vehicles.

3. Proposed Methodology

The outline of the proposed method has four modules as follows:

Module 1: initial data collection and analysis

TABLE 1: Evaluation parameters in related works.

Related work	Delay	Reliability	Data delivery	Trust
[18]	✓	×	✓	×
[19]	✓	✓	✓	×
[20]	✓	✓	✓	×
[21]	✓	✓	✓	×
[22]	✓		✓	×
[23]	✓	✓	×	×
[24]	✓	×	×	✓
[25]	✓	✓	×	✓
[26]	✓	×	✓	✓
[27]	×	✓	✓	×
[28]	✓	✓	×	✓
[29]	✓	×	✓	✓
[30]	×	✓	✓	×
[31]	✓	✓	×	✓
[32]	×	✓	✓	✓
[33]	✓	✓	✓	✓

Module 2: data preparation and initial analysis

Module 3: creating a pattern and discovering the path of cars

Module 4: vehicle clustering and data/information transmission routing

In the following, we will describe the modules and methods that are used in these modules. The flowchart of the proposed method is shown in Figure 1.

As shown in Figure 1, the proposed method flowchart has four modules as follows.

3.1. Initial Data Collection and Analysis. The first part of doing anything is collecting raw data. In the intelligent transportation system, the required data about the traffic on the routes and the speed of the vehicles are collected in different ways. Some of these methods include installing cameras on the road and installing sensors on the floor of the street, sensors in the body and inside the car, and so on. Some other data is collected via GPS used in vehicles, such as the exact geographical location of the vehicle.

All collected data must be stored in a single memory. The sensors themselves have a limited storage compartment, so it is necessary to connect this data to RSU roadside infrastructure using communication methods such as Wi-Fi, VANET, and Internet and send the data to the center through them for storage, analysis, and control.

The first step in the proposed intelligent system is to create and develop a sensing system. This sensing system consists of the following parts:

- (i) *Environmental data collection tools/sensors:* in the method proposed in this research, a set of environmental data collection tools (such as sensors) installed on street pavement, sensors installed on cars, cameras installed on the streets, sensors installed on roadside control, and communication

infrastructure is used to establish V2I communications to collect environmental data. This data can eventually be integrated together or transferred separately to control centers. In this research, the emphasis is on sensors installed under the pavement of the streets and on the car to measure the ambient temperature and identify glacial routes and GPS system to identify the location of cars and send it to the control center in line with other data. We used the collected NFCs for cases where we need to carefully monitor short but sensitive streets and routes, or we can say that the geographical area covered is less in this case. Since different sensors can be used simultaneously with these tools and sensors mentioned and considered by us, so we actually intend to use WSNs, NFC, and GPS as the main methods of collecting environmental data. In some cases, some very urgent and common operations can be performed by installing actuators in our proposed system. Like foggy air-sensitive actuators, the actuators can automatically alert the driver as soon as the environment detects that the environment is foggy, and in some smart and automatic vehicles, they automatically slow down their speed

- (ii) *Communication technologies:* after collecting data, the next step is how to exploit this data. Most of our tools and sensors are not able to fully process and exploit this data, so this data must be transferred to training centers. For this purpose, some communication technologies are used depending on the use (such as VANET, Wi-Fi, Bluetooth, DSRC, and GSM)
- (iii) *Communication tools/communication channel:* we use Internet communication channels as well as 5G mobile services to create communication
- (iv) *Web server/data storage and initial processing:* to store the collected data and maintain it for future processing or processing in real time or to separate the data and store the data in the database, available data such as location data in the GPS database, environmental data such as temperature or frost in the Ambience DB database, or location data in the commuter DB database is saved. This data in different databases can be used by the server when processing data, and the data contained in them is updated at certain intervals

3.2. Data Preparation and Initial Analysis. Some basic analyses such as identifying high-risk routes in terms of high accident statistics, identifying congested routes, detecting congestion hours during the day, and detecting congested days during the week and year are performed in this section. Finally, the normalized and desirable data for further analysis, along with the results obtained from the initial analysis performed in this module, are stored in a special main database.

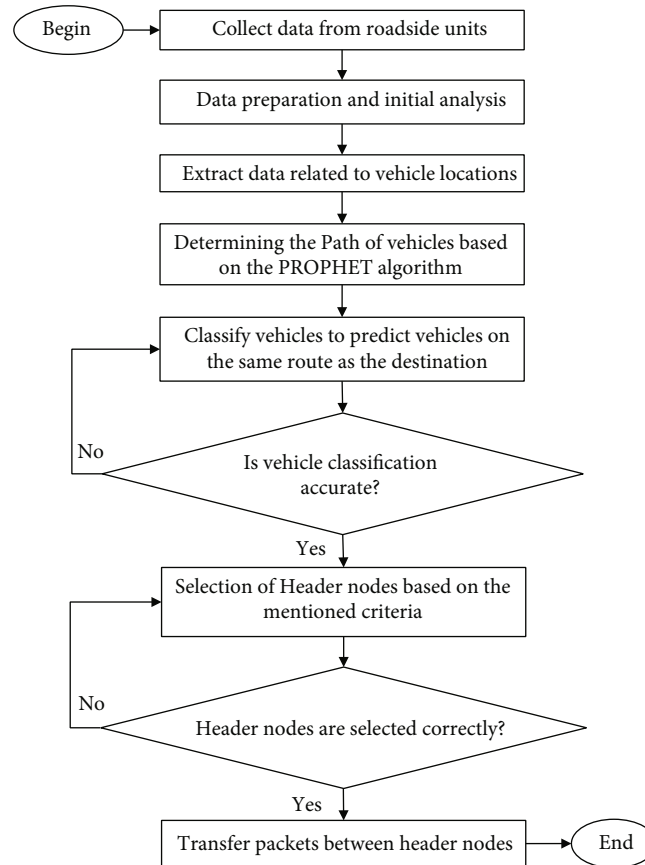


FIGURE 1: Flowchart of the proposed method.

Major tasks in data preparation or preprocessing performed on data include data comprehension, data cleansing, data integration, data conversion, and data size reduction.

3.3. Extraction of Pattern and Discovering the Path of Vehicles. In this module, two ways are applied to create a unique pattern for each vehicle:

- (i) *Applying the classification techniques in data mining:* classification techniques, such as decision tree, Naïve Bayesian, SVM, neural networks, rough set, and random forests, try to model each vehicle using attributes like the routes traveled, average speed, type of vehicle, age and gender of the vehicle driver, and number of offenses and extract the pattern of behavior for each vehicle. Because the decision tree technique is one of the simplest methods of classifying samples and can work with different types of data, in this study, this technique has been used to determine the behavioral patterns of drivers. Also, for each of the existing routes, a pattern is extracted based on the number of vehicle on that route and traffic situation at different hours of the day and at different days of the week and months of the year. In this model, the traffic situation and congestion

of vehicle at different hours, especially during rush hour congestion, are examined

- (ii) *Use of protocols to create vehicle trajectories in different routes:* the number of transit packages between vehicles and RSU off-road infrastructure and between vehicles and control centers is examined, and a pattern of the number of collisions, collisions of messages, or loss of messages is examined

In this research, using the PROPHET protocol, we have collected and created the vehicle trajectory so that we can adjust the routing tolerance with error.

3.4. Vehicle Clustering and Data/Information Transfer Routing. In this module, for all vehicles on the route, we put them in clusters based on various criteria such as the behavioral similarities or routes traveled, average speed, amount of traffic offenses, and number of accidents. Flowchart in Figure 2 is used to select the header in each cluster.

The data exchange paths will be creating from the header to the other nodes. And first, there is the priority with nodes that have a higher degree.

