IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 July 2023, accepted 6 September 2023, date of publication 15 September 2023, date of current version 16 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3315852

SURVEY

Machine Learning-Based Road Safety Prediction Strategies for Internet of Vehicles (IoV) Enabled Vehicles: A Systematic Literature Review

K. RAVEENDRA REDDY AND A. MURALIDHAR
Vellore Institute of Technology, Chennai, Tamil Nadu 600127, India
Corresponding author: A. Muralidhar (muralidhar.a@vit.ac.in)

ABSTRACT This systematic literature review aims to investigate the current state-of-the-art in machine learning (ML) based road traffic analysis, hindrance estimation, and predicting vehicle safety measures for the Internet of Vehicles (IoV). Specifically, we focus on the verification of the scope and need of federated learning in this field. Federated learning is a decentralized ML technique that allows multiple edge devices to collaboratively train a shared model while keeping the data locally. We searched various academic databases and selected peer-reviewed publications related to the topic. The review highlights the existing challenges and limitations of the traditional centralized ML approaches and presents the advantages and potential benefits of federated learning in road traffic analysis and vehicle safety. Furthermore, we analyzed the current state-of-the-art research in federated learning for road traffic analysis and identified the research gaps and future research directions. The findings of this review demonstrate the scope and need of federated learning in the field of road traffic analysis and vehicle safety, as well as the potential of federated learning to overcome the limitations of the centralized ML approaches.

INDEX TERMS Internet of Vehicles, driving behavior analysis, machine learning, collision prediction, traffic accidents, deep learning, risk assessment.

I. INTRODUCTION

The "Internet of Vehicles (IoV)" has gained increasing focus in recent years due to its potential to revolutionize the transportation industry. Machine learning (ML) has been identified as a promising technology that can be utilized in various aspects of the IoV, such as road traffic analysis, hindrance estimation, and predicting vehicle safety measures. However, traditional centralized ML approaches have limitations in handling the vast amounts of data generated by the IoV [1].

Federated learning has emerged as a promising decentralized ML technique that can overcome the limitations of centralized ML approaches and improve the accuracy and safety of road traffic analysis and vehicle safety measures for the IoV. Federated learning allows multiple edge devices to collaboratively train a shared model while keeping the data locally [2].

The proposed systematic literature review aims to investigate the current state-of-the-art in ML-based road traffic analysis, hindrance estimation, and predicting vehicle safety measures for the IoV, and to verify the scope and need of federated learning in this field. The review will conduct a comprehensive search of academic databases and select peer-reviewed publications related to the topic. The advantages and potential benefits of federated learning in road traffic analysis and vehicle safety, as well as the existing challenges and limitations of traditional centralized ML approaches will be analyzed. Additionally, research gaps and future research directions in the field will be identified [1].

This systematic literature review aims to contribute to the development of more accurate and efficient ML-based solutions for road traffic analysis, hindrance estimation, and vehicle safety measures for the IoV. It is expected that the

The associate editor coordinating the review of this manuscript and approving it for publication was Nafees Mansoor.

findings of this review will facilitate the adoption of federated learning in the IoV and improve the accuracy and safety of road traffic analysis and vehicle safety measures [3].

II. LITERATURE REVIEW

A review of the literature is an important aspect of research that offers a thorough overview of the body of already known data as well as gaps in a certain topic. In this section, we present a systematic literature review on "Machine Learning (ML)" based Road Traffic, Vehicle Hindrances Estimation, and Predicting Vehicle Safety Measures for the Internet of Vehicles (IoV).

The review is structured into three subsections: review strategy, research questions, and review. The review strategy outlines the methodology used to conduct the literature search and select relevant articles. The research questions provide a clear focus for the review and serve as a guide to the reader. Finally, the review section presents a detailed analysis of the selected articles, highlighting key findings, and identifying research gaps.

Our goal in conducting this review is to provide a comprehensive overview of the current state-of-the-art in ML-based road traffic analysis, hindrance estimation, and predicting vehicle safety measures for the IoV. Additionally, we aim to identify research gaps and potential directions for future research.

We hope that this literature review will serve as a valuable resource for researchers, practitioners, as well as policymakers interested in the application of ML in the transportation industry.

A. REVIEW STRATEGY

A possible review strategy for a systematic literature review on "Machine Learning based Road Traffic, Vehicle Hindrances Estimation and Predicting Vehicle Safety Measures for Internet of Vehicles" that verifies the scope and need of federated learning could be as follows:

1. Define the research questions: Define the research questions based on the research objectives and the gaps identified in the literature. For example:

- What is the current state-of-the-art in ML-based road traffic analysis, hindrance estimation, and predicting vehicle safety measures for the IoV?
- What are the advantages and potential benefits of federated learning in road traffic analysis and vehicle safety for the IoV?
- What are the existing challenges and limitations of traditional centralized ML approaches for road traffic analysis and vehicle safety for the IoV?
- What are the research gaps and future research directions in ML-based solutions for road traffic analysis, hindrance estimation, and vehicle safety measures for the IoV?

2. Define the inclusion and exclusion criteria: Define the inclusion and exclusion criteria for the literature search based on the research questions. For example, the inclusion criteria may include:

- Studies presented at conferences or printed in peer-reviewed publications
- Studies on ML-based road traffic analysis, hindrance estimation, and predicting vehicle safety measures for the IoV
- Studies that use federated learning as a technique for handling the vast amounts of data generated by the IoV

The exclusion criteria may include:

- Studies not published in English
- Studies not focused on ML-based road traffic analysis, hindrance estimation, and predicting vehicle safety measures for the IoV
- Studies that do not use federated learning as a technique for handling the vast amounts of data generated by the IoV

3. Conduct the literature search: Conduct the literature search using relevant academic databases such as "Google Scholar", "ACM Digital Library", as well as "IEEE Xplore". Use a combination of keywords related to the research questions and the inclusion criteria, such as "machine learning", "road traffic analysis", "vehicle hindrances estimation", "vehicle safety measures", "Internet of Vehicles", and "federated learning".
4. Screen the studies: Determine the inclusion as well as exclusion criteria before screening the studies. First, screen the studies based on the title and abstract, and then screen the full text of the selected studies.
5. Extract the data: Extract the relevant data from the selected studies, such as the research objectives, research methods, datasets used, ML techniques used, and results.
6. Analyze the data: Analyze the extracted data using a systematic approach such as a meta-analysis or a narrative synthesis. Compare and contrast the results of the selected studies and identify the trends, patterns, and gaps in the literature.
7. Write the literature review: Write the literature review based on the analysis of the selected studies. Organize the literature review into subsections such as "ML-based road traffic analysis", "Vehicle hindrances estimation", "Predicting vehicle safety measures", and "Federated learning". Summarize the key findings of each subsection and provide a critical evaluation of the strengths and limitations of the existing literature.
8. Verify the scope and need of federated learning: Based on the literature review, verify the scope and need of federated learning in ML-based road traffic analysis, hindrance estimation, and predicting vehicle safety measures for the IoV. Evaluate the advantages and potential benefits of federated learning compared to traditional centralized ML approaches and identify the

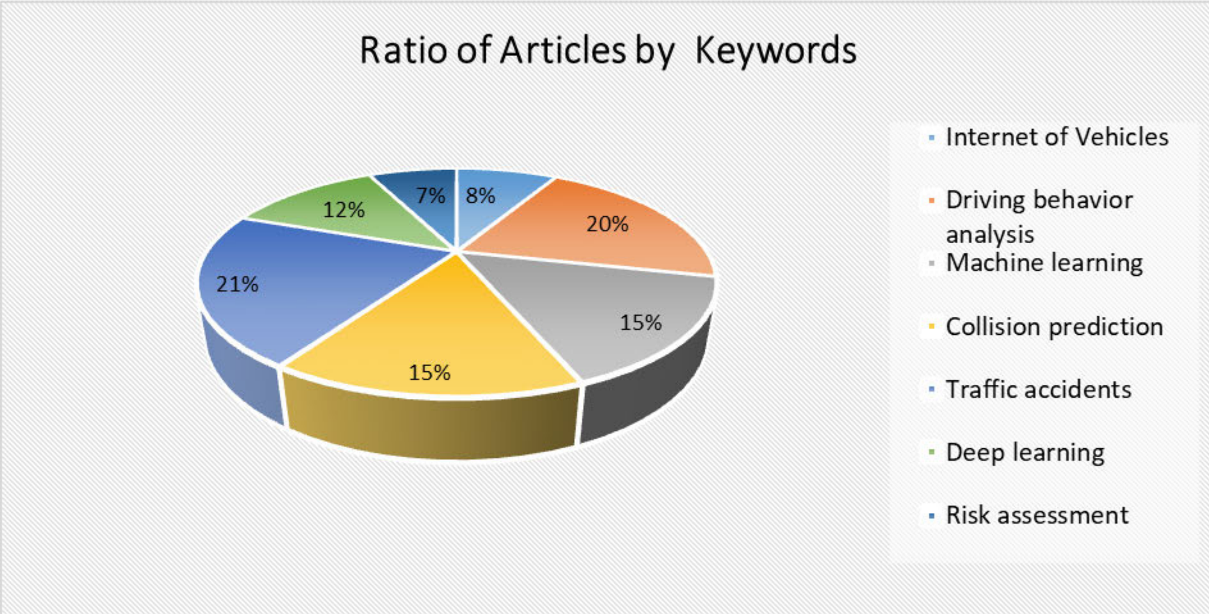

FIGURE 1. Distribution of articles according to keywords.

research gaps and future research directions in this field.

