Wiley
Journal of Advanced Transportation
Volume 2024, Article ID 3748345, 36 pages
https://doi.org/10.1155/atr/3748345

Review Article

Traffic Incident Duration Prediction: A Systematic Review of Techniques

Artur Grigorev, Adriana-Simona Mihaita, and Fang Chen

School of Computer Science, Faculty of Engineering and Information Technology, University of Technology Sydney, 61 Broadway St, Sydney, Australia

Correspondence should be addressed to Artur Grigorev; artur.grigorev@uts.edu.au

Received 31 March 2024; Revised 9 September 2024; Accepted 14 October 2024

Academic Editor: Long Truong

This systematic literature review investigates the application of machine learning (ML) techniques for predicting traffic incident durations, a crucial component of intelligent transportation systems (ITSs) aimed at mitigating congestion and enhancing environmental sustainability. Utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology, we systematically analyze literature that overviews models for incident duration prediction. Our review identifies that while traditional ML models like XGBoost and Random Forest are prevalent, significant potential exists for advanced methodologies such as bilevel and hybrid frameworks. Key challenges identified include the following: data quality issues, model interpretability, and the complexities associated with high-dimensional datasets. Future research directions proposed include the following: (1) development of data fusion models that integrate heterogeneous datasets of incident reports for enhanced predictive modeling; (2) utilization of natural language processing (NLP) to extract contextual information from textual incident reports; and (3) implementation of advanced ML pipelines that incorporate anomaly detection, hyperparameter optimization, and sophisticated feature selection techniques. The findings underscore the transformative potential of advanced ML methodologies in traffic incident management, contributing to the development of safer, more efficient, and environmentally sustainable transportation systems.

Keywords: artificial intelligence; intelligent transportation systems; machine learning; traffic accident; traffic incident classification; traffic incident duration prediction

1. Introduction to Traffic Accident Analysis

Today, artificial intelligence (AI) is being used to enhance the performance of various tasks, especially in the transport industry. AI technologies such as machine learning (ML) and deep learning (DL) models can be used to address transportation problems such as traffic management, urban mobility, and traffic safety. AI models are used to solve traffic prediction, traffic control, road safety planning, and traffic flow optimization problems [1, 2].

Traffic congestion, which arises in 60% of instances due to unforeseen events [3], poses a significant challenge for numerous urban centers globally. Various factors contribute to congestion, including population growth, the concentration of the workforce in central areas, and the absence of effective public transportation options. Generally, there are two primary types of traffic congestion: a) recurrent congestion, which occurs during peak hours when traffic demand surpasses road capacity, and b) nonrecurrent congestion, which results from unpredictable events such as car accidents, vehicle breakdowns, weather-related incidents, and more.

Intelligent transportation systems (ITSs) have become an essential component of urban transport networks in contemporary metropolises. By facilitating the observation and management of the transportation infrastructure, these systems not only increase travel safety but also augment overall transport network efficiency. Consequently, travel durations are minimized, air pollutant emissions are curtailed, and the city's economy and public health are notably improved.

The incorporation of AI techniques into the ITS has the potential to greatly reduce traffic congestion and its effects on the environment. The main data sources used by ITSs are as follows: vehicle detectors (magnetic, infrared, ultrasonic, and microwave), traffic cameras, Global Positioning Systems, and automatic vehicle identifiers (e.g., electronic toll collection, access control, and speed control) [4]. AI techniques were applied to these kinds of data previously [5, 6]. Multiple measures can be taken by ITS to reduce the impact of incidents (e.g., variable message signs (VMSs), toll roads, adaptive cruise control, adaptive traffic light control, and transport group priority management) [4].

Traffic incident management systems (TIMSs) collect data on traffic incidents, encompassing a multitude of factors that affect traffic incident duration. Accurate prediction of the total duration shortly after the incident occurs can potentially decrease operational costs and save end-users' time by influencing route planning decisions. Furthermore, the clearance time of accidents is inherently linked to persistent traffic congestion and a variety of external factors, each bearing differing levels of importance. As a result, it is essential to evaluate the significance of these incident-related factors to improve prediction accuracy. The majority of existing research in this domain has primarily concentrated on examining diverse ML models for specific road types, such as freeways or highways, and on distinct stages of the traffic incident duration, including clearance time, recovery time, and total duration [7]. However, there remains a notable absence of a comprehensive approach that can be universally applied to all road types, account for all accident categories, and remain relevant across various countries with unique driving behaviors.

Accurately predicting the total duration of a traffic incident shortly after it occurs is essential to saving operational costs and end-user time, as well as reducing traffic congestion. Understanding the importance of incident factors is key to improving the accuracy of predictions. In this study, we review the literature related to traffic incident duration prediction. Specifically, we discuss the challenges associated with each modeling step, the complexity of the task, and the most recent advances in this field, with a focus on the potential of ML for incident duration prediction.

1.1. Paper Structure. The paper organization is detailed as follows.

Section 1.2 presents the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) methodology that we have followed for our study, which has revised overall a total of almost 1300 papers on the topic of incident modeling, which have been further filtered and selected down to 200 final papers to provide a comprehensive structured analysis into current gaps and future research directions.

Section 4 gives an overview of all the required datasets that one needs to conduct a thorough incident modeling which ranges from accident logs, but also to traffic states such as flow, speed, occupancy, and external related information (weather, events, etc.). We also provide insights into public datasets that have been used for modeling, as many countries restrict access to such datasets due to privacy concerns.