3.5. Complexity. Due to the fact that the movement data of vehicles collected in a certain period of time is used to

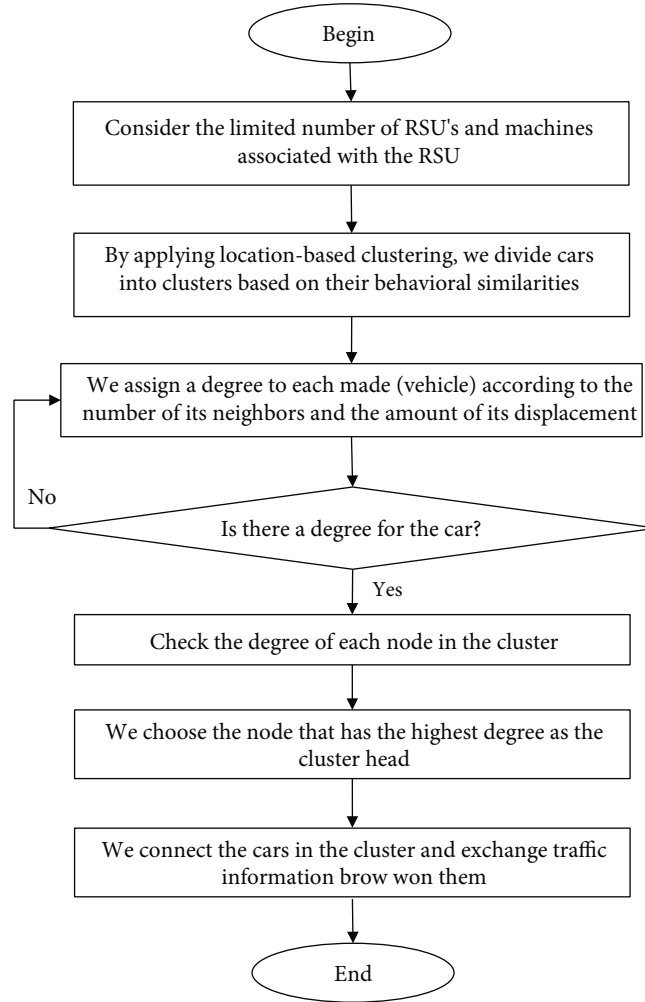


FIGURE 2: Flowchart of how to choose cluster head.

generate movement patterns, therefore, the generation of movement patterns will be performed offline and once during the proposed method. The time complexity of generating movement patterns based on the steps of the proposed method is $O(n)$ for clustering cars and $O(n^2)$ for comparing the movement of vehicles on the road. After generating movement patterns, to predict the movement path of new vehicles, it is enough to compare the location of the vehicles with the patterns. Due to the limited number of vehicle movement patterns, the time complexity of predicting the movement path of new vehicles can be ignored, and the time order of the proposed method is determined based on the number of vehicles in the network and of the order of $O(n)$. Finally, the total time cost of the proposed method will be of the order of $O(n^2)$.

4. Experimental Results

In order to simulate the proposed method in this research, NS-2 network simulator software version 2.35 was used. The simulation environment is a cross street with dimensions of 1000 by 1000 meters, which has four access points (AP) and two holes along the road. The street is located

between two intersections, at both ends of which are points simulated as traffic lights, with a total of 29 vehicles moving on the street. Vehicles in the network vary in speed from 10 to 30 km/h. Also, the settings related to the antenna and other infrastructures in this scenario have been applied in accordance with the standard settings in the simulation of previous works. More details about the simulation scenario are shown in Table 2.

4.1. Implementation of the First Module of the Proposed Method: Data Collection. In order to implement the first module and collect data, apply the scenario mentioned in Section 3 on the NS-2 emulator software and the result of the scenario on the NS-2 emulator in two output files with extensions save * .nam and * .tr. The file with the * .nam extension is related to the scenario visualization part. This file runs on the Nam console, which is part of the NS-2 simulator, and shows how vehicles move and transmit messages between vehicles. Figure 3 shows the graphical view of Nam software.

As shown in Figure 3, the cars are moving in the street and are communicating with each other wirelessly through built-in sensors. When a vehicle becomes aware of a hazard in the network for various reasons, it sends this message to

TABLE 2: Scenario details of the proposed method.

Simulation details	Initial values
Simulation space	1000*1000 m
Number of vehicles	29
Speed of vehicles	10-30 km/h
Package size	50
Routing protocol	DSDV
Communication protocol	FTP, TCP
Simulation time	35 s
Neighbor selection threshold	200 m

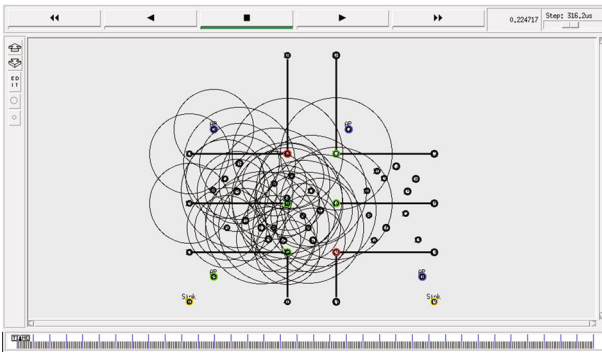


FIGURE 3: Appearance of street and vehicle simulation.

other vehicles within the network. According to the proposed method, before sending the message, the neighboring vehicles in the network should be identified and the message should be transmitted only between these vehicles to avoid redundancy and creating excessive duplicate messages and wasting time. For this purpose, in the event of an accident in the VANET, the originator first collects the distances of other vehicles in the network. Then, the position of each vehicle at any time relative to the starting point is calculated. Table 3 shows the values for the starting points and the speed and location of the new vehicles at different times.

As shown in Table 3, the position of each vehicle is calculated at different times. Now we transfer information in the network and collect data based on what happened in the network.

4.2. Implementation of the Second Module of Proposed Method: Data Preparation. The data collected in the first module is in the form of raw data that shows the position of vehicles in the network. In this module, we will prepare the data for presentation to the next modules. For this purpose, based on the defined scenario, we transfer information in the network between vehicles and roadside units and save the events in the trace file for later use. Hence, in Figure 4, transfer of information between vehicles present in the network is shown.

As shown in Figure 4, information is transmitted between vehicles on the network and reports the sending and receiving packets, sending and receiving times, vehicle locations, average vehicle speeds, and number of lost packets. This information is saved on the trace file. In this module, we will prepare this information for use in the next modules.

4.2.1. Data Preprocessing. In recent years, various training models have been found that, by performing an educational process on data, are able to predict and describe unknowns to the system. The important point here is the type of data used for each type of model. In fact, each model deals with a specific data type and explores a specific type of data. To use the models and benefit from the output results of the model, it is necessary to prepare the data in a specific model format. The data preparation process for each model is called data preprocessing. Data preprocessing has several steps that in this research, two types of these preprocessing steps are required that will be applied to the data extracted from the trace file in the simulation of the proposed scenario. Table 4 shows an example of the data stored in the trace file.

4.2.2. Data Cleaning. According to the set of features shown in Table 4, it can be seen that the data extracted from the trace file have many dimensions and the values of these features have been expanded in various ranges. The most important problem among data that needs to be cleaned is missing values. In order to overcome the problem of missing values, several solutions have been proposed in the sources, which are as follows:

- (i) Delete the sample with the missing value
- (ii) Replace the lost value with a random value
- (iii) Replace the lost value with the same value
- (iv) Replace the lost value with the average
- (v) Replace lost value based on specific function

The most common solution is the method of replacing the lost value with the mean, which in this research has been used in order to overcome the problem of lost values. Therefore, properties that have a missing value are cleared using the method of replacing the lost value with the mean.

4.2.3. Select a Subset of Features. The second preprocessing step used in this research is the selection of a subset of attributes that are directly related to the class label. As shown in Table 4, the data extracted from the trace file has 11 properties, and this large number of attributes can complicate the training model and the path discovery protocol which is used. Therefore, some of these attributes, which have little effect on node path detection, should be removed during the preprocessing step of selecting the feature subset.

The purpose of the preprocessing step is to select a subset of features to remove unrelated features and attributes and plugins so that in addition to reducing the data dimension and reducing the operational and spatial complexity of the system, the system accuracy can be increased. In addition, selecting a subset of data can detect implicit dependencies between data and path detection patterns to easily predict the path of test nodes that will be added to the scenario in the future. Therefore, in this study, irrelevant features and attributes that do not have many changes in the samples and naturally cannot have much effect on discovering the node path pattern

TABLE 3: Information on vehicle locations.

