IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Urban Traffic Safety: A Review of Data Sources and Machine Learning Models for Vehicle Accident Risk Assessment

Dr. Dinesh Yadav
Associate Professor, CSE Department
St. Andrews Institute of Technology & Management, Gurugram, Haryana, India
dinesh.yadav@saitm.ac.in

Abstract: *Traffic roads in urban areas have become a major issue as the rates of accidents involving vehicles increase in such high-population cities. This survey provides an informative overview of machine learning (ML) models and databases employed in vehicle accident risk evaluation. Classical and new data inputs can include sensor data, GPS, surveillance, crowd-sourced information which is divided into conventional and new data as well as assessed on how they can relate to risk modelling. The article discusses a wide range of ML techniques in the following way: decision trees, SVMs, ensemble methods, and deep learning models, such as graph neural networks (GNNs). Practical developments of smart cities and intelligent transport systems (ITS) are discussed, with the focus on cloud computing analytics and predictive safety models. Such issues as data integration, interpretability of the models, the ability to make predictions in real-time, and the desire to improve the standardization were mentioned as key challenges. The proposed review is expected to advise further researchers and urban planners on developing scalable, accurate, and transparent smart transportation ecosystem risk assessment systems.*

Keywords*:* Road Safety, Vehicle Accident Risk Assessment, Traffic Data Sources, Machine Learning Models, Smart Cities, Real-Time Prediction

I. INTRODUCTION

The collection and use of data have significantly driven advances across multiple sectors, including science, healthcare, and transportation. However, in the context of urban traffic systems, the growing reliance on data also introduces concerns related to data privacy, security, and ethical usage [1]. As Smart cities and intelligent transportation systems (ITS) become common, personal and locational sensitive information is often gathered through embedded system sensors, surveillance cameras, Global Positioning systems and vehicular communications networks. This sensitive information must be handled with privacy privacy-preserving mechanism and safe data governance in the process of collection, processing, and release.

The rising use of smart sensing (traffic cameras, LiDAR, infrared, and speech recognition devices) has led to tremendous amounts of heterogeneous and high-frequency traffic data. Such data can be of large volumes, high velocity, and variety and traditional machine learning (ML) algorithms can be challenged by such data [2]. As part of a counter-reaction, distributed machine learning frameworks and edge computing solutions have been introduced, which allow real-time data ingestion, processing, and learning via the usage of high-performance computing clusters or cloud-based infrastructure.

Safety-wise, even though the accidents that happen in urban traffic are sometimes random and do not occur that often, they nevertheless have spatial and temporal patterns with time [3]. Road geometry, built environment, traffic characteristics, road behaviour and weather conditions sway these patterns. The urban accident blackspots can be determined by performing a historical crash analysis along with the setting features of the surrounding area. Warning messages of accidents can then be sent to drivers via vehicle-to-infrastructure (V2I) and vehicle-to-roadside unit (V2R) communication devices in order to deliver better situational awareness and make drivers adhere to traffic laws. Instant

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

sharing of the information of accidents prevents subsequent collisions, enhances emergency management, and efficient scheduling of routes by apprising other travelers that a particular location is crowded or risky as indicated in Figure. 1.

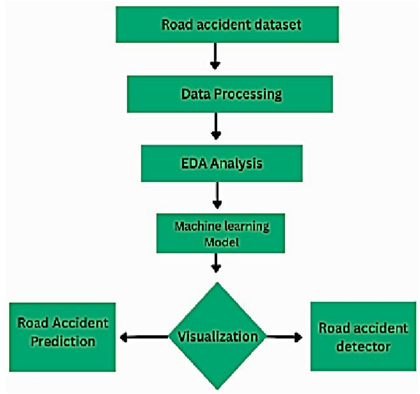

Fig. 1.Road Accident Assessment

To gain a better insight into the actual causes of urban traffic accidents it is necessary to produce use case taxonomies presenting such classifications as human error crashes, infrastructure problems related crashes, environmental conditions related crashes and vehicular dynamics related crashes [4]. Corresponding countermeasures, such as infrastructure redesign, signal timing optimization, or public awareness campaigns, can be extracted and categorized. A deep neural network (DNN)-based multilabel classification model can be trained to map specific crash scenarios to effective mitigation strategies, supporting decision-makers in deploying context-aware interventions according to risk prioritization scores.

Machine learning techniques are increasingly used to predict crash likelihood, classify accident severity, and model traffic safety risks under specific spatial and temporal conditions [5]. Predictive models rely on multimodal historical datasets, including collision reports, injury severity records, real-time traffic volume data, speed data, road surface conditions, and weather reports [6]. ML models, such as random forests, gradient boosting [7], support vector machines, and deep learning architectures, have demonstrated high accuracy in detecting patterns and forecasting crash events [8]. ML supports hotspot detection, severity prediction, behavior analysis, and policy evaluation [9]. Combining data from police, sensors, and connected vehicles enhances traffic safety modeling.

A. Structure of the paper

The paper is organized as follows: Section II reviews accident risk assessment for vehicles in urban traffic. Section III compares ML techniques in accident risk assessments. Section IV examines the sources of Traffic and Crash Data for urban safety risk assessment. Section V presents a summary of the literature, and Section VI concludes with key findings and future directions in real-time processing, scalability, and model interpretability.

