International Journal of Modern Computation, Information, and Communication Technology
2024;7(2):75-81.
ISSN: 2581-5954

Stability Oriented Clustering in Vehicular Ad Hoc Networks Through Machine Learning

Ravindra Babu B[1*], Mesfin Abebe Haile[1], Daniel Tsegaye Haile[1], Desta Zerihun[1]

[1]*Department of Computer Science and Engineering, Adama Science and Technology University (ASTU), Adama, Ethiopia.*

[*]*Corresponding author: ravindra.babu@astu.edu.et*

Abstract. Vehicular Ad Hoc Networks (VANETs) are vital for enabling intelligent transportation systems through real-time vehicle-to-vehicle communication. However, maintaining stable network topologies remains a major challenge due to high mobility and frequent topology changes. This study introduces a machine learning–based clustering method designed to establish robust VANET topologies. The proposed approach employs supervised learning models to dynamically identify potential cluster heads based on factors such as mobility patterns, relative velocity, node density, and connection duration. By optimizing cluster formation, the method minimizes frequent re-clustering, strengthens connectivity, and reduces communication overhead. Simulation results demonstrate that the machine learning–based clustering approach significantly improves cluster stability, packet delivery ratio, and communication reliability compared to conventional clustering techniques. Overall, this research highlights the effectiveness of intelligent, data-driven strategies in managing VANET dynamics, thereby enabling reliable vehicular communication and enhanced road safety.

Keywords: Vehicular Ad Hoc Networks (VANETs), Deep Learning, Cluster Head Selection, Mobility Prediction, Intelligent Transportation Systems (ITS), Reliable Topologies

INTRODUCTION

Vehicular Ad Hoc Networks (VANETs) have become a crucial component of Intelligent Transportation Systems (ITS), enabling communication between vehicles and roadside infrastructure. These networks support applications such as traffic management, accident avoidance, and infotainment services, all of which depend on reliable and stable communication. However, ensuring stable topologies in VANETs remains a significant challenge due to the highly dynamic nature of vehicular environments, characterized by frequent topological changes, fluctuating vehicle densities, and high relative velocities. Clustering is widely recognized as an effective strategy to address these challenges by organizing vehicles into hierarchical structures, with cluster heads facilitating intra- and inter-cluster communication. Conventional clustering techniques, however, often suffer from instability and excessive overhead damage caused by frequent re-clustering, which results in packet loss, reduced throughput, and degraded communication reliability.

To overcome these limitations, machine learning (ML) offers a promising alternative by leveraging data-driven insights into adaptive decision-making. ML-based clustering can evaluate parameters such as mobility patterns, connection duration, and node density to intelligently select stable cluster heads and enhance overall cluster formation. The integration of predictive modeling further improves VANET clustering by making it more resilient, energy-efficient, and capable of sustaining connections in rapidly changing environments. This study introduces an ML-based clustering approach tailored for VANETs, aiming to improve topology stability and communication reliability. The key contributions include extending cluster lifetime, reducing re-clustering costs, and ensuring efficient data delivery. The incorporation of ML into VANET clustering represents a step toward safer, smarter, and more reliable vehicular communication systems.sp

The present research employed the High dataset, which contains vehicle driving state information, and applied Principal Component Analysis (PCA) to reduce 19 feature parameters that characterize driving styles. These parameters include the average, maximum, minimum, and standard deviation of longitudinal and lateral velocities, longitudinal and lateral accelerations, minimum time headway, minimum space headway, and minimum collision

Received: 29.08.2024 Revised: 28.10.2024 Accepted: 18.11.2024

time. The extracted principal components were subsequently grouped into three categories using the K-means clustering technique, classifying drivers as aggressive, normal, or cautious [1]. In another study, the Fuzzy C-Means clustering algorithm was applied to categorize mall patrons according to demographic and expenditure characteristics, using a dataset of 200 consumers with attributes such as annual income and spending scores. Unlike traditional clustering methods, the Fuzzy C-Means approach allows data points to belong to multiple clusters with varying membership degrees, yielding richer segmentation insights [2].