Location at $t = 25$ s		Location at $t = 10$ s		Location at $t = 1$ s		Vehicle speed	y_0	x_0	Vehicle no.
454.448	349.448	382.015	277.015	363.955	258.955	20	330	225	1
436.893	426.893	357.217	347.217	337.3509	327.350	22	300	290	2
374.448	504.448	302.015	432.015	283.955	413.955	20	250	380	3
443.336	553.336	389.011	499.011	375.466	485.466	15	350	460	4
494.448	654.448	422.015	582.015	403.955	563.955	20	370	530	5
387.113	237.113	336.411	186.411	323.768	173.768	14	300	150	6
405.560	655.560	315.019	565.019	292.444	542.444	25	250	500	7
436.893	616.893	357.217	537.217	337.350	517.350	22	300	480	8
424.448	464.448	352.015	392.015	333.955	373.955	20	300	340	9
412.224	792.224	376.007	756.007	366.977	746.977	10	350	730	10
399.558	899.558	341.612	841.612	327.164	827.164	16	300	800	11
393.115	463.115	309.818	379.818	289.048	359.048	23	250	320	12
422.224	942.224	386.007	906.007	376.977	896.977	10	360	880	13
349.558	1029.558	291.612	971.612	277.164	957.164	16	250	930	14
368.226	868.226	299.415	799.415	282.257	782.257	19	250	750	15
524.669	964.669	481.209	921.209	470.373	910.373	12	450	890	16
624.448	1054.448	552.015	982.015	533.955	963.955	20	500	930	17
693.115	993.115	609.818	909.818	589.048	889.048	23	550	850	18
593.115	873.115	509.818	789.818	489.048	769.048	23	450	730	19
643.115	943.115	559.818	859.818	539.0486	839.048	23	500	800	20
654.448	894.448	582.015	822.015	563.955	803.955	20	530	770	21
543.336	593.336	489.011	539.011	475.466	525.466	15	450	500	22
590.891	500.891	543.810	453.810	532.070	442.070	13	510	420	23
482.224	462.224	446.007	426.007	436.977	416.977	10	420	400	24
560.891	430.891	513.810	383.810	502.070	372.070	13	480	350	25
593.115	343.115	509.818	259.818	489.048	239.048	23	450	200	26
703.115	353.115	619.818	269.818	599.048	249.048	23	560	210	27
636.893	286.893	557.217	207.217	537.350	187.350	22	500	150	28

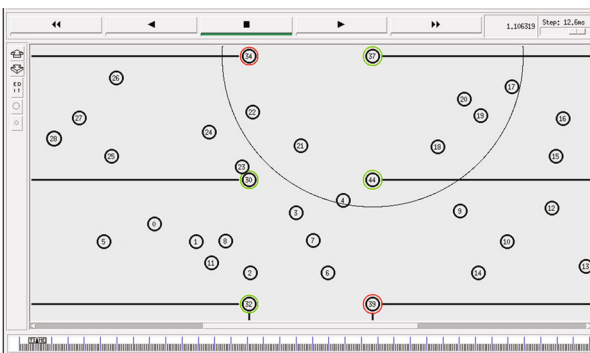


FIGURE 4: Transfer of information in the VANET.

are removed to impose additional complexity on the model to avoid suggested training.

Feature selection methods can be broadly divided into filter and wrapper methods. In the filter approach, the attribute selection method is independent of the training model that is applied to the selected attributes and evaluates the attribute weight only by considering the intrinsic properties of the data. In most cases, the amount of weight is calcu-

TABLE 4: An example of the data extracted from the trace.

Description	Value
Event	r
Time	0.032237058
Sender node	28
Protocol	RTR
Packet type	Message
Receiver node	30
Communication port	[0 ffffffff 1e 800]
Communication path	[30:255-1:255 32 0]
Initial location	930,500
Final location	10,450
Node speed	20.00

lated and the weak features are removed. The extracted features are presented as input to the training model. In this research, the filter approach is used to select features. After applying the filter method on the data obtained from the trace file, you can see that the features related to the

communication port, message type, and communication protocol are removed.

4.3. Implementation of the Third Module: Discover the Path of Vehicles and Create a Pattern. After preprocessing the data and extracting important features about the behavior of vehicles in the proposed network, in this module, we will explore the vehicle path and create patterns to identify vehicles moving on the existing route to the destination. The vehicle path refers to the direction in which the vehicle moves from the beginning of the simulation moment to the end of the simulation. This path represents the starting point of the vehicle x_0 , the end point of the vehicle x_n , and the direction of the car. For this purpose, we use the data recorded in the trace file and extract the starting points and locations of the vehicles at any time. Now, using the data related to the location of the cars, we will draw the path of the cars in the proposed network. Figure 5 shows an example of the path of vehicles in the network.

As shown in Figure 5, the vehicles in the proposed network start moving from point x_0 and continue moving up to point x_n . The trajectory of each vehicle is recorded at each moment according to the location at the beginning and end of the route and the location of the vehicle, and based on this, the trajectory of that vehicle is determined. Now, according to the trajectory of vehicles and according to the average speed of vehicles and the direction of movement of each vehicle, it is possible to create patterns about vehicles in the network.

To create patterns for vehicles on the network, we first classify vehicles in order to find vehicles moving in the direction of the destination. Thus, the vehicles in the network are divided into two classes: destination-oriented vehicles and non-destination-oriented vehicles. Routine vehicles are vehicles that, due to their behavior in the network, approach the destination node along their route and can exchange information with the destination node. Therefore, we apply the behavior of vehicles in the network to the decision tree as the characteristics of the vehicles so that the decision tree classifies the vehicles and the path patterns of the vehicles are determined. These characteristics include the average difference between the distance of each vehicle to the destination node, vehicle speed, distance of vehicles at the x_n point from the destination node, and total distance of the vehicle from other vehicles on the route.

4.3.1. Implement the Decision Tree. In this research, the decision tree, which is a rule-based classifier, has been used. The important point here is to choose a feature and a condition on the feature that divides the data well, so that the leaves

have the maximum degree of purity. Therefore, in this research, gain ratio criterion has been used to divide the samples based on features, so that important features can be identified and used as a condition for division. This criterion is calculated at each stage (level) of the decision tree production for all properties. Figure 6 shows the application of conditions to properties and the creation of a decision tree.

As can be seen in Figure 6, the decision tree applies conditions to the properties in the data set based on the importance of these properties and divides the entire data into subtrees that apply in these conditions. Each of these conditions represents a pattern that describes the vehicle's behavior on the road. Therefore, we make these patterns into rules in order to predict the trajectory of future vehicles. As mentioned earlier, in this study, nodes are defined in two groups of aligned and nonaligned vehicles, and all vehicles are placed in one of these two groups based on their behavioral patterns. Alignments are shown as class 1 and nonaligned vehicles as class 2. The patterns created are shown in Table 5.

As can be seen in Table 4, the patterns created based on the decision tree are based on the features in the data set, which will be used to predict the trajectory of future vehicles. After classifying the vehicles and creating patterns of vehicle behavior, we predict the trajectory of the next vehicles using the PROPHET protocol (probable routing protocol using encounter date and transfer protocol).

In PROPHET, the predictability of delivery between two nodes is calculated based on the date of contact between them, while the higher predictability of delivery increases the likelihood of further communication between them. In the PROPHET protocol, a message is copied to the contact node only when the transmission prediction capability for the destination node of the contact node is greater than the transfer node. By doing so, the PROPHET protocol is likely to deliver packets well, as well as meet latency and message overhead costs.

In this method, we use the PROPHET protocol to predict the trajectory of vehicles within the network. Vehicle lanes are designed to transmit information and warning data within the road network to destination nodes or roadside potholes. Vehicles that have a trajectory along the hole are a good option for transmitting information, and messages can be delivered to these vehicles. For this purpose, in the proposed method, the PROPHET protocol is used for vehicles that have been selected as aligned vehicles in the previous step through the decision tree, in which we use vehicle trajectory factors and vehicle speed to calculate delivery capability. Therefore, to calculate the deliverability, we use the proposed

$$DC(i) = \left| \frac{1}{\sum_{j=1}^t \sum_{i=1}^n \left(\left(\sqrt{(x_{ij}-x_s)^2 + (y_{ij}-y_s)^2} - \sqrt{(x_{(i-1)j}-x_s)^2 + (y_{(i-1)j}-y_s)^2} \right) / n \right)} \right| + \left| \frac{1}{\sum_{i=1}^n \sqrt{(x_n-x_s)^2 + (y_n-y_s)^2}} \right| + v_i, \quad (1)$$