II. UNDERSTANDING OF VEHICLE ACCIDENT RISK ASSESSMENT

Autonomous vehicles (AVs) are anticipated to enhance transportation comfort, performance, and safety compared to traditional human-driven vehicles, prompting car manufacturers to develop automated driving systems (ADS) that can surpass human capabilities [10]. To meet these expectations, an ADS must demonstrate natural driving behavior within flexible yet clearly defined safety margins that accommodate all road users and infrastructure boundaries [11]. Nevertheless, these safety ranges are very difficult to identify and quantify in real-life driving conditions facing individuals involved in AV safety evaluation. This is because maintaining and confirming the safe operation of the AVs prior to their configuration in real-life situations on the road is an essential issue in this regard, for example in Figure 2.

DOI: 10.48175/IJARSCT-28333

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

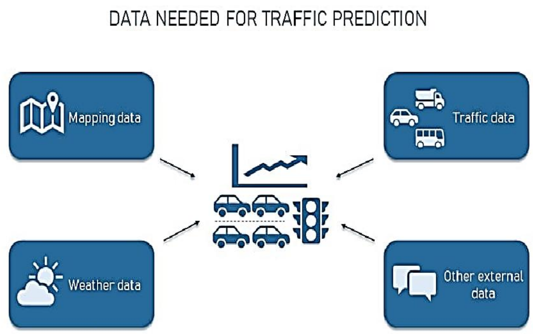

Fig. 2. Data Needed for Traffic Prediction

The accurate traffic prediction relies on four key data types: mapping data, traffic data, weather data, and other external factors such as events or incidents. These combined inputs help forecast traffic flow and improve road safety and efficiency.

A. Definition and Importance of Road Accident Risk Assessments

Road accident risk assessment involves evaluating and identifying areas with a high likelihood of traffic crashes to enhance safety and efficiency in urban transportation systems. Intersections, as critical convergence points in urban road networks, often become hotspots for accidents due to dense traffic, multidirectional flows, and complex environments. With advancements in vehicle-to-infrastructure (V2I) communication, the deployment of Roadside Units (RSUs) has emerged as a key strategy for mitigating risks at these locations. By analyzing historical crash patterns shaped by traffic organization and the built environment authorities can identify high-risk areas where RSUs can issue early warnings to drivers through vehicle-to-RSU (V2R) communication, helping prevent accidents and alerting drivers in real time. Additionally, timely dissemination of accident data supports dynamic route adjustments, enabling other road users to avoid congested or hazardous zones [12]. Thus, integrating accident risk assessment with RSU deployment not only enhances roadway safety but also improves traffic efficiency in complex urban settings.

B. Key Factors Influencing Accident Risks

Based on this, scholars have conducted extensive research on elevator safety, focusing on elevator safety supervision, risk analysis methods, risk warning methods, and other related aspects. In conjunction with the research theme of this paper, the study focuses on elevator safety risk analysis methods. Currently, elevator safety risk analysis methods primarily include qualitative analysis methods, safety checklist methods, system analysis methods, root cause analysis, and data mining, among others. Among the qualitative analysis methods, for example, an elevator safety evaluation method combines expert recommendations with objective factors. Evaluated elevator risk by employing a safety checklist to collect data and using machine learning methods. Constructed an AHP-YAAHP elevator safety risk assessment model based on hierarchical analysis and expert experience, which further improved the computational efficiency of elevator safety assessment methods. Proposed an elevator safety evaluation model based on Gray correlation analysis and hierarchical analysis method [13], and carried out model validation.

C. The Role of Smart Cities in IOT Accident Detection

It is known that cities are getting even more congested with regard to tourists, population and traffic. Traffic has increased as a result of increase in the number of vehicles thus resulting in the corresponding increase in road traffic accidents. A recent WHO report showed that every year, 1.35 million people die and 50 million people get injured. Road accidents are ranked as the eighth leading cause of death, with the Association for Safe International Road Travel (ASIRT) predicting that it may rise to the fifth leading cause of death soon unless drastic changes occur. In addition to the social harm caused by road traffic accidents, there is a significant financial cost. ASIRT states that road accidents have been robbing the annual budget of a given country between 1-2%.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, July 2025 Impact Factor: 7.67

III. MACHINE LEARNING TECHNIQUES IN ACCIDENT RISK ASSESSMENT

In the era of big data, ML offers powerful capabilities for accident risk assessment by uncovering hidden patterns and complex relationships within vast datasets. Traditional risk assessment approaches often reliant on questionnaire-based surveys are limited in scope and fail to provide a holistic and systematic evaluation of accident causation [14]. The ML-based models, however, have the ability to analyze highly complex causal networks that lie at the ground of construction and traffic safety incidents. Advanced frameworks such as complex networks, neural networks, and Bayesian Networks (BNs) are particularly well-suited for modeling and interpreting these intricate systems.

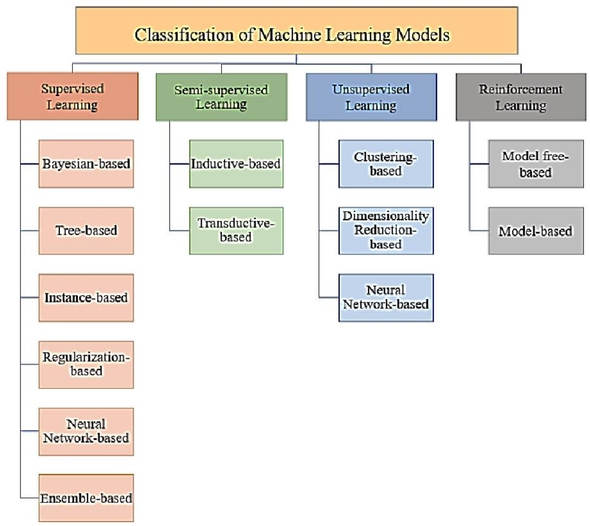

Fig. 3. Classification of Machine Learning Models

One particularly well-suited structure among them is represented by BNs (which have been used effectively as a means of dealing with uncertainty and to model probabilistic causalities between factors of accidents). By mapping the propagation of probabilities from root causes to actual accidents (as illustrated in Figure. 3), BNs enhance understanding of how various contributing factors interact, offering valuable insights for proactive risk mitigation and decision-making.