9. Conclude the literature review section by summarizing the key findings and highlighting the contributions of the study. Provide recommendations for future research and practical implications for the development of more accurate and efficient ML-based solutions for road traffic analysis, hindrance estimation, and vehicle safety measures for the IoV.

B. RESEARCH QUESTIONS

1. What are the challenges of using centralized machine learning approaches for road traffic and vehicle hindrances estimation and predicting vehicle safety measures for the Internet of Vehicles?
2. How can federated learning be used to address the challenges of centralized machine learning approaches in the context of the Internet of Vehicles?
3. What are the limitations and potential benefits of using federated learning for road traffic as well as vehicle safety in the context of the Internet of Vehicles?
4. What are the existing research studies that have used federated learning for road traffic and vehicle safety, and what are their findings and limitations?
5. What are the future research directions and opportunities for using federated learning in the context of vehicle safety and road traffic for the Internet of Vehicles?
6. What are the ethical and privacy concerns associated with federated learning in the context of the IoV, and how can they be addressed?
7. How can federated learning be integrated with other emerging technologies, such as blockchain, to enhance vehicle safety and road traffic in the context of the Internet of Vehicles?
8. How can federated learning be used to optimize traffic control and urban congestion reduction?

These research questions are essential in justifying the need and scope of the literature review towards the integration of federated learning with machine learning-based road traffic and vehicle hindrances estimation and predicting vehicle safety measures for the Internet of Vehicles. These research questions will provide a basis for identifying the gaps and limitations of existing study in this area, as well as inform the future study directions and opportunities for enhancing road safety, optimizing traffic management, and improving vehicle performance in the context of the IoV.

Out of the 72 articles chosen, each was selected based on relevant keywords. The frequency of the keywords and the corresponding article count were analyzed. Figure 1 illustrates the proportion of articles associated with each keyword.

Additionally, articles were filtered according to their publication type (journal, conference, book chapter, and other contents). Figure 2 presents the distribution of the selected articles based on their type. The remaining 72 articles were then assessed by publisher, with no articles being discarded during this process. Figure 2 also provides a detailed breakdown of the articles by publisher.

In the final stage, articles were filtered based on their year of publication, resulting in the removal of two articles. The remaining 72 articles were further assessed using qualitative synthesis factors, leading to the exclusion of two more articles. Figure 3 displays the distribution of the articles by their year of publication.

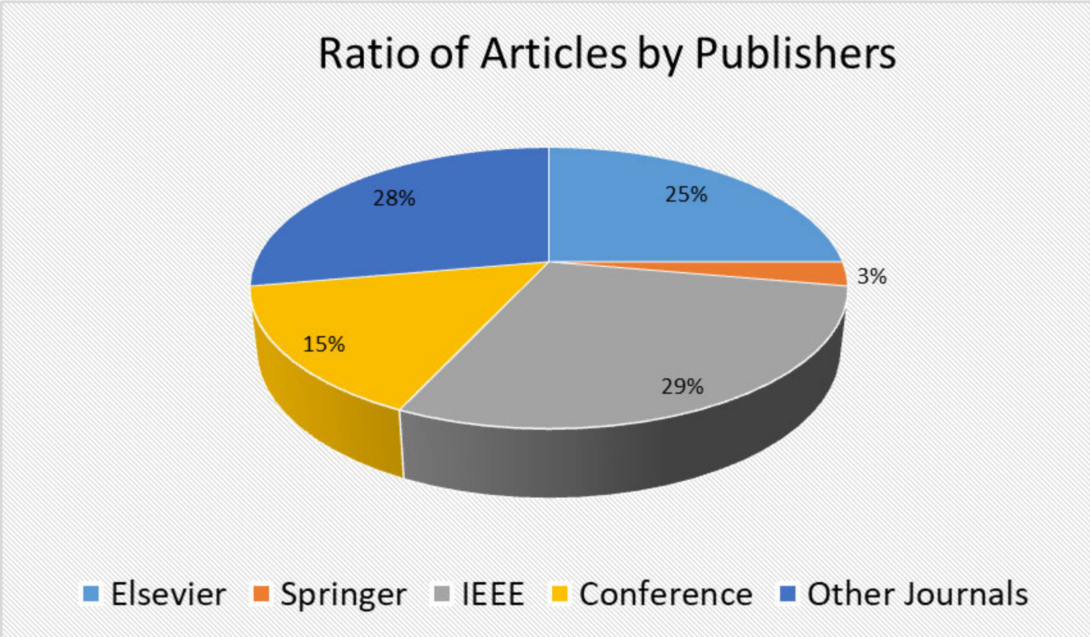

FIGURE 2. Distribution of articles by publisher.

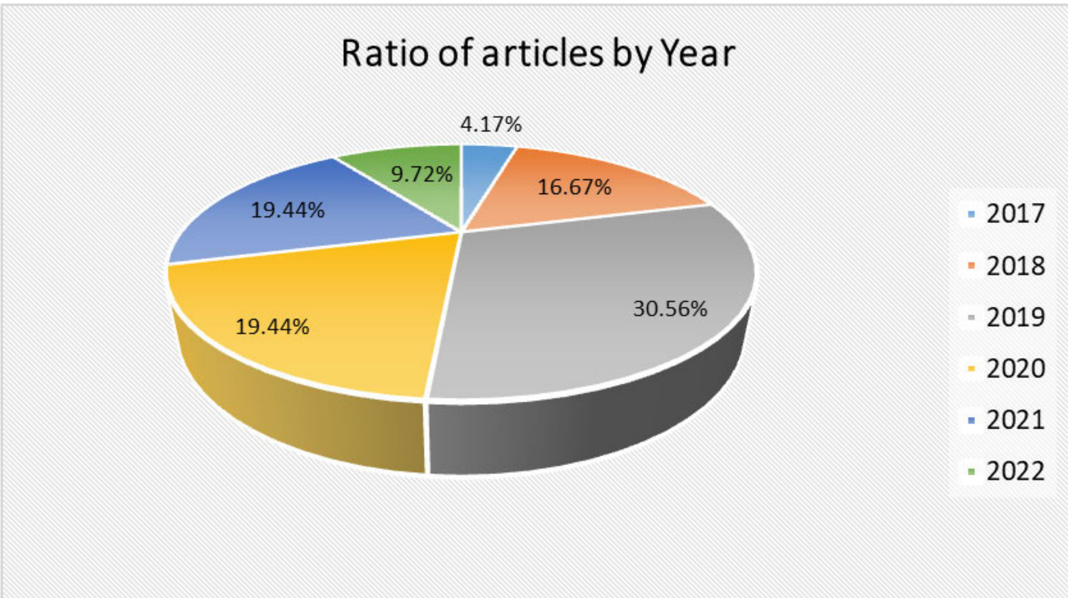

FIGURE 3. Fraction of articles by year of publication.

C. REVIEW

1) MACHINE LEARNING BASED DRIVER BEHAVIOR ASSESSMENT

Driver behavior analysis is an essential aspect of driving safety and the development of autonomous vehicles. This section summarizes and critiques recent articles that focus on various methods for driver behavior analysis Table 1.

This literature review highlights various machine learning (ML) applications in enhancing road safety. Xiong et al. [4] proposed a novel framework, CRTI, to predict road traffic accidents using ML and pattern recognition. Jahangiri et al. [5] developed an ML algorithm to detect hazardous driving actions by analysing basic safety message (BSM) data from connected vehicles. Saleh et al. [6] implemented stacked LSTM recurrent neural networks to classify driving behaviour through smartphone sensor data. Gwak et al. [7] employed ML algorithms to detect driver drowsiness using a combination of driving performance metrics, behavioural characteristics, and physiological responses. Huang et al. [8] presented an LSTM-NN model to replicate authentic traffic flow characteristics and asymmetric driving behaviour. Zhao et al. [9] investigated the travel behaviours of elderly drivers using probe vehicle (PV) data. Iranmanesh et al. [10] introduced an adaptive Forward Collision Warning (FCW) system that minimises the occurrence of false warnings and enhances the acceptability of warnings. The review shows that ML has the potential to enhance road safety by predicting accidents, detecting hazardous driving actions, classifying driving behaviour, detecting driver drowsiness, replicating authentic traffic flow characteristics, investigating travel behaviours of elderly drivers, and developing adaptive FCW systems.

Xie et al. [11] proposed a method for detecting driver distraction by using smartphone sensors. The data collected from the gyroscope, accelerometer, and magnetometer sensors of participants' smartphones was classified into distracted and non-distracted driving states using machine learning algorithms. The proposed methodology exhibited high accuracy, with an F1 score of 0.91, and was more precise than alternative methods. However, the sample size was limited, and the study did not account for extrinsic variables that may influence driver distraction.

Bejani et al. [12] developed the CADSE system, a driving style evaluation system that utilises smartphones. The system comprises three subsystems that calibrate smartphone sensors, classify manoeuvres, and evaluate driving styles. Ensemble learning is used to classify manoeuvres, and a rule-based fuzzy inference system is employed to identify hazardous and secure manoeuvres. The system achieved precision, accuracy, F-measure, and recall values greater than 94%, 92%, 92%, and 93%, respectively. The proposed system can be utilised by both insurance and enforcement subsystems and enables insurance companies to establish driver risk-based rates.