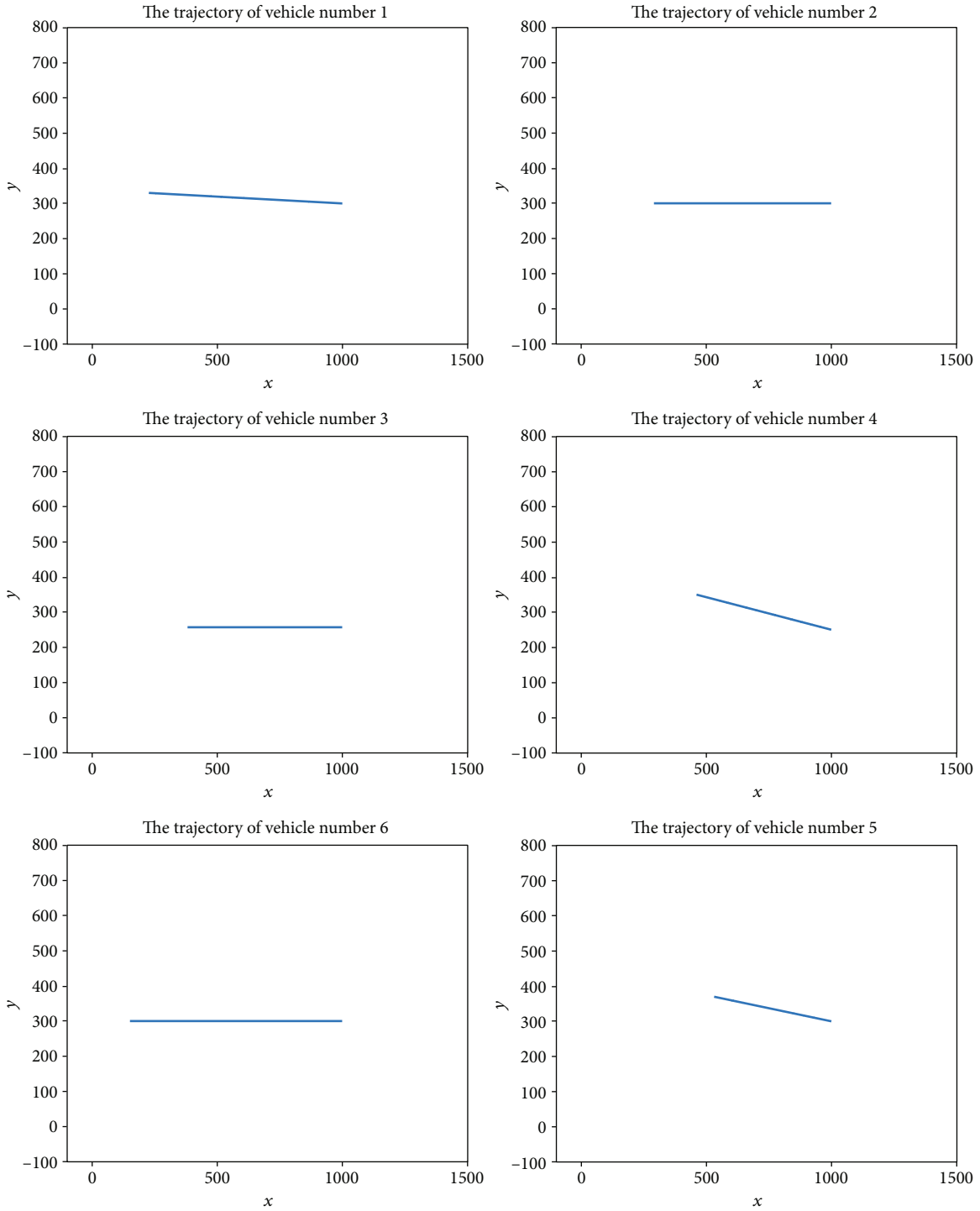


FIGURE 5: An example of the path of vehicles in the network (given the large number of vehicles in the network, only the path of a sample of vehicles is shown in the figure).

where x_{ij} and y_{ij} are the length and width of the i -th vehicle at the j -th moment, x_s and y_s are the length and width of the hole, n is the total vehicle number, and v_i is the average speed of each vehicle. The higher the DC value for a vehicle, the greater the delivery capacity for that vehicle, and the more likely it is to transmit data packets to the destination. Therefore, it is possible to select a vehicle with the highest amount of delivery capability among the vehicles that have

been designated as aligned vehicles and send data packets to that vehicle in order to transfer it to the destination hole or node. Table 6 shows the distance of each car to the hole node over time.

As can be seen in Table 6, the distance of aligned vehicle relative to the hole node decreases over time. Therefore, it can be said that aligned vehicles will approach the hole node at the end of the route. Table 7 also shows the difference

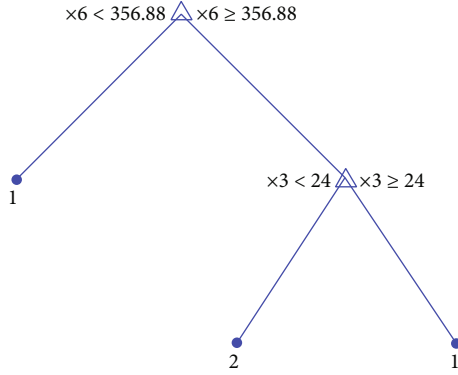


FIGURE 6: Steps to create a decision tree.

TABLE 5: Patterns created from the decision tree to discover the trajectory of vehicles.

Decision tree for classification
If "other vehicle distances" < 356.88, then class = 1
Else if "other vehicle distances" ≥ 356.88 and if "other vehicle distances" < 24, then node 4 else if x3 ≥ 24 then class = 2
Else if "other vehicle distances" ≥ 356.88 and if "other vehicle distances" < 24, then node 4 else if x3 < 24 then class = 1

between the distance of each car to the hole node at each moment of its travel path. In fact, the difference between the distance of each speed shows the tendency of each car towards the hole node.

As can be seen in Table 7, the difference in distances between aligned vehicles is gradually decreasing. Therefore, it can be said that aligned vehicles will approach the hole node at the end of the route.

Alignment vehicles according to the PRoPHET protocol will most likely deliver complete and secure packets to the destination vehicle or pit with high confidence, but the transmission of the packet by one vehicle along the entire route can be very delayed. Therefore, in this research, in order to reduce the delay of message transmission, the clustering method has been used, which we will describe and implement in the fourth module in the continuation of this chapter.

4.4. Implementation of the Fourth Module: Vehicle Clustering and Data/Information Transfer Routing. After predicting the trajectory of vehicles and shaping aligned vehicles and detecting nonaligned vehicles, in the proposed method, we will cluster the aligned vehicles. Vehicle clustering is based on the distance between vehicles and the accumulation of vehicles in road areas. The presence of vehicles in the road network may be complex or scattered. When vehicles are integrated, a clustered vehicle is first selected to cluster the vehicles on the road. The vehicle is selected based on three factors: alignment, maximum accumulation of vehicles, and minimum distance to other vehicles. Therefore, the following equation can be used to select the node [12]:

$$D(i, j) = \left| \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \right|, \quad (2)$$

$$HC(i) = \arg \max \left\{ \left(DC(i) + \left(\frac{1}{D(i, j)} \right) + \text{density}(i) \right) \right\} \text{ for } i \in n_{dc}, j \in C, \quad (3)$$

where $D(i, j)$ is the distance between the vehicles in the road network, $DC(i)$ is the amount of convergence tendency if the vehicle is aligned with the destination vehicle, and n_{dc} of the aligned vehicles and $\text{density}(i)$ is the density of the vehicles around the vehicle. Based on Equation (3), the vehicle with the highest amount of $HC(i)$ is selected as the threaded node. When a message is sensed on the network through sensors embedded in vehicles, the message is transmitted between aligned vehicles. Because aligned vehicles are clustered on the road network, messages are transmitted to the destination via clustered vehicles. The transmission of messages through threaded nodes according to Equation (3) assures us that the messages are transmitted to the destination at the appropriate speed in the direction of the destination and these messages are sent to the vehicles that are in the middle of the other vehicles and the crowd of vehicles around it is high. Therefore, if needed, that vehicles can send PAM to other vehicles in the network with the lowest communication cost and inform other cars about what is happening in the network.

In the case of vehicles, they may be spread across the road network and there are large distances between the vehicles in the road network. In this case, based on Equation (3), the vehicles that are in the direction of the destination node and move towards the destination with the highest speed have the highest amount of HC , and the appropriate option for selecting the vehicle for routing is selected. Table 8 shows the HC values for vehicles in the network.

As shown in Table 7, the value of the evaluation function is calculated to select the threaded node for all vehicles in the network. Now, based on the accumulation of nodes in the network sections, the nodes are clustered and the clustered nodes are identified according to Table 7. The vehicle with the highest value of the evaluation function is selected as the best vehicle in the desired cluster. Figure 7 shows the clustering of vehicles in the proposed method.