A. Overview of ML and DL Models Used

ML and DL being a subdivision of AI are widely popular within this industry and their application has been improving a lot as an automated intrusion detection mechanism. They have proven effective in identifying advanced threats including malicious websites, distributed denial of service attacks and botnet attacks. ML and DL models depend on representative network traffic data during the training phase to learn how to accurately distinguish between normal and malicious behaviours. These models enable automated, real-time detection and prevention [15]. Moreover, IDS based on ML and DL techniques are capable of detecting new and unknown attacks that traditional techniques are ineffective at countering. The study is based on a detailed examination of several articles published between 2017 and 2025. It highlights the challenges associated with the targeted networks. These challenges impose constraints that must be considered for effective future ML- and DL-based intrusion detection model propositions. This approach ensures optimal performance and robust security solutions. In this article, only interested in presenting current ML and DL-based approaches, but also more focused on the network for which these approaches are proposed [16]. The originality of the work lies in providing an integrated perspective on intrusion detection techniques across three distinct networks simultaneously: cloud computing, the IoT, and SDN. Identified different factors that affect the effectiveness of the proposed approaches related to the network for which they are proposed, the classifier applied, and the dataset used for training.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 2, July 2025 Impact Factor: 7.67

B. Supervised Vs Unsupervised Learning

Data analysis, pattern recognition, and decision-making with little to no human involvement are the hallmarks of ML, a subfield of artificial intelligence and a subfield of computer science. Automated future prediction is possible thanks to ML, which does not require human intervention or programming. For predictive analytics, ML offers a plethora of algorithms [17]. Data analysis and probability prediction are two main applications of ML algorithms, one of which is supervised and the other unsupervised. Predicting patterns and trends in business and industry for the future is the focus of this chapter's comparative assessment of the main types of ML algorithms.

C. Unsupervised and Semi-Supervised Models

In practical vehicle crash data, labelled data like cases where various known causes exist or levels of severity of the crash are frequently scarce or lacking. Unsupervised and semi-supervised learning methods are hence priceless when there is an attempt to mine valuable information out of partially labelled or completely unlabelled data. The models can discover obscure arrangements, flawed incidents as well as exceptional designs, which create perils of accidents and actionable findings in traffic security management as well as city planning.

1. Clustering Techniques

Clustering methods aggregate data by an intrinsic property of the data, making it easy to find spatial or temporal patterns of accidents. Such ways are very useful in doing such things as identification of crash hotspots, profiling of drivers and segmentation of roads.

- **K-means Clustering:** An algorithm based on the centroid which partitions the data on accidents into a specified number of clusters. It is widely applied to the finding of traffic areas that have similar crash tendency thus, interventions can be placed to address these areas and place resources.
- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):** Density-Based Clustering It is a density-based clustering method that figures out an arbitrary-shaped gathering and detaches noise or an outlier. DBSCAN is particularly convenient when analyzing spatial crash data, since it does not need to know beforehand the number of clusters and does not suffer in the presence of outliers, an advantage in locating accident hotspots in city road networks.

These clustering methods support visualization and policy formulation by revealing high-risk locations or temporal clusters of traffic incidents without requiring supervision.

2. Anomaly Detection

Anomaly detection methods help identify rare or abnormal patterns in traffic data, such as unusual driving behavior or severe but infrequent crashes [18]. Techniques such as Isolation Forest, One-Class SVM, and Autoencoders are effective in detecting outliers, particularly when labelled data is limited. These models support early warning systems and enhance real-time risk monitoring in intelligent transportation systems.

D. Deep Learning Approaches

The capacity of DL algorithms to learn hierarchical representations from high-dimensional raw data has made them useful tools for assessing the likelihood of car accidents. Images, videos, sensor streams, structured traffic records, and other data sources can have their spatial, temporal, and contextual patterns automatically extracted using DL architectures, as opposed to conventional ML models that depend substantially on human feature engineering.

[19]. This capacity is particularly advantageous for real-time accident prediction, traffic scene understanding, and modelling driver behaviour.

- **Multilayer Perceptron (MLP):** MLPs are feedforward neural networks commonly used for classification or regression tasks involving structured datasets [20]. In accident risk modeling, MLPs can process features such as weather, road conditions, traffic volume, and vehicle type to predict accident severity or likelihood.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

- **Recurrent Neural Networks (RNN):** RNNs are tailored for modeling sequences where the order of data points matters. In accident prediction, RNNs can process temporal data such as speed, acceleration, and historical crash records to forecast future events [21]. Though effective, standard RNNs suffer from vanishing gradients in long sequences, which is mitigated by architectures like LSTM and GRU.
- **Graph Convolutional Networks (GCN):** GCNs extend DL to non-Euclidean data structures, such as graphs representing road networks. In traffic safety, GCNs are employed to model spatial relationships between intersections, road segments, or sensor nodes, enabling the prediction of crash probabilities across the network. They can incorporate both static features and dynamic features for comprehensive risk assessment.
- **By integrating spatial and temporal information:** these DL models offer robust solutions for proactive crash detection, risk hotspot forecasting, and adaptive traffic management. The continued advancement of DL techniques, especially in multi-modal and graph-based modelling, holds great promise for developing intelligent transportation systems that are both predictive and preventive in nature.