Ramanishka et al. [13] presented the Honda Research Institute Driving Dataset (HDD), which provides a practical and demanding opportunity for investigating driver behaviour and devising algorithms for driver assistance and self-driving capabilities. The dataset aims to comprehend the interaction between human driving and traffic scenes. The authors introduced baseline algorithms for detecting driver behaviour and provided a detailed comparison of the HDD dataset with other driving datasets.

Li et al. [14] proposed a system that utilises a combination of dashcam and inertial sensors to identify atypical driving occurrences. The system's ability to identify anomalous situations enhances the automated vehicle testing process. The system employs an autoencoder-based method to identify unfamiliar views, and the atypical driving occurrences facilitate the process of re-education and enhancement of the autonomous driving model. The proposed system has the potential to expedite the advancement of resilient autonomous driving systems.

Fugiglando et al. [15] suggested a method for examining and categorising driver conduct using a particular subset of signals from the CAN bus. The application of unsupervised learning was employed to partition drivers into distinct groups, and a validation approach was implemented to assess the resilience of the clustering outcomes across diverse experimental scenarios. The present research presents a method for classifying driver behaviour in uncontrolled environments in near-real-time.

Bouhoute et al. [16] examined numerical abstractions in the context of analysing cyber-physical systems. Galois connections are utilised in the construction of continuous abstractions of system sounds, and the authors presented both static and dynamic numerical intervals for analysing driving data. The proposal outlined a methodology for the analysis of cyber-physical systems through the use of a continuous sound abstraction technique. Graph matching is utilised in similarity analysis for classification and pattern recognition purposes.

Sheykhfard et al. [17] conducted a naturalistic driving study to investigate driver-pedestrian interactions in Babol, Iran. The study analyzed the behavior of 66 individuals in 216 vehicle-pedestrian interactions on divided roads and 485 such interactions on undivided roads. The study found that the speed of the vehicle and the distance of the pedestrian primarily influenced vehicle-pedestrian interactions. The study also found that pedestrians' attentiveness to traffic patterns before crossing the road and group crossings positively impacted driver performance.

Arumugam et al. [18] examined the use of ML in the analysis of driver behavior for motor insurance telematics. The authors reviewed various learning techniques, such as supervised, unsupervised, deep, and reinforcement learning, that are employed for the analysis of driver behavior. The authors suggest that the implementation of telematics has the potential to enhance insurance risk evaluation, pricing strategies, and customer engagement.

Papadimitriou et al. [19] investigated driving behavior associated with the use of smartphones. The study utilized a mobile application and a cohort of 100 drivers to assess driving behavior and exposure. The study found that the use of mobile phones while driving negatively impacted driving behavior. The authors developed predictive models using machine learning algorithms to identify instances of mobile phone usage during driving.

Zhang et al. [20] proposed a methodology for recognizing driving behaviors using sensor data from in-vehicle CAN-BUS through the application of deep learning (DL). The proposed methodology facilitated automated acquisition of driving behaviors and temporal feature modeling without requiring expertise in feature modeling. The approach was capable of capturing complex temporal driving behaviors that cannot be captured by traditional methods.

Zhang et al. [21] introduced a defensive alerting mechanism to identify and provide notification of hazardous automobiles. The mechanism employed a cloud server and vehicle computation to evaluate the driving performance of all vehicles worldwide. The proposed mechanism could detect vehicles that were being driven recklessly and subsequently notify drivers to mitigate the likelihood of accidents.

Wan et al. [22] investigated the asymmetric behavior of car-following during traffic oscillation. The study employed an unmanned aerial vehicle (UAV) to capture footage in the vicinity of a constriction on a highway located in Nanjing, China. The research revealed that motorists exhibit a range of distinct yet uniform driving behaviors, and the oscillation response of each driver is contingent upon their pre-existing traits.

Moukafih et al. [23] introduced a Long Short Term Memory Fully Convolutional Network (LTSM-FCN) for identifying instances of aggressive driving behavior. The proposed methodology achieved high accuracy when evaluated using a 5-minute window, surpassing the performance of other methods. The authors observed that individuals with lower inhibitory capacity exhibit a higher propensity for violating traffic laws and experiencing vehicular accidents.

Chan et al. [24] conducted a comparative analysis of various sensing schemes and detection algorithms proposed by different authors, which utilize smartphones to identify instances of drowsiness and atypical driving actions. The study emphasizes the significance of active systems such as ADAS or IDAS in alerting drivers of potential hazards. However, the authors suggest that there are still novel methodologies in the literature that warrant further exploration.

Pawar et al. [25] investigated the effects of time constraints on drivers' braking behavior and accident risk. The study found that various factors, including gender, driving profession, age, driving history, approach speed, as well as driving condition, had a significant impact on braking behavior. The study sheds light on the impact of time constraints on the braking patterns of drivers and their susceptibility to accidents, which can inform the development of interventions aimed at mitigating the adverse effect of time constraints on drivers' behavior, thereby enhancing road safety.

Petraki et al. [26] analyzed the driving behavior of drivers at road segments and junctions by utilizing high-resolution smartphone sensor data. The study sheds light on the impact of traffic characteristics on driver behavior and suggests that the results may be utilized to formulate strategies aimed at enhancing road safety and mitigating aggressive driving behavior.

Abou Elassad et al. [27] utilized machine learning (ML) models to analyze driver behavior (DB). The authors present a theoretical framework that conceptualizes the phenomenon of driver behavior (DB) by considering various dimensions inherent in the "Driver-Vehicle-Environment (DVE)" system. The authors conducted a systematic literature review (SLR) on the concept of DB investigation and provided an overview of the literature on ML. The present review demonstrates that machine learning techniques have the ability to evaluate database performance and exhibit superior performance compared to alternative models.

Singhet et al. [28] provided a summary of naturalistic driving studies pertaining to driver behavior and road safety. The authors have presented key findings from prior naturalistic driving studies, including the fact that driver behavior is the primary cause of most road accidents. Additionally, the authors have contributed significant insights pertaining to naturalistic driving studies, which are instrumental in analyzing driver behavior and enhancing road safety.

Azadani et al. [29] conducted a comprehensive literature review on Driver Behavior Analysis (DBA), examining its challenges and future trends. The present proposal aims to facilitate researchers in comprehending the research direction and challenges pertaining to the field of DBA, which enhances road safety and mitigates occurrences of accidents.

Aboulola et al. [30] proposed MODAL-IoCV as a means of conducting driver behavior analysis through the use of DL techniques. The proposed approach endeavors to offer driver support and itinerary preparation with the objective of mitigating accidents and enhancing traffic safety. The proposed approach employs a multimodal methodology that integrates vehicle motion and lane change prediction, feature extraction, recommendation, and route planning to analyze drivers' behavior.

Uprety et al. [31] utilized Federated Machine Learning to develop an IoV misbehavior detection system that ensures privacy preservation. The detection of falsified data in Vehicular Ad hoc Networks (VANET) is achieved by the system while ensuring the preservation of user privacy. The present proposal aims to create a machine learning model that can identify instances of VANET position falsification attacks by utilizing the basic safety message (BSM) dataset of vehicles.

Brahim et al. [32] explored the use of smartphone sensors to categorize driver behavior in an effective and affordable way. They found that smartphone sensors, such as accelerometers, gyroscopes, and GPS, were more effective in categorizing behavior than CAN-bus sensors. The study compared various ML algorithms for time series classification to identify the optimal approach. Inter-axial information from multiple sensors was integrated to improve classification accuracy. The authors highlighted the importance of driver behavior profiling in insurance and fleet management.

In summary, the articles reviewed in this section present various methods for driver behaviour analysis. The proposed methodologies exhibit high accuracy, precision, and recall values, and have the potential to enhance driver safety in practical settings. However, certain limitations such as the limited sample size, the specific driving tasks employed, and the lack of consideration of extrinsic variables were observed. The datasets provided by the authors serve as valuable resources for investigating driver behavior and developing algorithms for driver assistance and self-driving capabilities.

TABLE 1. Summary of the articles reviewed towards driving and driver behavior.