As shown in Figure 7, vehicles within the network are divided into clusters due to their accumulation in the road network. After clustering the vehicles within the network, the evaluation function of selecting the clustered node for these clusters is executed, and among these clusters, for those clusters and vehicles that are in the direction of the destination node, the clustered nodes based on Equation (3) and Table 7 are selected. As shown in Figure 7, in transferring information from the source node to the destination in the first cluster of vehicle no. 1, which according to Table 7 has the highest value of the evaluation function of the cluster node among the other vehicles in the cluster, the cluster node is selected and receives the message from the source node. In the next step, in order to transmit the message to the destination, the vehicle number 4, which

TABLE 6: The distance of each vehicle to the hole node over time.

	$t = 0$	$t = 100$	$t = 1000$	$t = 5000$	$t = 10000$	$t = 15000$	$t = 20000$	$t = 30000$	$t = 35493$
V1	445	442.9247	424.0585	340.2086	235.3963	130.5839	25.7716	183.8532	299
V2	410	408.0223	390.0436	310.1383	210.2566	110.3750	10.4933	189.2700	299
V3	370	368.2734	352.5769	282.8147	195.6119	108.4092	21.2064	153.1991	249
V4	190	188.7755	177.6434	128.1674	66.3225	4.4776	57.3673	181.0572	249
V5	100	98.8870	88.7693	43.8014	12.4085	68.6183	124.8282	237.2479	299
V6	550	547.6318	526.1030	430.4195	310.8151	191.2107	71.6062	167.6026	299
V7	250	248.6081	235.9546	179.7165	109.4190	39.1215	31.1761	171.7711	249
V8	220	218.4129	203.9842	139.8571	59.6982	20.4607	100.6196	260.9374	349
V9	360	358.0223	340.0436	260.1383	160.2566	60.3750	39.5067	239.2700	349
V10	80	80.7503	87.5716	117.8883	155.7841	193.6800	231.5759	307.3676	349
V11	100	100.5551	105.6013	128.0289	156.0634	184.0979	212.1323	268.2013	299
V12	430	429.0795	420.7114	383.5199	337.0306	290.5412	244.0519	151.0732	100
V13	240	240.1395	241.4074	247.0424	254.0863	261.1301	268.1740	282.2616	290
V14	180	180.3319	183.3495	196.7610	213.5253	230.2897	247.0540	280.5827	299
V15	0	0.6946	7.0086	35.0713	70.1496	105.2280	140.3063	210.4629	249
V16	340	339.7462	337.4386	327.1828	314.3630	301.5432	288.7234	263.0838	249
V17	430	427.2943	402.6972	293.3768	156.7263	20.0758	116.5747	389.8757	540
V18	400	397.6569	376.3564	281.6871	163.3506	45.0141	73.3224	309.9955	440
V19	180	178.2706	162.5487	92.6738	5.3302	82.0134	169.3570	344.0443	440
V20	300	299.3027	292.9632	264.7878	229.5686	194.3494	159.1302	88.6918	50
V21	300	297.9359	279.1711	195.7720	91.5232	12.7257	116.9745	325.4722	440
V22	50	50.9512	59.5982	98.0294	146.0684	194.1074	242.1464	338.2244	391
V23	70	71.1743	81.8500	129.2973	188.6064	247.9156	307.2247	425.8430	491
V24	180	181.0042	190.1330	230.7055	281.4212	332.1368	382.8525	484.2838	540
V25	170	171.1715	181.8218	229.1564	288.3247	347.4930	406.6612	524.9977	590
V26	350	350.8089	358.1627	390.8461	431.7004	472.5547	513.4089	595.1175	640
V27	230	230.4463	234.5036	252.5358	275.0761	297.6164	320.1567	365.2372	390
V28	350	350.3905	353.9406	369.7188	389.4416	409.1643	428.8871	468.3326	490
V28	450	449.9721	449.7185	448.5915	447.1827	445.7740	444.3652	441.5477	440

has the highest value of the evaluation function of the heading node in the next cluster, is selected as the heading node and receives the message from the previous heading vehicle, which is vehicle number 1. Finally, node 10, which is located in the next cluster in the direction of the destination node (hole node), receives messages and sends them to the destination. Table 9 shows the process of selecting recruited vehicles from aligned vehicles within the road network. More about this source text is required for additional translation information.

As shown in Table 8, the nodes that have the highest value of the header node selection evaluation function are selected as header nodes in each cluster.

5. Evaluation

In the proposed method, a method is used to transfer data and information based on vehicle trajectory forecasting and routing based on vehicle clustering in line with the destination, with the aim of reducing the delay of messages sent on the network. In fact, the sending of messages in the

VANET started right after the accident and received the first message due to the danger on the road by the vehicle of origin. Accordingly, by publishing messages on the network and sending sensed messages to the destination node, the number of messages sent also increases. Therefore, with increasing simulation time, the number of messages sent in the network will also increase. Figure 8 shows a diagram of messages sent over the network based on increasing time.

As shown in Figure 8, the number of messages sent on the network increases with increasing network time and sending security messages on the network to the destination node. Due to the fact that the vehicles carrying the message of the clusters in the clusters are in the road network, the transmission of information in the network takes place only between the clustered nodes until these packets are sent to the destination node reach. Therefore, the number of messages sent on the network increases at a slow pace. In fact, in the proposed method, the transfer of data packets is purposeful and sending packets in the form of all broadcasts is not an unnecessary waste of data packets.

TABLE 7: The difference between the distances of each vehicle at each moment of the route.

	$t = 0$	$t = 100$	$t = 1000$	$t = 5000$	$t = 10000$	$t = 15000$	$t = 20000$	$t = 30000$	$t = 35493$
V1	445	442.9247	424.0585	340.2086	235.3963	130.5839	25.7716	183.8532	299
V2	410	408.0223	390.0436	310.1383	210.2566	110.3750	10.4933	189.2700	299
V3	370	368.2734	352.5769	282.8147	195.6119	108.4092	21.2064	153.1991	249
V4	190	188.7755	177.6434	128.1674	66.3225	4.4776	57.3673	181.0572	249
V5	100	98.8870	88.7693	43.8014	12.4085	68.6183	124.8282	237.2479	299
V6	550	547.6318	526.1030	430.4195	310.8151	191.2107	71.6062	167.6026	299
V7	250	248.6081	235.9546	179.7165	109.4190	39.1215	31.1761	171.7711	249
V8	220	218.4129	203.9842	139.8571	59.6982	20.4607	100.6196	260.9374	349
V9	360	358.0223	340.0436	260.1383	160.2566	60.3750	39.5067	239.2700	349
V10	80	80.7503	87.5716	117.8883	155.7841	193.6800	231.5759	307.3676	349
V11	100	100.5551	105.6013	128.0289	156.0634	184.0979	212.1323	268.2013	299
V12	430	429.0795	420.7114	383.5199	337.0306	290.5412	244.0519	151.0732	100
V13	240	240.1395	241.4074	247.0424	254.0863	261.1301	268.1740	282.2616	290
V14	180	180.3319	183.3495	196.7610	213.5253	230.2897	247.0540	280.5827	299
V15	0	0.6946	7.0086	35.0713	70.1496	105.2280	140.3063	210.4629	249
V16	340	339.7462	337.4386	327.1828	314.3630	301.5432	288.7234	263.0838	249
V17	430	427.2943	402.6972	293.3768	156.7263	20.0758	116.5747	389.8757	540
V18	400	397.6569	376.3564	281.6871	163.3506	45.0141	73.3224	309.9955	440
V19	180	178.2706	162.5487	92.6738	5.3302	82.0134	169.3570	344.0443	440
V20	300	299.3027	292.9632	264.7878	229.5686	194.3494	159.1302	88.6918	50
V21	300	297.9359	279.1711	195.7720	91.5232	12.7257	116.9745	325.4722	440
V22	50	50.9512	59.5982	98.0294	146.0684	194.1074	242.1464	338.2244	391
V23	70	71.1743	81.8500	129.2973	188.6064	247.9156	307.2247	425.8430	491
V24	180	181.0042	190.1330	230.7055	281.4212	332.1368	382.8525	484.2838	540
V25	170	171.1715	181.8218	229.1564	288.3247	347.4930	406.6612	524.9977	590
V26	350	350.8089	358.1627	390.8461	431.7004	472.5547	513.4089	595.1175	640
V27	230	230.4463	234.5036	252.5358	275.0761	297.6164	320.1567	365.2372	390
V28	350	350.3905	353.9406	369.7188	389.4416	409.1643	428.8871	468.3326	490
V28	450	449.9721	449.7185	448.5915	447.1827	445.7740	444.3652	441.5477	440

TABLE 8: Evaluation function values for selecting header nodes for vehicles in the network.