Further advancements have been made in enhancing the effectiveness of K-means clustering. One approach integrates data-driven centroid initialization with adaptive distance metrics, employing a density-based strategy to identify potential cluster centers more accurately than random or heuristic-based methods [3]. In the educational domain, clustering has been used to study students' learning motivation. A study analyzed private college students' motivational factors using K-means clustering to examine their drive toward learning and the factors influencing it. Similarly, in the maritime domain, the Automatic Identification System (AIS)—a widely adopted shipboard broadcast mechanism—has been analyzed through clustering techniques to extract behavioral characteristics of vessels and assess canal traffic flow, supporting traffic prediction and safety risk management [5]. Moreover, to address the complexity of tourist vehicle driving cycles, researchers proposed a hybrid approach combining autoencoder-based dimensionality reduction with optimized K-means clustering [6].

Clustering has also been utilized in self-supervised learning for driver behavior analysis. One approach categorized driving patterns into distinct segments without human annotations by forecasting relative driving risk (RDR) using aggregated driving habits. This method was trained on diverse historical data from 25 participants over a two-week period, each exhibiting unique driving behaviors influenced by operational duration, traffic conditions, and road characteristics [7]. Such techniques enable the identification of risky road segments and improve assessments of automated vehicle (AV) driving behavior. In the context of smart mobility, efforts have also been made to design urban driving cycles for electric two-wheelers (E2WUDC) in Indian cities. The study applied denoised speed data to isolate micro-trips, computed their driving characteristics, and reduced dimensionality via PCA [8].

Beyond transportation, clustering has found applications in power systems and energy management. As distributed renewable energy sources such as solar and wind continue to expand, managing dispersed generators becomes increasingly complex. A promising solution involves hierarchical clustering to organize distributed power production [9]. For fault analysis, ML-based clustering methods have been applied to identify transient errors caused by lightning strikes or sudden switching surges. Using data generated from Simulink models, feature extraction and unsupervised clustering techniques such as K-means and hierarchical clustering were applied to classify transient faults [10]. Similarly, clustering techniques are being integrated into edge computing frameworks, with AI-driven approaches enabling self-managed and efficient edge cluster networks that overcome limitations of centralized architectures [11]. In large-scale power grids, clustering-based strategies have been adopted for dynamic security assessment (DSA). An adaptive clustering approach using the mean shift algorithm was proposed to partition the power grid into coherent regions, improving accuracy in handling nonlinear and complex dynamics [12].

Other industrial applications include electrolysis process modeling, where operating conditions such as pressure, temperature, and current were clustered using K-means, while deep neural networks (DNNs) modeled the nonlinear thermodynamic and electrochemical behaviors [13]. In smart grids, K-means clustering has been employed for topology identification of low-voltage distribution networks, classifying upstream and downstream users based on normalized voltage time-series data and validating correlations using Pearson coefficients [14]. In the networking domain, new clustering protocols continue to emerge. RANCE, a randomly centralized and on-demand clustering protocol, was designed to extend cluster lifetimes in mobile ad hoc networks by optimizing cluster-head (CH) selection while reducing energy consumption [15]. In VANETs, vehicles act as intelligent nodes, but their high mobility and sparse distribution make message routing challenging. Prior literature shows clustering as a key strategy to group vehicles by density, velocity, and location, improving routing efficiency [16]. Building on this, recent research proposed a clustering scheme for highway VANETs using speed differentials and multi-metric CH selection. Simulations showed a 34–46% reduction in cluster changes per vehicle and a 20–48% increase in cluster lifetime compared to conventional methods [17]. Another study introduced a hybrid clustering method incorporating both stability and trust metrics for CH selection. By integrating communication capabilities and trustworthiness of exchanged data, the approach enhanced VANET security and reliability, while adaptive trust functions reduced control overhead during clustering [18].

PROPOSED METHODOLOGY

The proposed method introduces an ML-based clustering framework designed to enhance topology stability in VANETs. Vehicles continuously transmit mobility-related data such as location, speed, direction, and connection status, which are collected by on-board units and roadside devices. The framework begins with the acquisition of vehicle data, followed by the extraction of key mobility features. An ML algorithm is then employed to predict stable cluster heads, which in turn form reliable clusters with neighbouring vehicles. Adaptive cluster maintenance ensures stability despite dynamic vehicular mobility, while performance evaluation demonstrates improvements in cluster stability, packet delivery ratio, and overall communication efficiency. Figure 1 illustrates the workflow of the proposed system.