Author	Methodology	Application	Key Findings
Xiong et al. [4]	ML and pattern recognition (CRTI framework)	Predicting road traffic accidents	ML has the potential to predict road accidents
Jahangiri et al. [5]	ML algorithm analysing basic safety message (BSM) data from connected vehicles	Detecting hazardous driving actions	ML can detect hazardous driving actions
Saleh et al. [6]	Stacked LSTM recurrent neural networks using smartphone sensor data	Classifying driving behavior	ML can classify driving behavior through smartphone sensor data
Gwak et al. [7]	ML algorithms using driving performance metrics, behavioral characteristics, and physiological responses	Detecting driver drowsiness	ML can detect driver drowsiness using a combination of metrics
Huang et al. [8]	LSTM-NN model	Replicating authentic traffic flow characteristics and asymmetric driving behavior	ML can replicate authentic traffic flow characteristics
Zhao et al. [9]	Probe vehicle (PV) data analysis	Investigating travel behaviors of elderly drivers	ML can investigate travel behaviors of elderly drivers
Iranmanesh et al. [10]	Adaptive Forward Collision Warning (FCW) system	Enhancing FCW system performance	Adaptive FCW systems can minimize false warnings and enhance warning acceptability
Xie et al. [11]	Smartphone sensors classification using ML algorithms	Detecting driver distraction	ML can accurately detect driver distraction using smartphone sensors
Bejani et al. [12]	Ensemble learning and rule-based fuzzy inference system using smartphone sensors	Evaluating driving styles	Smartphone-based evaluation systems can be used to evaluate driving styles for insurance and enforcement purposes
Ramanishka et al. [13]	Honda Research Institute Driving Dataset (HDD) and baseline algorithms	Investigating driver behavior for driver assistance and self-driving capabilities	The HDD dataset provides a practical and demanding opportunity for investigating driver behavior
Li et al. [14]	Dashcam and inertial sensor data analysis	Identifying atypical driving occurrences for autonomous driving testing	ML can identify atypical driving occurrences for the advancement of autonomous driving systems
Fugiglando et al. [15]	Unsupervised learning using CAN bus signals	Classifying driver behavior in uncontrolled environments	Unsupervised learning can be used to classify driver behavior in uncontrolled environments
Bouhoute et al. [16]	Continuous sound abstraction technique using Galois connections and graph matching	Analyzing cyber-physical systems in driving data	Continuous sound abstraction techniques can be used to analyze cyber-physical systems in driving data
Sheykhfard et al. [17]	Naturalistic driving study	Investigating driver-pedestrian interactions	Speed and distance primarily influence vehicle-pedestrian interactions
Arumugam et al. [18]	Review of ML techniques for motor insurance telematics	Analyzing driver behavior for motor insurance telematics	Telematics can enhance insurance risk evaluation, pricing strategies, and customer engagement
Papadimitriou et al. [19]	Mobile application and ML algorithms	Identifying instances of mobile phone usage during driving	Mobile phone usage while driving negatively impacts driving behavior
Zhang et al. [20]	DL methodology using in-vehicle CAN-BUS sensor data	Recognizing complex driving behaviors	DL can capture complex temporal driving behaviors
Zhang et al. [21]	Defensive alerting mechanism using cloud server and vehicle computation	Notifying drivers of hazardous automobiles	Defensive alerting mechanisms can mitigate the likelihood of accidents
Wan et al. [22]	Employed an unmanned aerial vehicle (UAV) to capture footage	Investigated the asymmetric behavior of car-following during traffic oscillation	Motorists exhibit a range of distinct yet uniform driving behaviors, and the oscillation response of each driver is contingent upon their pre-existing traits.
Moukafih et al. [23]	Introduced a Long Short Term Memory Fully Convolutional Network	Identifying instances of aggressive driving behavior	Individuals with lower inhibitory capacity exhibit a higher propensity for violating traffic laws and experiencing vehicular accidents.
Chan et al. [24]	Conducted a comparative analysis of various sensing schemes	Identifying instances of drowsiness and atypical driving actions	Emphasizes the significance of active systems such as ADAS or IDAS in alerting drivers of potential hazards.
Pawar et al. [25]	Investigated the effects of time constraints	Drivers' braking behavior and accident risk	Various factors, including gender, driving profession, age, driving history, approach speed, as well as driving

TABLE 1. ***(Continued.)*** **Summary of the articles reviewed towards driving and driver behavior.**

			condition, had a significant impact on braking behavior.
Petraki et al. [26]	Analyzed the driving behavior of drivers	Driving behavior of drivers at road segments and junctions	The impact of traffic characteristics on driver behavior and suggests that the results may be utilized to formulate strategies aimed at enhancing road safety and mitigating aggressive driving behavior.
Abou Elassad et al. [27]	Utilized machine learning (ML) models	Analyze driver behavior (DB)	Machine learning techniques have the ability to evaluate database performance and exhibit superior performance compared to alternative models.
Singhet et al. [28]	Provided a summary of naturalistic driving studies	Driver behavior and road safety	Driver behavior is the primary cause of most road accidents.
Azadani et al. [29]	Conducted a comprehensive literature review	Driver Behavior Analysis (DBA) examining its challenges and future trends	Facilitate researchers in comprehending the research direction and challenges pertaining to the field of DBA, which enhances road safety and mitigates occurrences of accidents.
Aboulola et al. [30]	Proposed MODAL-IoCV as a means of conducting driver behavior analysis	Driver support and itinerary preparation with the objective of mitigating accidents and enhancing traffic safety	The proposed approach employs a multimodal methodology that integrates vehicle motion and lane change prediction, feature extraction, recommendation, and route planning to analyze drivers' behavior.
Uprety et al. [31]	Federated Machine Learning	Misbehavior detection in Vehicular Ad hoc Networks (VANET)	The proposed system detects falsified data in VANET while preserving user privacy, using basic safety message (BSM) dataset.
Brahim et al. [32]	Smartphone sensors and ML algorithms	Driver behavior profiling in insurance and fleet management	Smartphone sensors were found more effective than CAN-bus sensors in categorizing driver behavior, and integrating inter-axial information from multiple sensors improved classification accuracy.

2) MACHINE LEARNING BASED COLLISION SCOPE PREDICTION

Advancements in the Internet of Vehicles (IoV) have resulted in an increased focus on traffic safety research. This section examines recent studies on IoV-based traffic safety Table 2.

Road traffic injuries and fatalities are a significant public health concern worldwide, and governments need to take responsibility for ensuring road safety and providing resources for effective interventions to mitigate accidents, according to Hyder et al. [33]. Zong et al. [34] developed a severity causation network that uses information entropy and Bayesian networks to forecast severity indexes, aiding managers in scrutinising the severity of traffic accidents to enhance safety measures and minimise casualties and property damages. Yin et al. [35] proposed a hybrid model and algorithm that utilises ARIMA and GPSOWNN for short-term traffic flow prediction, showing superior performance in predictive accuracy. Chang et al. [36] presented a real-time object detection and dynamic prediction methodology for predicting vehicle positions in IoV frameworks. Zheng et al. [37] introduced the TASP-CNN method, a novel approach to predicting traffic accident severity that enhances predictive accuracy through the incorporation of combination relationships. The TrafficPredict model, developed by Ma et al. [38], utilises LSTM for forecasting traffic-agent trajectories in diverse traffic environments, exhibiting superior predictive accuracy compared to other contemporary techniques. Lv et al. [39] proposed the use of high-resolution traffic trajectories obtained from roadside LiDAR sensors to identify vehicle-pedestrian conflicts through the use of SDP spatial distribution. Peng et al. [40] discussed the evaluation of the significance of condition attributes within equivalence classes, generating decision rules based on equivalence classes to improve classification results. Katrakazas et al. [41] proposed an integrated method for assessing collision risk in autonomous vehicles, using traffic simulation, collision risk assessment, and machine learning techniques to forecast collisions on road segments. Gao et al. [42] developed a predictive model for insurance claims based on driver behaviour, while Ayuso et al. [43] presented a count data regression model for frequency, incorporating telematics data to accommodate driver behaviour and other variables that could potentially impact insurance expenses. The utilisation of telematics technology has the potential to enhance the calculation of insurance premiums by traffic regulatory bodies.

Several studies have been conducted to enhance traffic safety by predicting traffic accidents and mitigating their effects. Zhao et al. [44] proposed a traffic accident prediction method based on Convolutional Neural Networks (CNN). They employed deep learning techniques to extract multi-dimensional features autonomously from Vehicle Ad-hoc Networks (VANET) data. The CNN-based model was trained and tested on a dataset of simulated vehicle accidents, and it exhibited higher prediction accuracy and lower loss compared to other machine learning algorithms. The model's theoretical foundation is expected to facilitate the development of vehicle safety assisted driving, anti-collision guidance, and intelligent vehicle path planning optimization.

Lyu et al. [45] proposed a collision warning model for Advanced Driver Assistance Systems (ADAS) in a Vehicle-to-Vehicle (V2V) communication setting. The model incorporated lane-change behavior and trajectory prediction

TABLE 2. Summary of the articles reviewed towards Machine Learning based Collision Scope Prediction.