E. ML and DL Popular Models Used in Risk Management

Various ML algorithms have been used to identify and determine the risk of vehicle accidents. Models used commonly in the literature have the following characteristics: they are able to learn over large datasets, learn complex patterns and generalize to novel data:

- **Decision trees (DT):** Easy-to-interpret models that divide data (on thresholds of a feature). They are very popular in classifying accidents because they are transparent.
- **Random Forests (RF):** Decision trees ensemble, which increases accuracy in predictions and lowers overfitting. Works well on noisy traffic data and Variable importance analysis.
- **Support Vector Machines (SVM):** Good in classification problems where there are only two classes, particularly in problems that can be linearly separated. It has been applied in the areas of crash severity categorization and identifying risky driving behaviors.
- **Neural Networks (NN):** This is able to learn non-linear, complex relationships [22]. Variants of DL (CNNs, RNNs, LSTM) are used with image/video data (CCTV) and for sequential traffic flow forecasting.
- **Ensemble models are good at, e.g., XGBoost, LightGBM:** Rank the stacked models into the predictions to enhance the generalization. They are commonly run on massive accident data and demonstrate state-of-the-art results.

IV. SOURCES OF TRAFFIC AND CRASH DATA FOR URBAN SAFETY RISK ASSESSMENT

Urban traffic safety analysis utilizes a wide range of data sources to enhance accident risk assessment and mitigation. Public datasets such as NHTSA, STATS19, and CRSS provide structured records of reported crashes, while sensor and IoT data offer real-time insights into traffic flow and incidents. Crowdsourced platforms like Waze and Google Maps contribute user-reported data on congestion and road hazards [23]. Vehicle telematics systems capture in-depth driving behaviour and environmental data, including weather conditions and road characteristics, to help contextualize crash events. Additionally, social media and textual sources, such as Twitter and police reports, provide real-time, unstructured information that is valuable for situational awareness. However, integrating these heterogeneous data sources poses significant challenges related to interoperability, data quality, privacy, and completeness.

A. Common Causes and Patterns of Urban Vehicle Accidents

Urban traffic accidents are commonly caused by a mix of human error, infrastructural deficiencies, and environmental factors. Frequent contributors to road accidents include distracted driving, often due to mobile phone use or navigation systems, speeding in congested or residential zones, failure to yield at intersections and pedestrian crossings, poor visibility during night-time or adverse weather conditions, and inadequate infrastructure, such as missing signs or faulty traffic signals. These accidents often cluster near signalized intersections, arterial roads, school zones, and transit hubs. Temporal patterns also reveal higher crash frequencies during peak commuting hours and weekends. Recognizing these

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

causes and patterns is essential for designing targeted, data-driven safety interventions.

B. Importance of Data-Driven Accident Risk Assessment

Traditional traffic safety strategies have primarily relied on retrospective crash data and reactive interventions, which, while informative, are often insufficient for proactive prevention. With the advent of intelligent transportation systems powered by traffic sensors, GPS, surveillance cameras, connected vehicles, and mobile applications, urban environments now produce vast amounts of real-time, high-resolution data. Leveraging this data, data-driven accident risk assessment enables predictive analytics to identify high-risk locations before incidents occur, dynamic risk modelling that incorporates temporal and spatial traffic patterns and driver behaviour, and evidence-based policy making grounded in continuous monitoring and evaluation. It also supports the design of targeted safety interventions tailored to vulnerable user groups and accident-prone zones. By integrating ML, statistical modelling, and big data analytics, cities can transition from reactive to proactive traffic safety management, reducing crash rates, saving lives, minimizing economic losses, and enhancing overall transportation efficiency.

C. Applications and Real-World Implementations in Urban Traffic Safety

This section highlights the practical, real-world implementation of data-driven approaches and advanced technologies in urban traffic systems to enhance safety and reduce accident risks. These solutions are key components of smart city initiatives and Intelligent Transportation Systems (ITS), designed to enhance road safety, optimize traffic flow, and promote sustainable urban mobility. Various cities and transportation agencies have deployed systems that utilize real-time data from traffic sensors [24], GPS devices, surveillance cameras, and connected infrastructure to monitor traffic conditions, identify high-risk areas, and guide timely interventions. Dynamic traffic signal control, congestion monitoring dashboards, crash hotspot mapping, and crash hotspot mapping, as well as incident detection system programs are a few examples of real-world programs [25]. These implementations enable authorities to make informed decisions, allocate resources efficiently, and respond promptly to potential hazards, ultimately contributing to safer and more efficient urban transportation networks. They are given below:

1. New York City Vision Zero Program

Applies predictive analytics to determine which areas pose great danger and implements safety measures by use of speed cameras, traffic calming, and better signage.

2. Transportation Services of Toronto

Introduces real-time traffic conditions to ML models to determine the possibility of collisions and optimize the effectiveness of emergency response operations.

3. Integrated Traffic Management Systems (ITMS)

of India: Predict congestion and hotspot areas using CCTV, GPS, and sensor data so proactive policing can minimize traffic problems and the road planning can be improved[26].