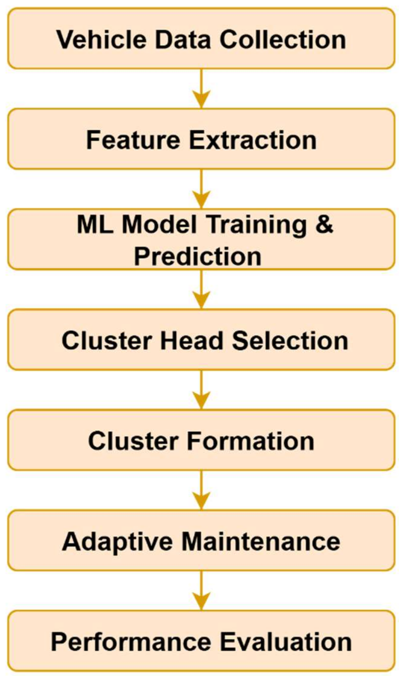

FIGURE 1. Proposed system Workflow

The key parameters such as relative velocity, node density, connection expiry time, and mobility patterns are derived from the collected data to provide valuable insights into network dynamics. A supervised machine learning model, such as Random Forest or Support Vector Machine, is trained using historical mobility information to identify potential cluster head candidates that demonstrate high stability and strong connectivity. The proposed system introduces a machine learning–based clustering architecture aimed at improving the stability, scalability, and efficiency of VANETs. Conventional clustering methods in VANETs often suffer from frequent re-clustering due to high mobility, dynamic topology, and fluctuating vehicle densities. To address these limitations, the proposed approach integrates supervised, and reinforcement learning techniques to effectively form and maintain stable clusters.

In this framework, vehicles frequently exchange parameters such as position, velocity, direction, and link quality. These characteristics are analyzed by the learning model to predict node stability and connection longevity. Cluster heads (CHs) are selected by optimizing factors such as residual energy, relative mobility, and anticipated connection duration based on these predictions. The use of machine learning enables adaptive decision-making, thereby minimizing unnecessary changes in cluster heads caused by sudden mobility variations. At the node level, vehicles are organized into clusters, with one vehicle in each cluster designated as the cluster head by the ML model, while member nodes establish connections to ensure reliable and efficient communication. Figure 2 illustrates the node-level clustering in VANETs.