Author	Methodology	Application	Key Findings
Hyder et al. [33]	Review	Road safety	Governments need to ensure road safety and provide resources for effective interventions to mitigate accidents.
Zong, Fang, et al. [34]	Information entropy, Bayesian networks	Traffic accident severity forecasting	Severity causation network aids managers in scrutinising the severity of traffic accidents to enhance safety measures and minimise casualties and property damages.
YIN, Lisheng et al. [35]	ARIMA, GPSOWNN	Traffic flow prediction	Hybrid model and algorithm exhibit superior performance in predictive accuracy.
Chang, Che-Cheng et al. [36]	Real-time object detection, dynamic prediction methodology	Vehicular position prediction	Real-time object detection and dynamic prediction methodology for predicting vehicle positions in IoV frameworks.
Zheng, Ming, et al. [37]	TASP-CNN method	Traffic accident severity prediction	TASP-CNN method enhances predictive accuracy through the incorporation of combination relationships.
Ma, Yuexin et al. [38]	LSTM	Traffic-agent trajectory forecasting	TrafficPredict model exhibits superior predictive accuracy compared to other contemporary techniques.
Lv, Bin et al. [39]	SDP spatial distribution	Vehicle-pedestrian conflict identification	High-resolution traffic trajectories obtained from roadside LiDAR sensors used to identify vehicle-pedestrian conflicts through the use of SDP spatial distribution.
Peng, Liqun, et al. [40]	Variable precision rough set (VPRS)	Emergency driving safety evaluation	Classification technique based on VPRS extracts a condensed core subset from the driving dataset, encompassing the most significant attributes for evaluating driving safety.
Katrakazas, Christos et al. [41]	Traffic simulation, collision risk assessment, machine learning	Collision risk assessment in autonomous vehicles	Integrated method for assessing collision risk in autonomous vehicles using traffic simulation, collision risk assessment, and machine learning techniques to forecast collisions on road segments.
Gao, Guangyuan , et al. [42]	Predictive model for insurance claims	Driver behaviour and insurance expenses	Predictive model for insurance claims based on driver behaviour.
Ayuso, Mercedes et al. [43]	Count data regression model	Driver behaviour and insurance expenses	Count data regression model incorporating telematics data to accommodate driver behaviour and other variables that could potentially impact insurance expenses.
Zhao et al. [44]	CNN	Traffic accident prediction	CNN-based model exhibits higher prediction accuracy and lower loss compared to other machine learning algorithms.
Lyu et al. [45]	Collision warning model	Advanced Driver Assistance Systems (ADAS)	Collision warning model for ADAS in a Vehicle-to-Vehicle (V2V) communication setting outperforms conventional models in terms of warning confusion matrix and warning time.
Zhao et al. [46]	Deep learning techniques	Risk of traffic accidents prediction	Algorithm utilises a CNN within an edge server to extract multi-dimensional features from a substantial amount of traffic data originating from the edge network of vehicles.
Peng et al. [47]	Variable precision rough set (VPRS)	Emergency driving safety evaluation	Suggested approach exhibits a notable level of precision and consistency, rendering it suitable for deducing prompt emergency braking maneuvers.
Bhalla et al. [48]	Counterfactual analysis	Vehicle safety interventions	Evaluation of road safety interventions estimating the advantages of enhancing vehicle design.
Walugembe et al. [49]	Investigation	Road safety in Tanzania	Investigation suggests enhancing the road transportation infrastructure to guarantee the
Wang et al.	DL techniques, CNN model	Rear-end collision prediction	RCPM exhibits superior performance compared to other algorithms.
Pawar et al.	Driving simulator	Braking behaviour and likelihood of accidents	Various factors influence the braking behaviour of drivers.
Lin et al.	Crowdsourced data	Traffic flow prediction	Enhancing traffic condition data leads to improved precision.
Gutierrez-Osorio et al.	Review of contemporary ML techniques	Accident analysis	Bayesian networks are identified as the most precise models for accident analysis.
Yu et al.	DSTGCN model	Traffic accident forecasting	DSTGCN exhibits superior performance compared to traditional and contemporary techniques.
Fang et al.	SCAFNet model	Driver attention forecasting	SCAFNet demonstrates exceptional performance on three distinct datasets.
Lee et al.	Driving behaviour model	Collision avoidance for two-wheeled vehicles	Theoretical framework offers a predictive model for avoiding collisions by riders who engage in risky behaviour.
Li et al.	Risk assessment algorithm	Autonomous vehicle decision-making	The proposed approach addresses the needs of both drivers and passengers by implementing a collision avoidance strategy that takes into account driving style preferences.
Zhao et al.	Integration of online and offline data	Driver risk classification for UBI	Integrating both types of data enhances the accuracy of risk assessments, with offline consumer behavior variables being particularly significant.
Lin et al.	Predictive model using Bayes' theorem	Intersection accident risk assessment	Road dimensions, designated speed limits, and markings along the roadside are significant determinants of the likelihood of traffic

TABLE 2. ***(Continued.)*** **Summary of the articles reviewed towards Machine Learning based Collision Scope Prediction.**

			accidents at intersections.
Chang et al.	Machine learning architecture	Vehicular position prediction for collision avoidance	Their algorithm outperforms a prior study's linear algorithm, demonstrating the potential of machine learning techniques and vehicle dynamics principles to enhance future vehicular position prediction.
Lin et al. [60]	Predictive model using Bayes' theorem and traffic accident data	Risk assessment for intersection accidents	Road dimensions, designated speed limits, and roadside markings significantly impact likelihood of intersection accidents
Chang et al. [61]	Vehicle dynamics and machine learning architecture	Predicting vehicular positions for collision avoidance in IoV	Machine learning techniques and vehicle dynamics principles can enhance future vehicular position prediction
Yang et al. [62]	Transfer learning based on Internet of Vehicles data	Forecasting vehicular collisions	High degree of precision in forecasting vehicular collisions
Gutierrez-Osorio et al. [63]	Ensemble deep learning approach with "Gated Recurrent Units" and "CNN"	Predicting traffic collisions using social media and open data	Deep learning ensemble model exhibits superior performance compared to other methods
Morimoto et al. [64]	Comparative analysis of global traffic safety objectives and strategies	Enhancing road traffic safety	Comprehensive strategy needed encompassing infrastructure, technology, culture, and behaviour
Malawade et al. [65]	Spatio-temporal scene graph embedding using visual perception, GNN, and LSTM	Forecasting collisions involving autonomous vehicles	SG2VEC model demonstrates superior predictive accuracy and earlier detection capabilities compared to current leading approach
Suat-Rojas et al. [66]	Named entity recognition and geocoding	Extracting data related to traffic accidents from Twitter	Twitter has the potential to furnish information regarding traffic accidents, with commercial and industrial areas of the city being impacted by Twitter
Pan et al. [68]	Communication network architecture using LoRa technology and LSTM/ANN	Early warning system for pedestrian-vehicle collision in uncertain situations	Probability of confidence is a determining factor for issuing warnings for pedestrian-vehicle collisions

models, which were evaluated through both driving simulator and real-world vehicle tests. The proposed model outperformed conventional models in terms of warning confusion matrix and warning time, offering novel modeling concepts and theoretical backing for the optimization of ADAS cut-in functionality.

Zhao et al. [46] introduced an algorithm for predicting the risk of traffic accidents in vehicular edge networks using deep learning techniques. The algorithm utilizes a CNN within an edge server to extract multi-dimensional features from a substantial amount of traffic data originating from the edge network of vehicles. The model predicts vehicular collisions and notifies drivers to reduce speed and exercise caution. The simulations demonstrated that the algorithm under consideration exhibits reduced loss and improved prediction accuracy compared to other machine learning algorithms, and a comparative analysis was conducted to predict the severity of collision injuries using four commonly used ML techniques and two statistical methods.

Peng et al. [47] suggested utilizing actual driving data for the evaluation of emergency driving safety. They employed a classification technique based on variable precision rough set (VPRS) to extract a condensed core subset from the driving dataset, which encompasses the most significant attributes for evaluating driving safety. The suggested approach exhibits a notable level of precision and consistency, rendering it suitable for deducing prompt emergency braking maneuvers, thereby enhancing the efficacy of collision avoidance systems (CASs).

Bhalla et al. [48] conducted an evaluation of road safety interventions in the Latin America and Caribbean (LAC) region, estimating the advantages of enhancing vehicle design. The study conducted a counterfactual analysis to evaluate the potential decrease in fatalities and disability-adjusted life years lost in LAC nations if eight established vehicle safety technologies were more extensively accessible. The investigation projected the number of lives saved and disability-adjusted life years averted if these technologies were universally implemented in all vehicles.

Walugembe et al. [49] investigated the rates of fatalities resulting from road traffic accidents (RTAs) during pre-historical times in Ilala and two other municipalities located in the Dar es Salaam Region of Tanzania. The research suggests enhancing the road transportation infrastructure to guarantee the safety of road users through the implementation of current policies, reinforcement of law enforcement, and the introduction of immediate and effective penalties.

Wang et al. [50] introduced a Rear-end Collision Prediction Mechanism (RCPM) that utilizes DL techniques to enhance the prediction of rear-end collisions, which are a significant contributor to traffic accidents. The CNN model is trained through the utilization of both training and testing sets derived from the preprocessed dataset. The study reveals that RCPM exhibits superior performance compared to other algorithms in the prediction of rear-end collisions.

Pawar et al. [25] used a driving simulator to investigate the connection between driver braking patterns and the likelihood of accidents in response to heightened time constraints. Their research showed that various factors, including gender, driving occupation, approach speed, age, driving experience, and driving conditions, exert a significant influence on the braking behaviour of drivers. Lin et al. [51] utilised crowdsourced data to forecast post-accident traffic flow patterns and found that enhancing traffic condition data leads to an improvement in precision. Gutierrez-Osorio et al. [52] conducted a comprehensive review of contemporary machine learning techniques for analysing and predicting road accidents and concluded that Bayesian networks have been identified as the most precise models for accident analysis.

Yu et al. [53] proposed a "Deep Spatio-Temporal Graph Convolutional Network (DSTGCN)" for forecasting traffic accidents. Their model exhibits superior performance compared to both traditional and contemporary techniques when applied to real-world datasets. Fang et al. [54] introduced the "Semantic Context-Induced Attentive Fusion Network (SCAFNet)" as a means of forecasting driver attention in the context of "driving accident scenarios (DADA)". Their approach demonstrated exceptional performance on three distinct datasets. Lee et al. [55] formulated a driving behaviour model aimed at forecasting the hazardous collision avoidance of riders of two-wheeled vehicles. Their theoretical framework offers a predictive model for the avoidance of collisions by riders who engage in risky behaviour.

Li et al. [56] presented an algorithm for autonomous vehicle decision-making that is based on risk assessment. Their proposed approach for addressing the needs of both drivers and passengers involves the implementation of a collision avoidance strategy that takes into account driving style preferences. Finally, Zhao et al. [57] investigated driver behaviour in incidents involving collisions and near-misses between cars and cyclists. They found that the "Bus Rapid Transit (BRT)" parameter holds significant importance in preventing collisions. Li et al. [58] analysed road traffic accidents in Shenzhen, China, and suggested that China would benefit from the implementation of traffic management countermeasures and "advanced driver assistance systems (ADASs)".