Vehicle no.	The value of the header node evaluation function	Vehicle no.	The value of the header node evaluation function	Vehicle no.	The value of the header node evaluation function
1	12605.0025230644	11	11450.8443351707	21	10199.0068933721
2	9762.84175823016	12	9420.60520826243	22	8167.08652900570
3	9357.58973116927	13	12413.5949490017	23	8973.49306016971
4	18301.4087332702	14	1445.61204611270	24	8214.88333766602
5	8525.81733039176	15	11292.5886888899	25	9094.22389569419
6	12256.4692509958	16	12411.8153697601	26	11010.5772186310
7	9170.01142069164	17	12516.0434281110	27	12089.7135617565
8	8758.10006652234	18	11499.9662504847	28	12442.5178217940
9	9300.62029930559	19	9212.94159417572	29	11234.9592498159
10	14237.0772406787	20	10526.5165883803		

Therefore, another criterion that has been evaluated in the present study is the number of lost packets in the network. An ACK message has been received from the vehicle when the source vehicle sends a message to the vehicle. After

receiving the confirmation message, the packets are sent to the header node. The same thing is repeated for other vehicles until the message reaches its destination. Due to the limited capacity of the queue in the vehicles, the received

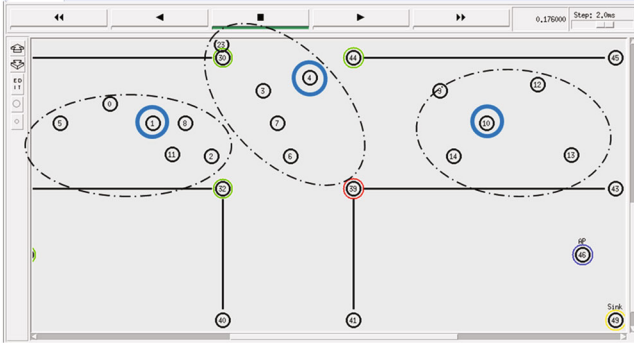


FIGURE 7: Clustering of vehicles within the network.

TABLE 9: The process of selecting header vehicles.

The cluster head in first connection is node number 4
The cluster head in second connection is node number 1
The cluster head in third connection is node number 10

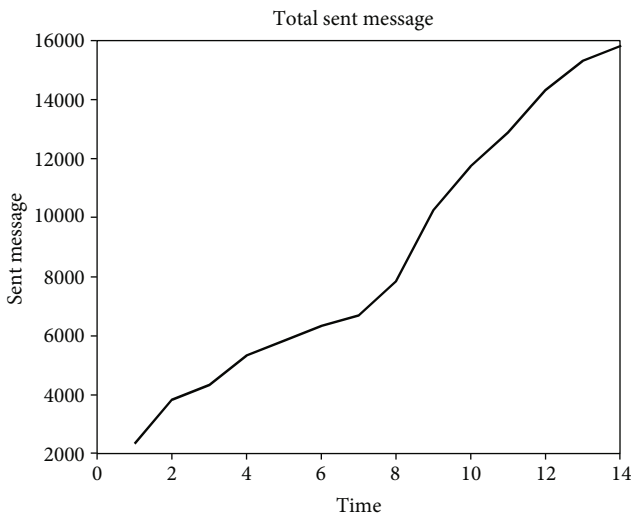


FIGURE 8: Graph of messages sent in the network.

messages should be stored in the queue of the vehicle to be examined in turn. When the number of these messages exceeds the queue size, the excess messages are lost and the accuracy of the information transmitted in the network may be damaged. This event will increase the number of lost packets in the VANET, which is considered as one of the weaknesses of routing methods in VANETs. Figure 9 shows the number of lost message packets as the VANET increases with time. More about this source text is required for additional translation information.

As shown in Figure 9, the number of messages lost in the proposed vehicle network increases slightly with increasing time, and finally, the total number of messages lost during the simulation of the proposed scenario only 130 messages, which is only 0.8111% of the total messages sent on the network.

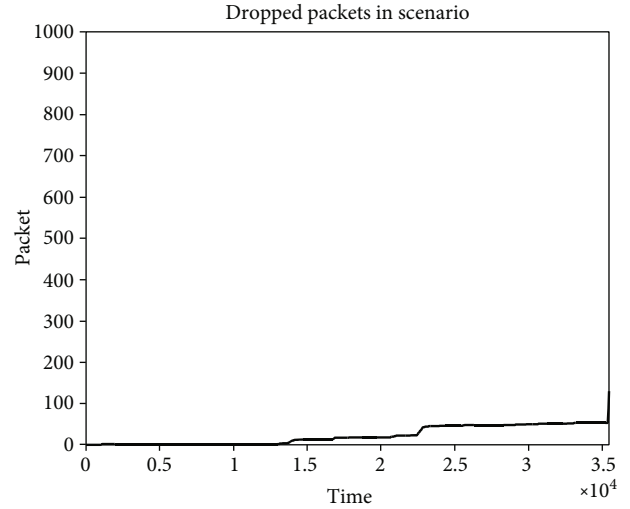


FIGURE 9: Number of lost message packets.

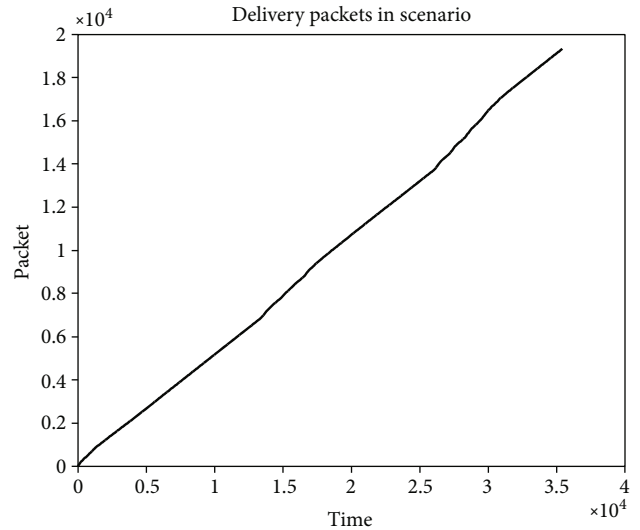


FIGURE 10: Diagram of delivered packages.

Due to the small number of lost packets, another criterion that has been examined in this method is the delivery rate of data packets. Package delivery rate is the ratio of the number of delivered messages to the sent messages. Figure 10 shows the diagram of packages delivered based on the proposed scenario with increasing time. Figure 11 also shows the package delivery rate in the proposed scenario.

As shown in Figures 10 and 11, a large number of healthy packages sent have been delivered to the destination and the packet delivery rate tends to be a fixed amount. By calculating the average package delivery rate over time, this amount is about 88.56%.

The last criterion that has been examined in this research is the delay of packets sent in the VANET, which is one of the most important criteria in this type of network. Due to the fact that delays in VANETs cause a loss of time for drivers to react to road accidents and disrupt real-time applications in VANETs, so any delay in less VANETs, the

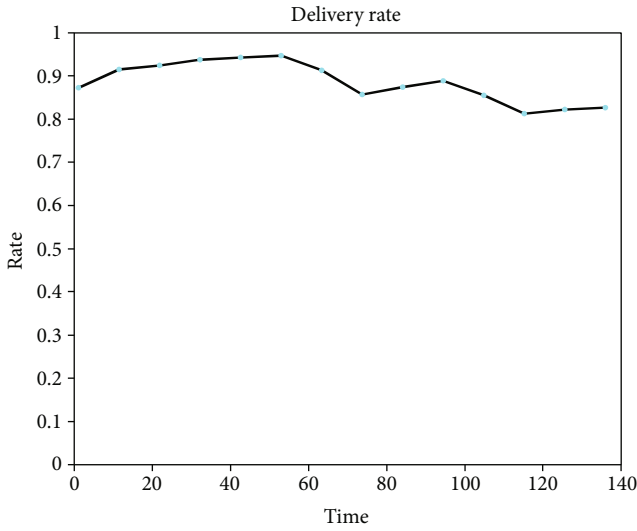


FIGURE 11: Package delivery rate.