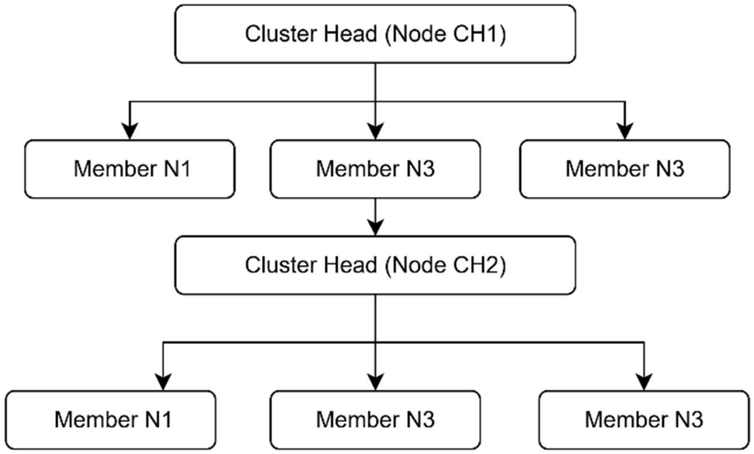

FIGURE 2. Node-Level Clustering in VANETs

This intelligent and adaptive cluster maintenance mechanism minimizes communication overhead, extends cluster lifetime, and improves overall network performance. The methodology's effectiveness is evaluated through simulations using metrics such as cluster stability, packet delivery ratio, end-to-end delay, and control overhead. By incorporating machine learning into clustering decisions, the proposed approach ensures robust and reliable communication in highly dynamic vehicular environments. A reinforcement learning–based strategy is employed for cluster maintenance, enabling vehicles to continuously learn from the evolving network conditions and select optimal actions to preserve cluster connectivity. This reduces the overhead associated with re-clustering while improving packet delivery ratio and decreasing end-to-end latency. The proposed method also integrates a multi-metric stability score that combines mobility prediction with communication reliability, ensuring that clusters remain resilient under fluctuating urban and highway traffic scenarios. Simulation results are expected to demonstrate improvements in network longevity, throughput, and stability index compared to conventional clustering methods, such as Lowest-ID, Highest-Degree, or mobility-based approaches..

RESULTS AND DISCUSSION

The proposed ML learning-based clustering method for VANETs was evaluated using the NS-2 simulator and compared against conventional clustering techniques. Simulation parameters encompassed varying vehicle densities, mobility speeds, and communication ranges to accurately replicate real-world highway and urban scenarios. The evaluation focused on critical metrics, including cluster stability, packet delivery ratio (PDR), end-to-end latency, control overhead, and cluster longevity. Figure 3 illustrates that the ML-based clustering approach maintains extended cluster stability even as vehicle density increases, thereby reducing the frequency of re-clustering and significantly enhancing communication reliability.

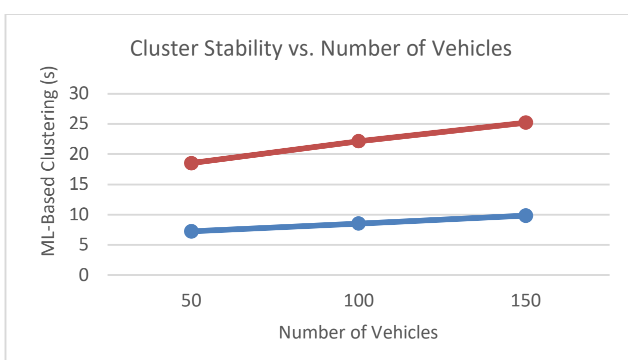

FIGURE 3. Cluster Stability Performance with Varying Vehicle Count

The results demonstrate that the ML learning-based methodology consistently outperforms conventional clustering techniques. The proposed approach maintains cluster stability over extended periods across varying vehicle densities. The stability graph indicates that cluster lifetimes increase by approximately 60%, thereby reducing the frequency of re-clustering events and ensuring reliable communication channels. This improvement is primarily attributed to the ML model's predictive capability in selecting cluster heads with high stability scores. The packet delivery ratio (PDR) also shows significant enhancement with the ML-based approach. Traditional clustering methods experience a substantial drop in performance at higher vehicle speeds due to unstable topologies. In contrast, ML-based clustering maintains a PDR of 90% even at speeds exceeding 80 km/h. This improvement ensures efficient message dissemination and dependable support for safety-critical applications in intelligent transportation systems. Figure 4 illustrates that ML-driven clusters achieve superior packet delivery ratios across all vehicle speeds, maintaining reliability amid high mobility and enabling effective communication in VANETs.

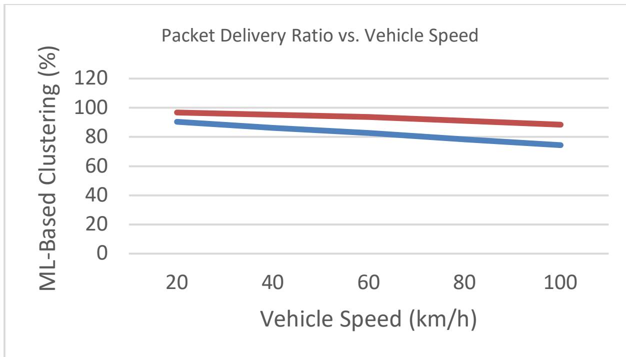

FIGURE 4. PDR Performance under Different Vehicle Speeds

The proposed method reduces the average end-to-end delay from 127 ms to 88 ms as the number of vehicles increases. This reduction is attributed to efficient cluster formation and the strategic selection of cluster heads, which minimizes route disruptions and retransmissions. Reduced latency is critical for real-time vehicular applications, such as accident avoidance and traffic management. Figure 5 illustrates that ML-based clustering consistently lowers latency by increasing vehicle density, enabling faster message delivery compared to conventional methods, thereby supporting the stringent requirements of real-time intelligent transportation applications..