Zhao et al. [59] investigated the impact of integrating both online and offline channel data on driver risk classification for usage-based insurance (UBI) products. By utilising driving behavior data obtained from On-Board Diagnostics (OBD) loggers and offline consumer behavior data obtained from 4S dealerships, LR, NN, RF, and SVM were used to categorise driver risk. The study found that integrating both types of data enhanced the accuracy of risk assessments, with offline consumer behavior variables being particularly significant. These findings have implications for the pricing and cost management of UBI in the insurance industry.

Lin et al. [60] developed a predictive model for the risk of intersection accidents using Bayes' theorem and traffic accident data. The study found that road dimensions, designated speed limits, and markings along the roadside were significant determinants of the likelihood of traffic accidents at intersections. The predictive model has the capability to assess potential hazards to mitigate occurrences of vehicular collisions and provide insights for the design of intersections and environmental enhancements.

Chang et al. [61] proposed a novel approach to predicting vehicular positions for collision avoidance in the IoV. The authors used vehicle dynamics and machine learning architecture to improve the accuracy and reliability of future vehicular position estimation, building upon the YOLOv4 real-time object detection project. The study found that their algorithm outperformed a prior study's linear algorithm, demonstrating the potential of machine learning techniques and vehicle dynamics principles to enhance future vehicular position prediction.

Yang et al. [62] presented a vehicle collision prediction model that utilises Internet of Vehicles data and is based on transfer learning. The model utilised historical performance data to predict future collision occurrences and their respective timings by analysing and simulating the driving mode of the vehicle. The model exhibited a high degree of precision in forecasting vehicular collisions, which enhances traffic safety and forthcoming driving mechanisms.

Gutierrez-Osorio et al. [63] proposed using social media and open data to predict traffic collisions, using an ensemble deep learning approach composed of "Gated Recurrent Units" and "CNN". The study found that their Deep Learning ensemble model exhibited superior performance compared to baseline algorithms and alternative deep learning methods.

Morimoto et al. [64] presented a theoretical framework for enhancing road traffic safety through a comparative analysis of global traffic safety objectives and strategies. The study highlighted the necessity of adopting a comprehensive strategy towards ensuring road safety, encompassing factors such as infrastructure, technology, culture, and behaviour.

Malawade et al. [65] created a model called SG2VEC that utilises spatio-temporal scene graph embedding to forecast collisions involving autonomous vehicles (AVs). The model employs visual scene perception in conjunction with GNN and LSTM layers to forecast collisions. The SG2VEC model demonstrated superior predictive accuracy and earlier detection capabilities compared to the current leading approach.

Suat-Rojas et al. [66] proposed a methodology for extracting data related to traffic accidents in Spanish from Twitter. The study found that Twitter has the potential to furnish information regarding traffic accidents, with commercial and industrial areas of the city being impacted by Twitter. The proposed method utilises named entity recognition and geocoding to address issues related to informal language and misspellings.

Lastly, Pan et al. [67] presented a communication network architecture and early warning system for Vehicle-to-Pedestrian (V2P) using LoRa technology. The system uses LSTM and ANN to predict the area of risk for pedestrian-vehicle collision in situations where pedestrian trajectories are uncertain. The probability of confidence is a determining factor for issuing warnings for pedestrian-vehicle collisions.

The studies presented in this section provide valuable insights into various aspects of road safety and accident prevention, the use of machine learning techniques to analyse and predict accidents, and the development of models for predicting hazardous collision avoidance. These studies can inform the development of effective strategies and technologies for improving road safety and reducing accidents.

3) MACHINE LEARNING BASED TRAFFIC CONGESTION PREDICTION

In recent years, several studies have focused on developing methodologies for predicting traffic collisions and congestion using machine learning techniques. Zhao et al. [59] integrated online and offline data to improve the accuracy of driver risk assessment in usage-based insurance products. Kothai et al. [68] proposed a hybrid BLSTME and CNN model for predicting traffic congestion levels that effectively addressed the issue of overfitting and attained a high level of accuracy. Onyeneke et al. [69] examined the factors contributing to and consequences of congestion in road traffic in metropolitan regions, and proposed several policy measures for mitigation. Du et al. [70] introduced a hybrid multimodal deep learning method for predicting short-term traffic flow that surpassed various baseline methods. Wei et al. [71] proposed an AutoEncoder Long Short-Term Memory approach for traffic flow prediction that demonstrated superior performance compared to prior techniques. Hébert et al. [72] utilised big data analytics to develop high-resolution accident forecasting techniques in Montreal, Canada, and identified several significant indicators of vehicle collisions. Moses et al. [73] presented a methodology for vehicular traffic prediction utilising machine learning techniques. Ranjan et al. [74] proposed a novel neural network architecture for predicting traffic congestion levels across an entire urban road network that exhibited superior computational efficacy and prediction performance compared to other DNNs.

While these studies offer promising solutions for addressing traffic collision and congestion problems, each has certain limitations that should be taken into account when applying these models in real-world scenarios. Further research is necessary to address these limitations and optimise the efficacy and feasibility of these proposed methodologies in intelligent transportation systems Table 3.

III. OBSERVATIONS

Centralized machine learning approaches for vehicle safety and road traffic in the context of the Internet of Vehicles (IoV) face significant challenges due to centralized data collection and storage. These approaches may not be scalable for large datasets, require significant computational resources, and may not be robust against cyber threats.

Federated learning, a distributed machine learning strategy, can address these challenges by enabling data to be processed and trained locally on individual devices, thereby addressing privacy and security concerns. Federated learning also improves scalability, reduces communication costs, and is more robust against adversarial attacks.

Existing research studies have demonstrated the effectiveness of federated learning for vehicle safety and road traffic in the context of the IoV, but limitations include the need for significant computational resources, heterogeneity of data, and limited data sharing among devices. Future research directions for federated learning in the context of vehicle safety and road traffic include investigating real-time decision-making, optimizing communication protocols and models, and addressing ethical and privacy concerns.

Federated learning can be integrated with other emerging technologies, such as blockchain, to enhance vehicle safety and road traffic in the context of the IoV. Blockchain technology can offer a decentralized platform for sharing and storing data among multiple entities, including vehicles and traffic management systems, and can ensure secure and private data sharing.

One potential application is the development of a blockchain-based platform for sharing data among vehicles to improve traffic flow and reduce congestion. Federated learning can be used to develop predictive models based on the shared data, which can be used to optimize traffic flow in real-time. Urban regions can benefit from improved traffic management and reduced congestion through federated learning.

Predictive models developed through federated learning can optimize traffic light timings, reroute vehicles, and suggest alternative modes of transportation. One potential application is the development of a federated learning-based platform for optimizing public transportation in urban areas by integrating data from multiple sources to develop predictive models for public transportation demand and suggest optimal routes and timings for buses and train.

A. RESEARCH SCOPE

Based on the research questions, some possible research objectives for a study on federated learning in the context of vehicle safety and road traffic for the Internet of Vehicles are:

1. To identify the ethical and privacy concerns associated with federated learning in the context of the IoV and develop strategies to address them.
2. To explore the potential of integrating federated learning with other emerging technologies, such as blockchain, to enhance road traffic and vehicle safety in the context of the Internet of Vehicles.

TABLE 3. The summary of the articles reviewed under Machine Learning based Traffic Congestion Prediction.

Author	Methodology	Application	Key Findings
Kothai et al. [68]	Hybrid BLSTME and CNN model for predicting traffic congestion levels	Traffic congestion prediction	The proposed model effectively addressed the issue of overfitting and attained a high level of accuracy.
Onyeneke et al. [69]	Examination of factors contributing to and consequences of congestion in road traffic in metropolitan regions	Policy measures for congestion mitigation	The study proposed several policy measures for mitigating congestion.
Du et al. [70]	Hybrid multimodal deep learning method for predicting short-term traffic flow	Short-term traffic flow prediction	The proposed method surpassed various baseline methods.
Wei et al. [71]	AutoEncoder Long Short-Term Memory approach for traffic flow prediction	Traffic flow prediction	The proposed approach demonstrated superior performance compared to prior techniques.
Hébert et al. [72]	High-resolution accident forecasting using big data analytics	Accident forecasting	Several significant indicators of vehicle collisions were identified.
Moses et al. [73]	Methodology for vehicular traffic prediction using machine learning techniques	Vehicular traffic prediction	The proposed methodology utilised machine learning techniques.
Ranjan et al. [74]	Neural network architecture for predicting traffic congestion levels across an entire urban road network	Urban road network congestion prediction	The proposed neural network architecture exhibited superior computational efficacy and prediction performance compared to other DNNs.

3. To investigate the feasibility and effectiveness of using federated learning to optimize traffic management and reduce congestion in urban areas.
4. To conduct a comprehensive literature review on existing research studies that have used federated learning for vehicle safety and road traffic and analyze their findings and limitations.
5. To identify the future research directions and opportunities for using federated learning in the context of vehicle safety and road traffic for the Internet of Vehicles.

B. RESEARCH OBJECTIVES

The research objectives recommended for further research are to develop machine learning strategies to estimate the hindrances caused by traffic, vehicle health, driver cognitive levels, and driving behavior to avoid consequences such as accidents and mortality caused by road transportation accidents. In order to achieve this, the following key challenges need to be addressed:

Initially, using blockchain-based Federated Learning to manage and access the data that accumulates from vehicles in an unstructured format.