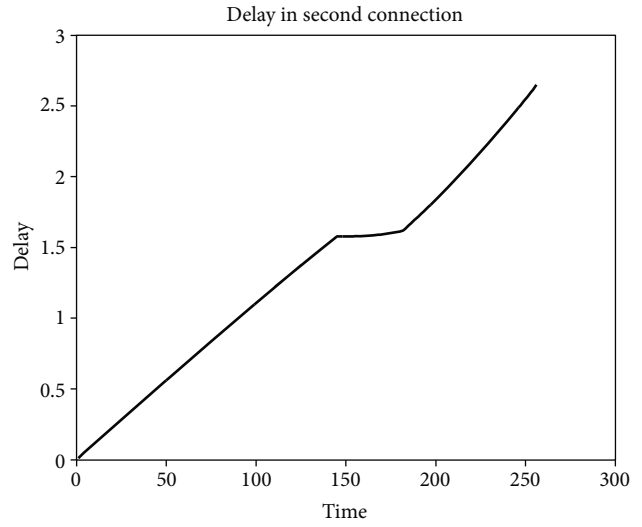


FIGURE 13: Delay in message transmission in the second stage of communication.

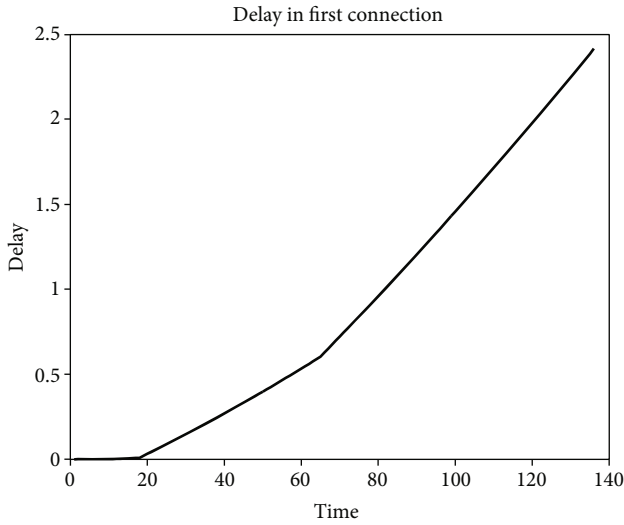


FIGURE 12: Message transmission delay in the first stage of communication.

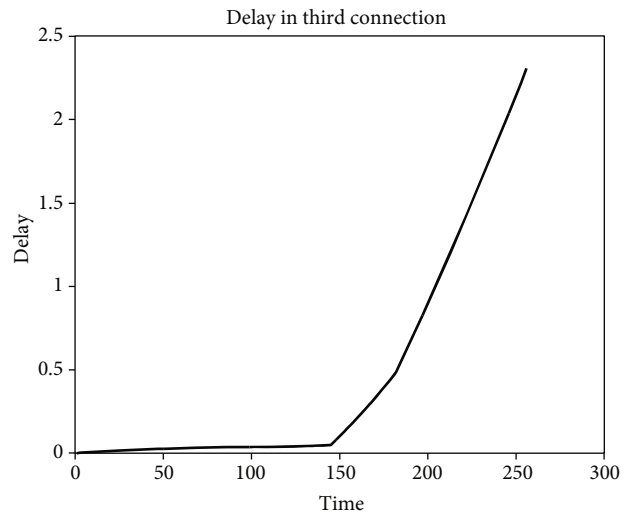


FIGURE 14: Delay in message transmission in the last stage of communication.

routing method used is more effective. In the proposed method, since the messages are transmitted between the node clusters, therefore, the delay of message transmission between the clusters must be calculated. Figure 12 shows the message transmission delay in the first stage, Figure 13 shows the message transmission delay in the second stage, and finally, Figure 14 shows the message transmission delay in the last stage.

As shown in Figures 12–14, the message transmission delay between threaded nodes tends to be constant. Therefore, by calculating the maximum delay during message transmission between headers, the total network message delay can be expressed as 24.566 ms.

5.1. Comparing the Proposed Method with Previous Methods. Routing methods in VANET can be evaluated and compared

based on different evaluation criteria. Since reducing the delay and increasing the data delivery rate are the main goals in the proposed method, we compare the proposed method with other state-of-the-art methods based on these criteria. The proposed method tries to reduce the delay in message transmission by choosing the optimal cluster head vehicles based on density criteria and close distance to other vehicles. On the other hand, by finding aligned vehicles and predicting the movement path of neighbor vehicles, it tries to transmit messages with high reliability, which reduces the loss rate of packets and increases the data delivery rate. Therefore, in Figure 15, the data delivery rate in the proposed method is compared with previous methods, including the method based on the farthest distance (FD) [34], the expected progress distance (EPD) [35], and the two-way stable communication (BDSC) [19].

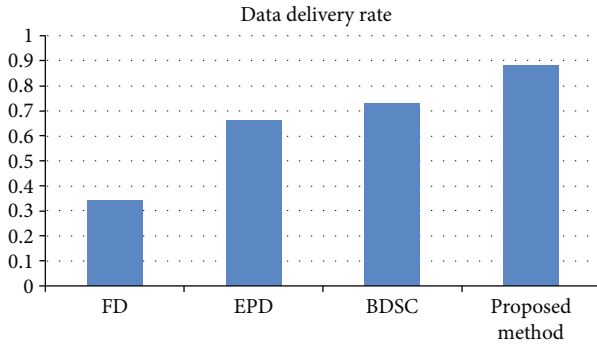


FIGURE 15: Comparing the delivery rate of the proposed method with previous methods.

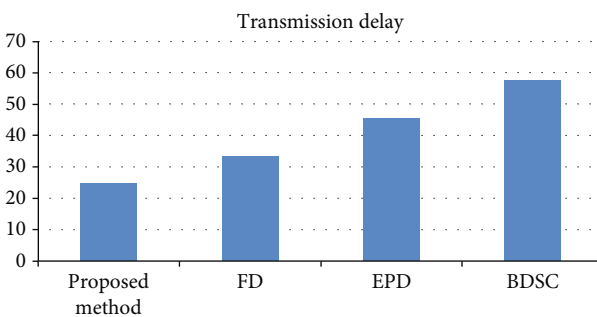


FIGURE 16: Comparing the delay of message transmission in the proposed method with previous methods.

As shown in Figure 15, the proposed method has a significant improvement in terms of package delivery rate compared to previous methods. Figure 16 also compares the message transmission delay in the proposed method with the previous methods included.

As shown in Figure 16, the proposed method has a lower value in terms of message transmission delay in the network compared to previous methods.

6. Conclusion and Discussion

The dissemination of real-time information in automotive case networks (VANET), due to the dynamic nature and rapid movement of vehicles, has become one of the research challenges in this field, which has attracted the attention of many researchers in this field. The data published in this network may not reach the destination in time due to changes in distances, redirects, and unforeseen movements by vehicles and may cause excessive delays. Therefore, informing the vehicles in the network quickly will be very useful in preventing accidents on the road. Therefore, in order to prevent disruption of the timely information process in automotive networks, strategies should be considered to reduce network latency. Also, previous related methods in the literature for data transmission in VANET have focused more on reliability and trust, and little attention has been paid to the prediction of vehicle movement and path pattern discovery in these types of networks. Therefore, in order to overcome this issue, the proposed method has presented

an approach based on clustering and frequent pattern discovery in predicting the movement path of vehicles. In this research, a new method for data transfer and information in VANETs was proposed, which is based on discovering the trajectory of vehicles within the network and predicting the trajectory of future vehicles and sending messages based on clustering of aligned vehicles. The contributions of this paper are included, using the decision tree classification method to classify vehicles in aligned and not aligned vehicles, using sequential pattern mining to discover vehicle movement patterns, clustering vehicles in the network based on the current position to detect neighbor vehicles, and finding best cluster head to transfer messages using combination of above three steps. This research proposed a new method to discover the trajectory and clustering of vehicles, from the distance of each vehicle to the destination node at any time in order to extract the acceleration of the vehicle to the destination, vehicle speed, and vehicle distance to another vehicle, and the density of vehicles around the vehicle is used. The simulation results show that the proposed method with a delivery rate of 88.56% has significantly improved in terms of package delivery rate compared to previous methods. Also, the proposed method with a total delay of 24.566 ms in terms of message transmission delay in the network has a lower value compared to previous methods.