4. Real-time Crash Risk Prediction Systems

Advanced ML and AI-powered models enable the development of real-time crash risk prediction systems. By integrating data from traffic cameras, GPS devices, weather sensors, and vehicular telematics, these systems provide timely alerts and forecasts for potential accident-prone scenarios. This facilitates proactive interventions by traffic authorities and helps drivers avoid high-risk zones.

5. Smart Traffic Management and Routing

Data-driven traffic management systems use predictive analytics and optimization algorithms to dynamically control traffic signals, reroute vehicles, and minimize congestion. Real-time data inputs from road sensors and connected vehicles enhance the accuracy of traffic flow predictions, reduce bottlenecks, and improve road safety, especially during peak hours and emergencies.

6. Urban Planning and Infrastructure Development

Urban planners leverage crash data analytics to identify high-risk areas and design safer infrastructure. Visualization tools and spatial crash pattern analysis inform decisions on where to add pedestrian crossings, traffic calming measures, road lighting, or redesign intersections. Data-driven planning ensures safety-centric urban growth.

7. Insurance and Policy Making

Crash data and predictive models support insurers in risk assessment, premium pricing, and fraud detection. Policymakers utilize these insights to inform the development of evidence-based traffic laws, safety regulations, and transportation policies. Actuarial models enriched with real-world data foster more equitable and effective road safety regulations.

8. Public Awareness and Preventive Campaigns

Visual crash analytics and data storytelling enhance public engagement and awareness, promoting a deeper understanding of road safety. Interactive platforms and targeted campaigns based on crash trends educate road users about risk behaviors, promoting safer driving habits. These efforts support community-level behavior change and contribute to a long-term reduction in accident rates.

D. Urban Traffic Prediction Tools with Smart City Safety Frameworks

Urban traffic prediction tools are key components of smart city safety systems, using real-time and historical data to forecast traffic flow and assess accident risk [27]. Integrated with Intelligent Transportation Systems (ITS), these tools help city authorities manage congestion, identify high-risk areas, and implement timely safety measures to enhance urban mobility and road safety:

- Traffic API and Waze of Google: Combine real-time data in one source and use crowd-sourced resources to warn the user of possible dangers or congestion that could cause accidents.
- World Bank + Map box: Open Traffic Platform: Relying on anonymized GPS data, it is able to assist cities in low- and middle-income countries to keep track of and predict traffic patterns.
- Artificial intelligence-powered Dashcam systems: Applied to a car in the fleet with regard to examining habits of the driver and making warnings of possible crashes and violations of the lanes and dangerous manoeuvres.

In current smart city safety systems, ML is a crucial aspect. These systems are based on the incorporation of traffic management, city planning, as well as emergency response:

- Intelligent Traffic Control programs: Predictive-based adaptive traffic signals minimize chances of crashes and congestion during peak times.
- Artificial Intelligence in Road Designing: Use the accident data and propose infrastructure changes [28],e.g. speed bumps, pedestrian area.
- Autonomous Integration of Vehicles: ML modelling is enabling smart cities to make their autonomous vehicles communicate instantaneously to allow reaching an ideal coordination that eliminates collisions and other accidents.

V. LITERATURE REVIEW

The section contains a systematic overview of the latest works in the areas of data architecture, big data integration, and ML in application-related risk assessment of vehicle accidents. This study, which was reviewed, proposes a combination of theoretical innovation and practical use that can contribute to the enhancement of safe transport systems founded on data, as shown in Table I.

Gavric et al. (2025) propose the Urban Surface and Air Mobility (USaAM) concept that enables door-to-door travel, utilizing the same fly-drive vehicle for both surface and airborne transportation. The proposed USaAM concept provides a traffic control system for surface and airborne traffic, allowing for the integration of fly-drive vehicles into the existing surface traffic. To simulate the proposed concept, a first-of-its-kind microsimulation platform is developed. Multiple traffic demand profiles were assessed on an urban corridor, revealing that USaAM has significant potential. However, their study also shows that as surface traffic demand increases, landing performance deteriorates, while increases in both surface and airborne traffic demands lead to declines in takeoff performance. This paper contributes to the understanding of the feasibility and limitations of integrating fly-drive vehicles into urban transportation systems and is the first step toward a state-of-the-art microsimulation tool for future 3D transportation.[29]

Gao et al. (2025) vehicle speed and acceleration probabilistically modeled using GD, while entropy theory is introduced to quantify risk uncertainty. A risk assessment model based on graph neural networks (GNNs) is then designed to capture the spatiotemporal dynamics of multivehicle interactions and predict the potential risk levels of driving strategies. The results demonstrate that the framework accurately quantifies collision risks in multivehicle interactions in complex traffic scenarios, with high accuracy and robustness across typical situations such as cruising, cut-ins, lane changes, overtaking, and different density traffic. By thoroughly analyzing traffic risk characteristics and incorporating them into intelligent driving decision-making, this study provides significant technical insights and theoretical support for enhancing the safety and decision-making efficiency of autonomous driving systems [30].