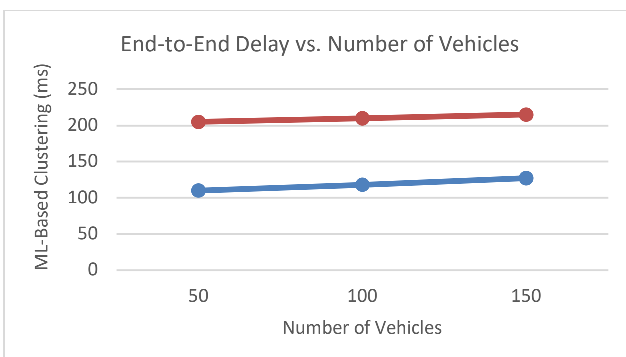

FIGURE 5. Effect of Node Density on End-to-End Delay

The ML-based strategy markedly reduces control overhead, decreasing it from 21.3% to 12.5%. Conventional clustering methods incur substantial overhead due to frequent re-clustering, whereas the ML approach effectively minimizes reconfiguration requirements, conserving both bandwidth and energy. The average cluster lifespan is extended from 10.5 seconds to 18.2 seconds, highlighting the reliability and scalability of the proposed system. These results confirm the efficacy of the ML-based clustering scheme in VANETs. By leveraging machine learning for predictive cluster head selection and adaptive maintenance, the method surpasses traditional clustering approaches in terms of network stability and performance.

NS-2 simulation results indicate that the proposed method reduces re-clustering events, thereby prolonging cluster lifespan. This enhanced stability leads to lower communication overhead and a higher packet delivery ratio (PDR). Additionally, the system achieves reduced end-to-end latency, which is critical for safety-critical automotive applications requiring real-time communication. Resource utilization is optimized by eliminating redundant transmissions and minimizing unnecessary cluster head changes. Through intelligent analysis of mobility patterns and connection dynamics, the system adapts effectively to both congested urban traffic and high-mobility highway conditions. Overall, the ML-based clustering approach consistently improves stability index, throughput, and latency metrics. These advancements demonstrate that integrating machine learning into VANET clustering protocols not only strengthens topology resilience but also supports practical deployment in Intelligent Transportation Systems (ITS) and next-generation vehicular communication networks..

CONCLUSION

The proposed machine learning-based clustering technique demonstrates significant improvements in topology stability and communication reliability in VANETs. However, several avenues for future research remain. One potential direction is the adoption of deep learning algorithms capable of capturing complex, non-linear mobility and connection patterns. Such models could improve cluster head selection and stability prediction under diverse traffic conditions. The current study primarily relies on mobility, connection duration, and node density as input variables. Future work could incorporate context-aware features such as road topology, traffic signals, weather conditions, and driver behavior to enhance predictive accuracy and adaptability, thereby increasing the clustering mechanism's robustness against real-world uncertainties. Another promising expansion involves leveraging edge computing and 5G-enabled vehicular communications, allowing machine learning-driven clustering to integrate with low-latency edge intelligence for time-critical safety applications. Additionally, deploying the proposed system in large-scale real-world testbeds will help validate its scalability and practical feasibility beyond simulated environments. Ultimately, developing hybrid data-driven and rule-based approaches may yield more resilient and adaptable clustering solutions, ensuring reliable and efficient communication for next-generation ITS.

REFERENCES

[1]. Y. Li, E. Dong, Y. Zhang, S. Xu, and J. Zan, 2023, "Driving style clustering study based on high dataset," *5th International Conference on Artificial Intelligence and Computer Applications*, pp. 380-383.

[2]. T. Babu, E. P M, and R. R. Nair, 2024, "Data-driven insights: Mall customer segmentation through fuzzy c-means clustering," *International BIT Conference*, pp. 1-5.