Next, performing feature engineering under federated learning on the data that IoV collects on a continuous basis, which is used to predict road transportation scenarios, traffic and vehicle impediments, driver cognitive levels, and driving behavior. These predictions will be used to forecast the accidents, vehicle accident risk, and safe routes.

Finally, the challenges of safe route assessment, road congestion management, and intelligent traffic signaling will be attempted to address by using Federated Learning approaches.

The eventual goal of this research is to improve road safety and reduce accidents and mortality caused by road transportation accidents. By addressing these challenges through machine learning, the researchers aim to s before they occur.

IV. CONCLUSION

The use of machine learning algorithms for road traffic and vehicle safety in the context of the Internet of Vehicles (IoV) has gained significant attention in recent years. However, centralized data collection and storage present significant challenges, including privacy concerns and communication costs. Federated learning has emerged as a promising solution for addressing these limitations. This literature review provides an overview of recent advancements in the use of federated learning for road traffic and vehicle safety in the context of the IoV. The review identified potential benefits of using federated learning, including improved privacy and security, scalability, and robustness against adversarial attacks. However, there are also limitations, including the need for significant computational resources and the risk of model divergence or overfitting. The review highlights several existing research studies that have used federated learning for road traffic and vehicle safety, but further research is needed to validate its effectiveness and scalability in real-world scenarios. Overall, the paper emphasizes how federated learning has the ability to overcome the drawbacks of centralized machine learning techniques and offers recommendations for further study and improvement in this field.

REFERENCES

[1] M. N. Sadiku, M. Tembely, and S. M. Musa, "Internet of Vehicles: An introduction," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 8, no. 1, p. 11, 2018.

[2] O. Kaiwartya, A. H. Abdullah, Y. Cao, A. Altameem, M. Prasad, C.-T. Lin, and X. Liu, "Internet of Vehicles: Motivation, layered architecture, network model, challenges, and future aspects," *IEEE Access*, vol. 4, pp. 5356–5373, 2016.

[3] S. Abbasi, A. M. Rahmani, A. Balador, and A. Sahafi, "Internet of Vehicles: Architecture, services, and applications," *Int. J. Commun. Syst.*, vol. 34, no. 10, p. e4793, Jul. 2021.

[4] X. Xiong, L. Chen, and J. Liang, "A new framework of vehicle collision prediction by combining SVM and HMM," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 699–710, Mar. 2018.

[5] A. Jahangiri, V. J. Berardi, and S. G. Machiani, "Application of real field connected vehicle data for aggressive driving identification on horizontal curves," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 7, pp. 2316–2324, Jul. 2018.

[6] K. Saleh, M. Hossny, and S. Nahavandi, "Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6.

[7] J. Gwak, M. Shino, and A. Hirao, "Early detection of driver drowsiness utilizing machine learning based on physiological signals, behavioral measures, and driving performance," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1794–1800.

[8] X. Huang, J. Sun, and J. Sun, "A car-following model considering asymmetric driving behavior based on long short-term memory neural networks," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 346–362, Oct. 2018.

[9] Y. Zhao, T. Yamamoto, and T. Morikawa, "An analysis on older driver's driving behavior by GPS tracking data: Road selection, left/right turn, and driving speed," *J. Traffic Transp. Eng.*, vol. 5, no. 1, pp. 56–65, Feb. 2018.

[10] S. M. Iranmanesh, H. N. Mahjoub, H. Kazemi, and Y. P. Fallah, "An adaptive forward collision warning framework design based on driver distraction," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 3925–3934, Dec. 2018.

[11] J. Xie, A. R. Hilal, and D. Kulic, "Driver distraction recognition based on smartphone sensor data," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 801–806.

[12] M. M. Bejani and M. Ghatee, "A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data," *Transp. Res. C, Emerg. Technol.*, vol. 89, pp. 303–320, Apr. 2018.

[13] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko, "Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7699–7707.

[14] H. Li, H. Wang, L. Liu, and M. Gruteser, "Automatic unusual driving event identification for dependable self-driving," in *Proc. 16th ACM Conf. Embedded Networked Sensor Syst.*, Nov. 2018, pp. 15–27.

[15] U. Fugiglando, E. Massaro, P. Santi, S. Milardo, K. Abida, R. Stahlmann, F. Netter, and C. Ratti, "Driving behavior analysis through CAN bus data in an uncontrolled environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 737–748, Feb. 2019.

[16] A. Bouhoute, R. Oucheikh, K. Boubouh, and I. Berrada, "Advanced driving behavior analytics for an improved safety assessment and driver fingerprinting," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2171–2184, Jun. 2019.

[17] A. Sheykhfard and F. Haghighi, "Performance analysis of urban drivers encountering pedestrian," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 62, pp. 160–174, Apr. 2019.

[18] S. Arumugam and R. Bhargavi, "A survey on driving behavior analysis in usage based insurance using big data," *J. Big Data*, vol. 6, no. 1, pp. 1–21, Dec. 2019.

[19] E. Papadimitriou, A. Argyropoulou, D. I. Tselentis, and G. Yannis, "Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving," *Saf. Sci.*, vol. 119, pp. 91–97, Nov. 2019.

[20] J. Zhang, Z. Wu, F. Li, C. Xie, T. Ren, J. Chen, and L. Liu, "A deep learning framework for driving behavior identification on in-vehicle CAN-BUS sensor data," *Sensors*, vol. 19, no. 6, p. 1356, Mar. 2019.

[21] L. Zhang, L. Yan, Y. Fang, X. Fang, and X. Huang, "A machine learning-based defensive alerting system against reckless driving in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 12227–12238, Dec. 2019.

[22] Q. Wan, G. Peng, Z. Li, F. Inomata, Y. Zheng, and Q. Liu, "Using asymmetric theory to identify heterogeneous drivers' behavior characteristics through traffic oscillation," *IEEE Access*, vol. 7, pp. 106284–106294, 2019.

[23] Y. Moukafih, H. Hafidi, and M. Ghogho, "Aggressive driving detection using deep learning-based time series classification," in *Proc. IEEE Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Jul. 2019, pp. 1–5.

[24] T. K. Chan, C. S. Chin, H. Chen, and X. Zhong, "A comprehensive review of driver behavior analysis utilizing smartphones," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4444–4475, Oct. 2020.

[25] N. M. Pawar, R. K. Khanuja, P. Choudhary, and N. R. Velaga, "Modelling braking behaviour and accident probability of drivers under increasing time pressure conditions," *Accident Anal. Prevention*, vol. 136, Mar. 2020, Art. no. 105401.

[26] V. Petraki, A. Ziakopoulos, and G. Yannis, "Combined impact of road and traffic characteristic on driver behavior using smartphone sensor data," *Accident Anal. Prevention*, vol. 144, Sep. 2020, Art. no. 105657.

[27] Z. E. A. Elassad, H. Mousannif, H. Al Moatassime, and A. Karkouch, "The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103312.

[28] H. Singh and A. Kathuria, "Analyzing driver behavior under naturalistic driving conditions: A review," *Accident Anal. Prevention*, vol. 150, Feb. 2021, Art. no. 105908.

[29] M. N. Azadani and A. Boukerche, "Driving behavior analysis guidelines for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 6027–6045, Jul. 2022.

[30] O. Aboulola, M. Khayyat, B. Al-Harbi, M. S. A. Muthanna, A. Muthanna, H. Fasihuddin, and M. H. Alsulami, "Multimodal feature-assisted continuous driver behavior analysis and solving for edge-enabled Internet of Connected Vehicles using deep learning," *Appl. Sci.*, vol. 11, no. 21, p. 10462, Nov. 2021.

[31] A. Uprety, D. B. Rawat, and J. Li, "Privacy preserving misbehavior detection in IoV using federated machine learning," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–6.

[32] S. B. Brahim, H. Ghazzai, H. Besbes, and Y. Massoud, "A machine learning smartphone-based sensing for driver behavior classification," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2022, pp. 610–614.

[33] A. A. Hyder, "Measurement is not enough for global road safety: Implementation is key," *Lancet Public Health*, vol. 4, no. 1, pp. 12–13, Jan. 2019.

[34] F. Zong, X. Chen, J. Tang, P. Yu, and T. Wu, "Analyzing traffic crash severity with combination of information entropy and Bayesian network," *IEEE Access*, vol. 7, pp. 63288–63302, 2019.

[35] L. Yin, S. Tang, S. Li, and Y. He, "Traffic flow prediction based on hybrid model of auto-regressive integrated moving average and genetic particle swarm optimization wavelet neural network," vol. 41, no. 9, pp. 2273–2279, 2019.

[36] C.-C. Chang, W.-M. Lin, and C.-A. Lai, "Using real-time dynamic prediction to implement IoV-based collision avoidance," *Appl. Sci.*, vol. 9, no. 24, p. 5370, Dec. 2019.

[37] M. Zheng, T. Li, R. Zhu, J. Chen, Z. Ma, M. Tang, Z. Cui, and Z. Wang, "Traffic accident's severity prediction: A deep-learning approach-based CNN network," *IEEE Access*, vol. 7, pp. 39897–39910, 2019.

[38] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, and D. Manocha, "TrafficPredict: Trajectory prediction for heterogeneous traffic-agents," in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2019, vol. 33, no. 1, pp. 6120–6127.

[39] B. Lv, R. Sun, H. Zhang, H. Xu, and R. Yue, "Automatic vehicle-pedestrian conflict identification with trajectories of road users extracted from roadside LiDAR sensors using a rule-based method," *IEEE Access*, vol. 7, pp. 161594–161606, 2019.