Chai et al. (2025) carried out a detailed statistical study of the rising trend of hazmat accident frequencies in U.S. highway transportation sector. Relying on the findings, came up with a superior model that would act as a strong source of data to support risk warning efforts. The trend is an indication that prompt measures should be increased to improve the precision of accident risk alerts to reduce losses in the economy. This research also shows that the majority of the accidents happen between 6:00 and 11:00 and 91.7 percent of all those accidents lead to spillage. The discovery highlights the importance of a proper emergency response strategy that is unique and specific to incidents of spillages. The challenge of performance curve reduction of models in large volumes was solved by creating the model process named SF-T0.25 on the stacking algorithm, which was ensured by validation with more than 70,000 records of spillage accidents in 2021-2023. The results show that the prediction accuracy of the model reaches 0.9628, which is better than the parameter-adjusted ET model (0.94981). The SF-T0.25 model likewise has good results in such indicators as Jaccard similarity coefficient and the cross-entropy [31].

Almutairi et al. (2025) present an innovative accident prediction and prevention model that is referred to as A-LAPPM (Attention-based Long- and Short-Term Memory Autoencoder). This model aims at enhancing safety. The data received from vehicle sensors, Vehicle-to-Vehicle (V2V) communication, and ambient variables are all incorporated into the model during the process of identifying and responding to potential accident hazards. Critical temporal patterns and danger indicators are captured by the model through the utilization of sequential learning through Long- and Short-Term Memory (LSTM) units and the enhancement of focus through the utilization of an attention mechanism. Massive experiments do assess the utility of the A-LAPPM model which is proposed. Such tests are based on more important performance indicators, namely the correctness of the predictions made, the speed of the response, the decrease in the incidence of accidents, the lack of failure in the decision-making process, and the robustness to untruthful information as well as the model leads to approximately 11.8 percent increase in the accuracy of the predictions, 28.5 percent faster rate of response, and 50 percent lower rates of accidents, which ultimately contribute to the overall paradigm of the autonomous vehicles to endlessly perform better in terms of denominator in driving situations being complex [32].

Goswami et al. (2024) aim to design a model for the machine learning assessment of traffic accidents and severity based on weather, time, and location. These include the numerical such as the temperature and wind speed while the other is categorical such as weather conditions. For missing values and categorical inputs, the pre-processing step is taken, for temporal and interaction features the feature engineering step is performed and out of all these steps the most crucial factors for the accident severity are filtered from all the variables. This study proves that XGBoost has the fastest and highest accuracy in classifying the level of seriousness of the accident. Such a discovery opens up the possibilities of applying ML methods at a higher advanced level to improve traffic safety. It is crucial to point out that this research pays great attention to the application of extended analytical models and big data in increasing traffic safety and outlining prevention measures [33].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 2, July 2025
Impact Factor: 7.67

Wu et al. (2024) propose a driving risk assessment approach based on ML on the basis of the driver 's timely response to the accident and the successful braking of the vehicle. This approach combines multi-dimensional driving data and ML to establish a vehicle braking deceleration and braking distance prediction model to predict the braking distance of the vehicle. Then, combined with the safety distance threshold and historical braking data, the data are compared and analyzed to predict the driving risk level. Through simulation experiments, it can be seen that compared with the traditional linear regression prediction method, the braking distance error is reduced by 0.74 m, and the error is relatively reduced, which can effectively improve driving safety [34]

Table 1: Literature Summary on Urban Traffic Safety: Data Sources and Machine Learning Models for Vehicle Accident Risk Assessment

Reference	Study On	Approach	Key Findings	Challenges	Future Direction
Gavric, et al. (2025)	Integration of fly-drive vehicles into urban transportation systems	Microsimulation platform for USaAM (Urban Surface and Air Mobility)	Traffic demand profiling; analysis of surface-air coordination	USAAM shows high potential for 3D urban travel; increased surface traffic reduces landing performance; both air and surface loads impact takeoff performance	Scalability in real-world traffic; coordinating air-ground interactions
Gao, et al. (2025)	Collision risk quantification in multivehicle interactions using GNN	Graph Neural Network (GNN), Gaussian Distribution (GD), Entropy theory	Spatiotemporal modeling of vehicle dynamics and risk prediction	Accurate prediction of collision risks in various driving scenarios (e.g., cut-ins, lane changes); robust risk assessment model	Computational complexity; data sparsity in certain scenarios
Chai, et al. (2025)	Hazardous materials accident prediction and spillage risk in U.S. highways	Stacking algorithm-based model “SF-T0.25”; large dataset analysis	Time-based trend analysis; emergency response planning	High accuracy (0.9628) in predicting spillage accidents; most incidents occur 6:00–11:00 with 91.7% involving spillage	Performance degradation on large datasets; emergency response modeling
Almutairi, et al. (2025)	Accident prediction and prevention using LSTM and attention mechanisms	Attention-based LSTM Autoencoder (A-LAPPM); sensor and V2V data	Sequential learning for temporal hazard detection; attention mechanism for focus	11.8% higher prediction accuracy; 28.5% faster response; 50% accident rate reduction; improves autonomous vehicle safety	Vulnerability to false sensor data; generalizability to varied road conditions
Goswami, et al. (2024)	Traffic accident severity prediction based on weather, time, and location	ML models (e.g., XGBoost); feature engineering	Handling of categorical/missing data; benchmarking models on accuracy	XGBoost outperforms others in speed and accuracy; highlights weather and temporal features as key predictors	Data imbalance; real-time applicability
Wu, et al. (2024)	Risk assessment through	ML-based prediction using multi-	Comparison with historical braking records and safety	Reduced braking distance error by 0.74m compared to linear regression; enhanced	Limited model to braking response only; may not

	braking distance and deceleration prediction	dimensional driving data	thresholds	safety through better distance prediction	generalize across vehicle types