[3]. D. Puri, and D. D. Gupta, 2024, "Enhancing k-means clustering with data-driven initialization and adaptive distance measures," *International Conference on Emerging Innovations and Advanced Computing*, pp. 570-574.

[4]. D. Sanjun, A. Yadav, A. Khan, and D. Yadav, 2023, "Analysis of learning internal drive of private university students based on k-means clustering algorithm," *OITS International Conference on Information Technology*, pp. 703-707.

[5]. C. Han, S. Zhou, C. Xu, and H. Qian, 2024, "Visualized node detection-driven shipping network construction and ship trajectory clustering," *Sixth International Conference on Next Generation Data-driven Networks*, pp. 328-333.

[6]. Z. Chu, B. Liu, E. Guo, and M. Xu, 2024, "Construction of sightseeing vehicle driving cycles using autoencoder-based dimensionality reduction and optimized k-clustering," *8th International Conference on Electrical, Mechanical and Computer Engineering*, pp. 1059-1062.

[7]. M. Abdullah Eissa, and P. Chen, 2023, "Personalized electric vehicle range prediction based on self-supervised driving pattern clustering," *IEEE International Automated Vehicle Validation Conference*, pp. 1-6.

[8]. A. Gurusamy, A. Bokdia, H. Kumar, A. Radhika, B. Ashok, and C. Gunavathi, 2024, "Expediency analysis of clustering algorithms for electric two-wheeler driving cycle development under Indian smart city driving conditions," *IEEE Access*, 12, pp. 180279-180300.

[9]. X. Qi, Q. Hou, J. Gao, A. Chen, and L. Chen, 2023, "Data-driven cluster method for photovoltaic power stations," *IEEE 7th Conference on Energy Internet and Energy System Integration*, pp. 4409-4413.

[10]. S. Vithalani, and V. Shukla, 2023, "Data-driven fault detection in power systems: Clustering analysis for lightning and switching overvoltage," *IEEE Technology and Engineering Management Conference -*

Asia Pacific, pp. 1-7.

[11]. C. S. Mbacke Babou, Y. Owada, M. Inoue, K. Takizawa, and T. Kuri, 2024, “AI-driven automation for optimal edge cluster network management,” *IEEE INFOCOM 2024 - IEEE Conference on Computer Communications Workshops*, pp. 1-6.

[12]. H. Qi, R. Sun, P. Wang, J. Cao, Y. Zhu, and C. Li, 2024, “An adaptive clustering-based partitioning method of large-scale power grid for data-driven dynamic security assessment,” *3rd International Conference on Power Systems and Electrical Technology*, pp. 588-593.

[13]. S. Jeon, and S. Bae, 2023, “Data-driven electrolyzer modeling: Adaptive model considering operating conditions using k-means clustering,” *11th International Conference on Power Electronics and ECCE Asia*, pp. 2455-2460.

[14]. S. Shi, X. Han, T. Li, H. Zhang, and H. Yan, 2024, “A data-driven topology identification method for low-voltage distribution network,” *3rd International Conference on Power Systems and Electrical Technology*, pp. 522-526.

[15]. X. Chen, G. Sun, T. Wu, L. Liu, H. Yu, and M. Guizani, 2022, “RANCE: A randomly centralized and on-demand clustering protocol for mobile ad hoc networks,” *IEEE Internet of Things Journal*, 9(23), pp. 23639-23658.

[16]. R.S. Bali, N. Kumar, and J.J. Rodrigues, 2014, “Clustering in vehicular ad hoc networks: taxonomy, challenges, and solutions,” V*ehicular Communications*, 1(3), pp. 134-52.

[17]. E. Sakhaee, and A. Jamalipour, 2008, “Stable clustering and communications in pseudolinear highly mobile ad hoc networks,” *IEEE Transactions on Vehicular Technology*. 57(6), pp. 3769-3777.

[18]. S. Oubabas, R. Aoudjit, J.J. Rodrigues, S. Talbi, 2018, “Secure and stable vehicular ad hoc network clustering algorithm based on hybrid mobility similarities and trust management scheme,” *Vehicular Communications*. 13, pp. 128-138.