The limitation of the proposed method is to store the history of vehicle's movement on any type of road, and based on that, it produces movement patterns of vehicle and predicts the path of new vehicles. The proposed method stores the history of vehicle movement on any type of road, and based on that, it produces movement patterns of vehicle and predicts the path of new vehicles. It is natural that the more data related to vehicle movements in the proposed method, the more accurate in path pattern discovery. This method is more useful on busy roads where there are more vehicles than quiet roads. The accuracy of the proposed method will be lower in predicting the path pattern of the vehicles on quiet roads.

Data Availability

The simulated data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by 03 Special Project and 5G Project of Jiangxi Province under Grant 20204ABC03A18.

References

- [1] B. Jarupan and E. Ekici, "Mobility management for efficient data delivery in infrastructure-to-vehicle networks," *Computer Communications*, vol. 35, no. 18, pp. 2274–2280, 2012.
- [2] M. Bakhouya, J. Gaber, and P. Lorenz, "An adaptive approach for information dissemination in vehicular ad hoc networks,"

- Journal of Network and Computer Applications*, vol. 34, no. 6, pp. 1971–1978, 2011.
- [3] Z. Peng, M. S. Jabloo, Y. D. Navaei et al., “An improved energy-aware routing protocol using multiobjective particular swarm optimization algorithm,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6677961, 16 pages, 2021.
 - [4] X. Zhao, W. Zhong, and Y. D. Navaei, “A novel energy-aware routing in wireless sensor network using clustering based on combination of multiobjective genetic and cuckoo search algorithm,” *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6939868, 14 pages, 2022.
 - [5] Z. Peng, M. Rastgari, Y. D. Navaei et al., “TCDABCF: a trust-based community detection using artificial bee colony by feature fusion,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 6675759, 19 pages, 2021.
 - [6] S. Einy, C. Oz, and Y. D. Navaei, “The anomaly- and signature-based IDS for network security using hybrid inference systems,” *Mathematical Problems in Engineering*, vol. 2021, Article ID 6639714, 10 pages, 2021.
 - [7] S. Einy, C. Oz, and Y. D. Navaei, “Network intrusion detection system based on the combination of multiobjective particle swarm algorithm-based feature selection and fast-learning network,” *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 6648351, 12 pages, 2021.
 - [8] S. Einy, C. Oz, and Y. D. Navaei, “IoT cloud-based framework for face spoofing detection with deep multicolor feature learning model,” *Journal of Sensors*, vol. 2021, Article ID 5047808, 18 pages, 2021.
 - [9] Y. Nanehkaran, Z. Licai, J. Chen et al., “Anomaly detection in heart disease using a density-based unsupervised approach,” *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6913043, 14 pages, 2022.
 - [10] S. Jamali and Y. D. Navaei, “A two-level product recommender for E-commerce sites by using sequential pattern analysis,” *International Journal of Integrated Engineering*, vol. 8, no. 1, 2016.
 - [11] Y. D. Navaei and M. Afzali, “Dihedral product recommendation system for E-commerce using data mining applications,” *International Journal of Computer & Information Technologies (IJOCIT)*, vol. 3, pp. 610–631, 2015.
 - [12] Y. D. Navaei and M. Afzali, “A survey on product recommendation system in E-commerce,” *International Journal of Computer & Information Technologies (IJOCIT)*, pp. 585–600, 2014.
 - [13] R. Zhang, H. Shu, and Y. D. Navaei, “Load balancing in edge computing using integer linear programming based genetic algorithm and multilevel control approach,” *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 6125246, 22 pages, 2022.
 - [14] Y. Nanehkaran, Z. Licai, J. Chen et al., “Diagnosis of chronic diseases based on patients’ health records in IoT healthcare using the recommender system,” *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 5663001, 14 pages, 2022.
 - [15] H. Shahwani, S. Attique Shah, M. Ashraf, M. Akram, J. (. P.). Jeong, and J. Shin, “A comprehensive survey on data dissemination in vehicular ad hoc networks,” *Vehicular Communications*, vol. 34, article 100420, 2022.
 - [16] T. Nadeem, P. Shankar, and L. Iftode, “A comparative study of data dissemination models for VANETs,” in *2006 Third Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services*, pp. 1–10, San Jose, CA, USA, 2006.
 - [17] M. M. Hamdi, O. A. R. Al-Dosary, O. A. S. Alrawi, A. S. Mustafa, M. S. Abood, and M. S. Noori, “An overview of challenges for data dissemination and routing protocols in VANETs,” in *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, pp. 1–6, Ankara, Turkey, 2021.
 - [18] M. Kakkasageri and S. S. Manvi, “Regression based critical information aggregation and dissemination in VANETs: a cognitive agent approach,” *Vehicular Communications*, vol. 1, no. 4, pp. 168–180, 2014.
 - [19] O. Rehman, M. Ould-Khaoua, and H. Bourdoucen, “An adaptive relay nodes selection scheme for multi-hop broadcast in VANETs,” *Computer Communications*, vol. 87, pp. 76–90, 2016.
 - [20] A. Louazani and L. Sekhri, “Petri net model for connectivity maintenance in VANET clustering-based routing algorithm,” in *2016 International Conference on Advanced Aspects of Software Engineering (ICAASE)*, pp. 92–97, Constantine, Algeria, 2016.
 - [21] A. Jalooli, M. Song, and X. Xu, “Delay efficient disconnected RSU placement algorithm for VANET safety applications,” in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, San Francisco, CA, USA, 2017.
 - [22] A. Abuashour and M. Kadoch, “Performance improvement of cluster-based routing protocol in VANET,” *Ieee access*, vol. 5, pp. 15354–15371, 2017.
 - [23] R. Shahidi and M. H. Ahmed, “On the analytical calculation of the probability distribution of end-to-end delay in a two-way highway VANET,” *IEEE access*, vol. 6, pp. 1109–1125, 2018.
 - [24] M. M. Mohammed Nasr, A. Abdelgader, Z. G. Wang, and L. F. Shen, “VANET clustering based routing protocol suitable for deserts,” *Sensors*, vol. 16, no. 4, p. 478, 2016.
 - [25] S. P. Ardakani, “ACR: a cluster-based routing protocol for VANET,” 2018, <https://arxiv.org/abs/1805.07565>.
 - [26] G. Rathee, A. Sharma, R. Iqbal, M. Aloqaily, N. Jaglan, and R. Kumar, “A blockchain framework for securing connected and autonomous vehicles,” *Sensors*, vol. 19, no. 14, p. 3165, 2019.
 - [27] M. Poongodi, M. Hamdi, A. Sharma, M. Ma, and P. K. Singh, “DDoS detection mechanism using trust-based evaluation system in VANET,” *IEEE Access*, vol. 7, pp. 183532–183544, 2019.
 - [28] A. Sharma and N. Kumar, “Third eye: an intelligent and secure route planning scheme for critical services provisions in Internet of vehicles environment,” *IEEE Systems Journal*, vol. 16, no. 1, pp. 1217–1227, 2022.
 - [29] C. Wu, S. Ohzahata, and T. Kato, “Flexible, portable, and practicable solution for routing in VANETs: a fuzzy constraint Q-learning approach,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 9, pp. 4251–4263, 2013.
 - [30] O. S. Oubbati, N. Chaib, A. Lakas, P. Lorenz, and A. Rachedi, “UAV-assisted supporting services connectivity in urban VANETs,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3944–3951, 2019.
 - [31] O. Sami Oubbati, N. Chaib, A. Lakas, S. Bitam, and P. Lorenz, “U2RV: UAV-assisted reactive routing protocol for VANETs,” *International Journal of Communication Systems*, vol. 33, no. 10, article e4104, 2020.
 - [32] L. Zhao, Z. Li, A. Y. al-Dubai et al., “A novel prediction-based temporal graph routing algorithm for software-defined vehicular networks,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13275–13290, 2022.

- [33] L. Zhao, T. Zheng, M. Lin, A. Hawbani, J. Shang, and C. Fan, "SPIDER: a social computing inspired predictive routing scheme for softwarized vehicular networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 9466–9477, 2022.
- [34] W. B. Jaballah, M. Conti, M. Mosbah, and C. E. Palazzi, "Fast and secure multihop broadcast solutions for intervehicular communication," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 433–450, 2014.
- [35] H. Okada, A. Takano, and K. Mase, "A Proposal of Link Metric for Next-Hop Forwarding Methods in Vehicular Ad Hoc Networks," in *2009 6th IEEE Consumer Communications and Networking Conference*, pp. 1–5, Las Vegas, NV, USA, 2009.