VI. CONCLUSION AND FUTURE WORK

This paper has examined the present state of vehicle accident risk evaluation as a combination of both the existing data sources of traffic information in a city and ML models that can be used to predict accidents. Observed a wide variety of data inputs, traditional ones such as traffic crash reports as well as fresh, real-time streams, such as sensor data, GPS trajectories, and even video surveillance feeds, which all have crucial roles in efficient risk analysis. As well, talked about other ML algorithms that have been found to be robust in their predictive power in this area, such as decision trees, random forests, support vector machines, ensemble techniques, and DL structures. Through these models, there has been potential accuracy in the prediction of the occurrence of accidents and their severity levels. But the sensitivity and accuracy of these predictive models largely depend upon the quality, the resolution, and the time sensitivity of the input data. Further, it was reported that real deployment settings, particularly those based on Smart City projects and Intelligent Transportation Systems (ITS) and improving the availability of standardized traffic data and using state-of-the-art ML techniques.

In future research, which should improve on real-time, explainable, and scalable accident risk prediction systems. With the combination of IoT, vehicle communication (V2V, V2I), and spatiotemporal DL models, the accuracy and context awareness of the predictions can be considerably improved. It is also important to outline the interpretability of models, ethical use of AI, and data fusion across domains so that the use of the smart city is reliable and fair.

REFERENCES

[1] A. Jenefa, J. M. D. B, C. J. S. K, E. S, S. J. W, and B. M. Kuriakose, "Utilizing RL and Web-Enhanced Commuting for Traffic Congestion Mitigation and Public Transportation Enhancement," in *7th International Conference on Electronics, Communication and Aerospace Technology, ICECA 2023 - Proceedings*, 2023. doi: 10.1109/ICECA58529.2023.10395723.

[2] N. Malali, "Next-Generation Augmented Data Science Platform for Autonomous Model Generation, Continuous Testing, Adaptive Deployment, and Real-Time Performance Optimization Using AI-Driven …," 2025

[3] S. R. Sagili and T. B. Kinsman, "Drive Dash: Vehicle Crash Insights Reporting System," in *2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA)*, IEEE, Oct. 2024, pp. 1–6. doi: 10.1109/ICISAA62385.2024.10828724.

[4] A. Balasubramanian, "AI-Driven Optimization of Urban Mobility: Integrating Autonomous Vehicles with Real-Time Traffic and Infrastructure Analytics," *Int. J. Innov. Res. Creat. Technol.*, vol. 5, no. 5, pp. 1–13, 2019.

[5] A. Borucka and S. Sobczuk, "The Use of Machine Learning Methods in Road Safety Research in Poland," *Appl. Sci.*, vol. 15, no. 2, p. 861, Jan. 2025, doi: 10.3390/app15020861.

[6] S. P. Ardakani *et al.*, "Road Car Accident Prediction Using a Machine-Learning-Enabled Data Analysis," *Sustainability*, vol. 15, no. 7, Mar. 2023, doi: 10.3390/su15075939.

[7] A. Balasubramanian, "Improving Air Quality Prediction Using Gradient Boosting," *Int. J. Sci. Technol.*, vol. 13, no. 2, pp. 1–9, 2022.

[8] M. Megnidio-Tchoukouegno and J. A. Adedeji, "Machine Learning for Road Traffic Accident Improvement and Environmental Resource Management in the Transportation Sector," *Sustain.*, vol. 15, no. 3, 2023, doi: 10.3390/su15032014.

[9] R. Q. Majumder, "Machine Learning for Predictive Analytics : Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, 2025.

[10] H. Muslim *et al.*, "Cut-Out Scenario Generation With Reasonability Foreseeable Parameter Range From Real

Highway Dataset for Autonomous Vehicle Assessment," *IEEE Access*, vol. 11, pp. 45349–45363, 2023, doi: 10.1109/ACCESS.2023.3268703.

[11] P. Khare, S. Arora, and S. Gupta, "Integration of Artificial Intelligence (AI) and Machine Learning (ML) into Product Roadmap Planning," in *2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP)*, IEEE, Aug. 2024, pp. 1–6. doi: 10.1109/ICECSP61809.2024.10698502.

[12] S. Zhang, S. Wang, S. Huang, X. Liu, X. Wang, and N. Chen, "Optimized Deployment Strategy for Roadside Units Based on Accident Risk Assessment and Simulation Validation," *IEEE Access*, vol. 12, pp. 83330–83339, 2024, doi: 10.1109/ACCESS.2024.3413018.

[13] Q. Liu *et al.*, "Elevator Group Usage Risk Assessment Model using IoT Platform Information," in *IEEE Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, 2022. doi: 10.1109/ITAIC54216.2022.9836657.

[14] J. Jiang, G. Liu, and X. Ou, "Risk Coupling Analysis of Deep Foundation Pits Adjacent to Existing Underpass Tunnels Based on Dynamic Bayesian Network and N–K Model," *Appl. Sci.*, vol. 12, no. 20, Oct. 2022, doi: 10.3390/app122010467.

[15] N. Prajapati, "The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications," *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[16] M. K. Ngueajio, G. Washington, D. B. Rawat, and Y. Ngueabou, "Intrusion Detection Systems Using Support Vector Machines on the KDDCUP'99 and NSL-KDD Datasets: A Comprehensive Survey," in *Lecture Notes in Networks and Systems*, 2023. doi: 10.1007/978-3-031-16078-3_42.