[40] L. Peng, M. A. Sotelo, Y. He, Y. Ai, and Z. Li, "Rough set based method for vehicle collision risk assessment through inferring driver's braking actions in near-crash situations," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 2, pp. 54–69, Summer. 2019.

[41] C. Katrakazas, M. Quddus, and W.-H. Chen, "A new integrated collision risk assessment methodology for autonomous vehicles," *Accident Anal. Prevention*, vol. 127, pp. 61–79, Jun. 2019.

[42] G. Gao, M. V. Wüthrich, and H. Yang, "Evaluation of driving risk at different speeds," *Insurance, Math. Econ.*, vol. 88, pp. 108–119, Sep. 2019.

[43] M. Ayuso, M. Guillen, and J. P. Nielsen, "Improving automobile insurance ratemaking using telematics: Incorporating mileage and driver behaviour data," *Transportation*, vol. 46, no. 3, pp. 735–752, Jun. 2019.

[44] H. Zhao, H. Cheng, T. Mao, and C. He, "Research on traffic accident prediction model based on convolutional neural networks in VANET," in *Proc. 2nd Int. Conf. Artif. Intell. Big Data (ICAIBD)*, May 2019, pp. 79–84.

[45] N. Lyu, J. Wen, Z. Duan, and C. Wu, "Vehicle trajectory prediction and cut-in collision warning model in a connected vehicle environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 966–981, Feb. 2022.

[46] H. Zhao, H. Cheng, Y. Ding, H. Zhang, and H. Zhu, "Research on traffic accident risk prediction algorithm of edge Internet of Vehicles based on deep learning," vol. 42, no. 1, pp. 50–57, 2020.

[47] L. Peng, M. Angel Sotelo, Y. He, Y. Ai, and Z. Li, "A method for vehicle collision risk assessment through inferring driver's braking actions in near-crash situations," 2020, *arXiv:2004.13761*.

[48] K. Bhalla and K. Gleason, "Effects of vehicle safety design on road traffic deaths, injuries, and public health burden in the Latin American region: A modelling study," *Lancet Global Health*, vol. 8, no. 6, pp. 819–828, Jun. 2020.
[49] F. Walugembe, F. Levira, B. Ganesh, and D. W. Lwetoijera, "A retrospective study on the epidemiology and trends of road traffic accidents, fatalities and injuries in three municipalities of dar es Salaam Region, Tanzania between 2014–2018," *Pan Afr. Med. J.*, vol. 36, pp. 1–10, May 2020.
[50] X. Wang, J. Liu, T. Qiu, C. Mu, C. Chen, and P. Zhou, "A real-time collision prediction mechanism with deep learning for intelligent transportation system," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9497–9508, Sep. 2020.
[51] Y. Lin and R. Li, "Real-time traffic accidents post-impact prediction: Based on crowdsourcing data," *Accident Anal. Prevention*, vol. 145, Sep. 2020, Art. no. 105696.
[52] C. Gutierrez-Osorio and C. Pedraza, "Modern data sources and techniques for analysis and forecast of road accidents: A review," *J. Traffic Transp. Eng.*, vol. 7, no. 4, pp. 432–446, Aug. 2020.
[53] L. Yu, B. Du, X. Hu, L. Sun, L. Han, and W. Lv, "Deep spatio-temporal graph convolutional network for traffic accident prediction," *Neurocomputing*, vol. 423, pp. 135–147, Jan. 2021.
[54] J. Fang, D. Yan, J. Qiao, J. Xue, and H. Yu, "DADA: Driver attention prediction in driving accident scenarios," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 6, pp. 4959–4971, Jun. 2022.
[55] J. Lee, S. Kishino, and K. Suzuki, "Prediction of collision avoidance ability of two-wheeled vehicle riders using driving behaviors and emotional states," *Int. J. Automot. Eng.*, vol. 12, no. 2, pp. 32–40, 2021.
[56] G. Li, Y. Yang, T. Zhang, X. Qu, D. Cao, B. Cheng, and K. Li, "Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios," *Transp. Res. C, Emerg. Technol.*, vol. 122, Jan. 2021, Art. no. 102820.
[57] Y. Zhao, T. Miyahara, K. Mizuno, D. Ito, and Y. Han, "Analysis of car driver responses to avoid car-to-cyclist perpendicular collisions based on drive recorder data and driving simulator experiments," *Accident Anal. Prevention*, vol. 150, Feb. 2021, Art. no. 105862.
[58] G. Li, Y. Liao, Q. Guo, C. Shen, and W. Lai, "Traffic crash characteristics in Shenzhen, China from 2014 to 2016," *Int. J. Environ. Res. Public Health*, vol. 18, no. 3, p. 1176, Jan. 2021.
[59] X. Zhao, T. Lu, and Y. Dai, "Individual driver crash risk classification based on IoV data and offline consumer behavior data," *Mobile Inf. Syst.*, vol. 2021, pp. 1–10, Jun. 2021.
[60] D.-J. Lin, M.-Y. Chen, H.-S. Chiang, and P. K. Sharma, "Intelligent traffic accident prediction model for Internet of Vehicles with deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 3, pp. 2340–2349, Mar. 2022.
[61] C.-C. Chang, Y.-M. Ooi, and B.-H. Sieh, "IoV-based collision avoidance architecture using machine learning prediction," *IEEE Access*, vol. 9, pp. 115497–115505, 2021.
[62] L. Yang, Z. Wang, L. Ma, and W. Dai, "Transfer learning-based vehicle collision prediction," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–9, Aug. 2022.
[63] C. Gutierrez-Osorio, F. A. González, and C. A. Pedraza, "Deep learning ensemble model for the prediction of traffic accidents using social media data," *Computers*, vol. 11, no. 9, p. 126, Aug. 2022.
[64] A. Morimoto, A. Wang, and N. Kitano, "A conceptual framework for road traffic safety considering differences in traffic culture through international comparison," *IATSS Res.*, vol. 46, no. 1, pp. 3–13, Apr. 2022.
[65] A. V. Malawade, S.-Y. Yu, B. Hsu, D. Muthirayan, P. P. Khargonekar, and M. A. A. Faruque, "Spatiotemporal scene-graph embedding for autonomous vehicle collision prediction," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9379–9388, Jun. 2022.
[66] N. Suat-Rojas, C. Gutierrez-Osorio, and C. Pedraza, "Extraction and analysis of social networks data to detect traffic accidents," *Information*, vol. 13, no. 1, p. 26, Jan. 2022.
[67] R. Pan, L. Jie, X. Zhang, S. Pang, H. Wang, and Z. Wei, "A V2P collision risk warning method based on LSTM in IOV," *Secur. Commun. Netw.*, vol. 2022, Jul. 2022, Art. no. 7507573.
[68] G. Kothai, E. Poovammal, G. Dhiman, K. Ramana, A. Sharma, M. A. AlZain, G. S. Gaba, and M. Masud, "A new hybrid deep learning algorithm for prediction of wide traffic congestion in smart cities," *Wireless Commun. Mobile Comput.*, vol. 2021, pp. 1–13, May 2021.
[69] C. Onyeneke, C. Eguzouwa, and C. Mutabazi, "Modeling the effects of traffic congestion on economic activities-accidents, fatalities and casualties," *Biomed. Statist. Informat.*, vol. 3, no. 2, pp. 7–14, 2018.
[70] S. Du, T. Li, X. Gong, and S.-J. Horng, "A hybrid method for traffic flow forecasting using multimodal deep learning," 2018, *arXiv:1803.02099*.
[71] W. Wei, H. Wu, and H. Ma, "An AutoEncoder and LSTM-based traffic flow prediction method," *Sensors*, vol. 19, no. 13, p. 2946, Jul. 2019.
[72] A. Hébert, T. Guédon, T. Glatard, and B. Jaumard, "High-resolution road vehicle collision prediction for the city of Montreal," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2019, pp. 1804–1813.
[73] A. Moses and R. Parvathi, "Vehicular traffic analysis and prediction using machine learning algorithms," in *Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng. (IC-ETITE)*, Feb. 2020, pp. 1–4.
[74] N. Ranjan, S. Bhandari, H. P. Zhao, H. Kim, and P. Khan, "City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN," *IEEE Access*, vol. 8, pp. 81606–81620, 2020.

K. RAVEENDRA REDDY received the bachelor's and master's degrees from the Sri Sai Institute of Technology and Science Affiliated to Jawaharlal Nehru Technological University Anantapur. He is currently pursuing the Doctor of Philosophy degree in computer science engineering with the Vellore Institute of Technology, Chennai, under the guidance of Dr. A. Muralidhar. He is an Assistant Professor with the Guru Nanak Institute of Technology, Hyderabad, India. He has totally professional experience of more than six years working in various prestigious institutions. He has published two papers in various national conferences. His research interests include machine learning, federated learning, network security, database technologies, and big data analytics.

A. MURALIDHAR received the bachelor's degree from Sri Krishna Devaraya University, the master's degree from Jawaharlal Nehru Technological University, Hyderabad, and the Doctor of Philosophy degree in computer science from the Vellore Institute of Technology, Chennai, India. He is currently an Associate Professor with the Vellore Institute of Technology. He has a total professional experience of more than 18 years working in various prestigious institutions. He has published 20 papers in various national and international peer-reviewed journals and conferences. His research interests include machine learning, knowledge discovery, data mining, database technologies, and big data analytics.