[17] N. Mearaj and M. A. Wani, "Zero-day Attack Detection with Machine Learning and Deep Learning," in *Proceedings of the 17th INDIACom; 2023 10th International Conference on Computing for Sustainable Global Development, INDIACom 2023*, 2023.

[18] R. Q. Majumder, "A Review of Anomaly Identification in Finance Frauds Using Machine Learning Systems," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 10, pp. 101–110, Apr. 2025, doi: 10.48175/IJARSCT-25619.

[19] S. Mathur and S. Gupta, "An Energy-Efficient Cluster-Based Routing Protocol Techniques for Extending the Lifetime of Wireless Sensor Network," in *2023 International Conference on the Confluence of Advancements in Robotics, Vision and Interdisciplinary Technology Management (IC-RVITM)*, IEEE, Nov. 2023, pp. 1–6. doi: 10.1109/IC-RVITM60032.2023.10434975.

[20] S. Rongala, S. A. Pahune, H. Velu, and S. Mathur, "Leveraging Natural Language Processing and Machine Learning for Consumer Insights from Amazon Product Reviews," in *2025 3rd International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES)*, 2025, pp. 1–6. doi: 10.1109/ICSSES64899.2025.11009528.

[21] N. Patel, "Enhanced Network Security: Real-Time Malicious Traffic Detection in SD-WAN Using LSTM-GRU Hybrid Model," in *2024 9th International Conference on Communication and Electronics Systems (ICCES)*, IEEE, Dec. 2024, pp. 826–833. doi: 10.1109/ICCES63552.2024.10860215.

[22] S. Nokhwal, P. Chilakalapudi, P. Donekal, S. Nokhwal, S. Pahune, and A. Chaudhary, "Accelerating Neural Network Training: A Brief Review," *ACM Int. Conf. Proceeding Ser.*, pp. 31–35, 2024, doi: 10.1145/3665065.3665071.

[23] N. Patel, "Sustainable Smart Cities: Leveraging IoT and Data Analytics For Energy Efficiency And Urban Development," *J. Emerg. Technol. Innov. Res.*, vol. 8, no. 3, pp. 313–219, 2021.

[24] H. S. Chandu, "Enhancing Manufacturing Efficiency : Predictive Maintenance Models Utilizing IoT Sensor Data," *IJSART*, vol. 10, no. 9, pp. 58–66, 2024.

[25] N. Moustafa, N. Koroniotis, M. Keshk, A. Y. Zomaya, and Z. Tari, "Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions," *IEEE Commun. Surv. Tutorials*, vol. 25, no. 3, pp. 1775–1807, 2023, doi: 10.1109/COMST.2023.3280465.

[26] Y. Rbah *et al.*, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems in IoMT: A survey," in *2022 2nd International Conference on Innovative Research in Applied Science, Engineering and*

Technology (IRASET), IEEE, Mar. 2022, pp. 1–9. doi: 10.1109/IRASET52964.2022.9738218.

[27] M. M. Belal and D. M. Sundaram, "Comprehensive review on intelligent security defences in cloud: Taxonomy, security issues, ML/DL techniques, challenges and future trends," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 10, pp. 9102–9131, Nov. 2022, doi: 10.1016/j.jksuci.2022.08.035.

[28] M. Bouke, A. Abdullah, N. Udzir, and N. Samian, "Overcoming the Challenges of Data Lack, Leakage, and Dimensionality in Intrusion Detection Systems: A Comprehensive Review," *J. Commun. Inf. Syst.*, 2024, doi: 10.14209/jcis.2024.3.

[29] S. Gavric, A. Stevanovic, N. Mitrovic, F. Netjasov, and J. Rakas, "Urban Surface and Air Mobility Control: A Microsimulation Integrating Intelligent Fly-Drive Vehicles Into Surface Traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 7, pp. 9893–9906, Jul. 2025, doi: 10.1109/TITS.2025.3565916.

[30] H. Gao *et al.*, "Driving Risk Assessment for Intelligent Vehicles Based on Entropy-Informed Graph Neural Networks and Gaussian Distributions," *IEEE Trans. Neural Networks Learn. Syst.*, pp. 1–14, 2025, doi: 10.1109/TNNLS.2025.3569826.

[31] H. Chai, K. Dong, Y. Liang, Z. Han, and R. He, "Machine learning-based accidents analysis and risk early warning of hazardous materials transportation," *J. Loss Prev. Process Ind.*, vol. 95, Jun. 2025, doi: 10.1016/j.jlp.2025.105594.

[32] A. Almutairi, A. F. Al Asmari, F. Alanazi, T. Alqubaysi, and A. Armghan, "Deep learning based predictive models for real-time accident prevention in autonomous vehicle networks," *Sci. Rep.*, vol. 15, no. 1, p. 20844, Jul. 2025, doi: 10.1038/s41598-025-04867-8.

[33] A. Goswami, M. K. Ranjan, Prince, R. Rikame, A. Patil, and A. M. Sattar, "Accident Severity Unveiled: How Weather, Roads, and Time Conspire in Machine Learning Predictions," in *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)*, 2024, pp. 1–6. doi: 10.1109/ARIIA63345.2024.11051606.

[34] S. Wu *et al.*, "Driving Risk Assessment Approach with Machine Learning under the Influence of Braking Rate," in *2024 12th International Conference on Information Systems and Computing Technology (ISCTech)*, IEEE, Nov. 2024, pp. 1–8. doi: 10.1109/ISCTech63666.2024.10845714.