Section 5 describes methods of statistical analysis for traffic accident modeling. Section 4 is devoted to the use of ML in traffic accident analysis including classification and regression tasks, feature selection, imbalanced dataset management techniques, anomaly detection, dimensionality reduction, novel ML methods, and frameworks. Finally, in Section 5, we provide a summary of the challenges we have detected as well as future research gaps to be filled.

1.2. Literature Review Material and the PRISMA Method. The PRISMA [8, 9] is the review method that has been applied when organizing and analyzing literature for this study. The PRISMA method is designed to disclose all elements of the review process, from the identification and selection of studies to the synthesis of results (search strategy, criteria for inclusion and exclusion of studies, and analysis process) in order to reduce the risk of bias in performing a literature review. The process of reviewing the literature is shown in Figure 1. In the first stage, relevant literature has been identified by using publication databases based on keyword search (see Table 1). The list of used keywords is as follows.

Since traffic incidents can be studied across different research areas, it is necessary to always specify the "traffic" to narrow the search result to the relevant area. The word "traffic" can also be referred to as a "molecular transfer process" from biology. Keywords can include "clearance time" since this term is very specific to the task of incident duration modeling. The keyword "machine learning" is the main method used in incident duration prediction. The use of "traffic incident prediction using random forest (RF)" is clearly related to the tasks of classification and regression related to the traffic incident duration modeling. By using very specific terms and specifying area, it is possible to locate relevant literature quickly.

The process of creating search queries using linguistic and logical perspectives involves various transformation techniques. Here is a concise explanation of each type of variation and their technical aspects:

- Synonym: Replaces words with similar meanings, e.g., "Traffic Incident" to "Traffic Accident," capturing articles with alternative terms.
- Specification: Refines queries by adding specifics, e.g., "ML for traffic incident duration estimation" instead of "ML," targeting more relevant articles. Another example is "ML for Incident Duration Prediction" and "XGBoost for Accident Clearance Time Prediction."
- Generalization: Broadens queries by using general terms or removing constraints, e.g., "AI techniques in traffic incident management" instead of "Traffic Incident prediction using AI," covering related topics.
- Contextualization: Queries regarding a specific context, e.g., "PRISMA method in traffic incident duration prediction research," allow finding articles with similar research approaches or contexts.
- Focus shift: Alters queries by shifting focus to a different aspect, e.g., "Traffic incident impact estimation" instead of "Traffic Incident duration prediction," exploring subtopics or aspects of the research area.

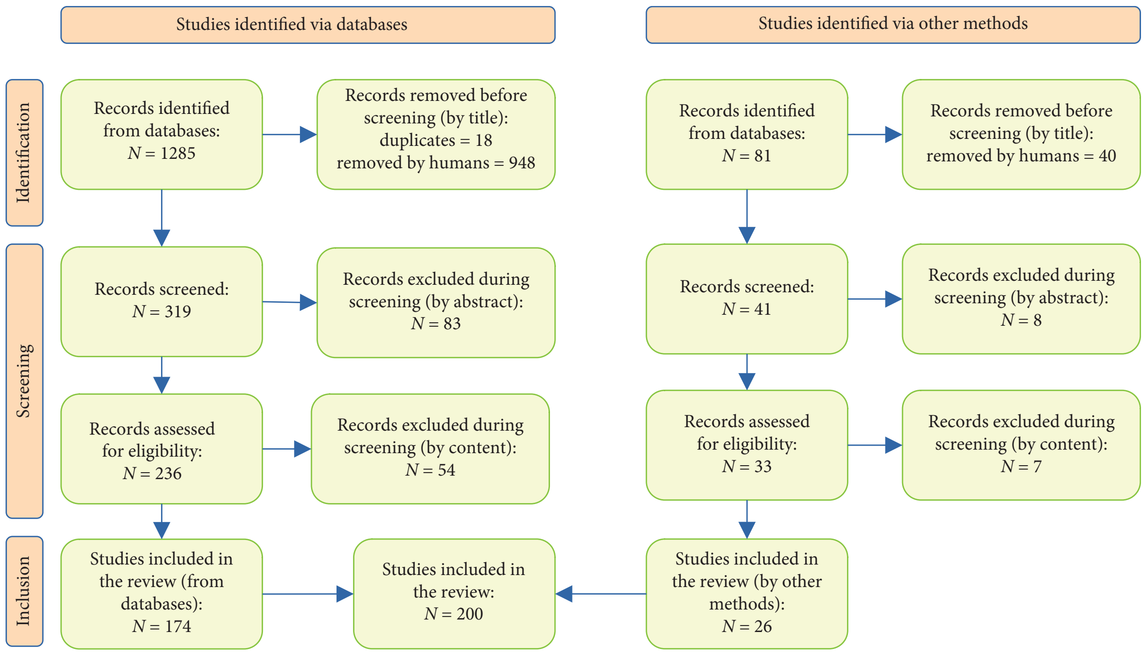

Figure 1: Flow diagram for systematic review based on the PRISMA approach.

Table 1: Search queries and the corresponding transformation techniques.

No.	Search query	Technique used
1	Traffic incident duration prediction	Base query
2	Traffic incident clearance time prediction using machine learning	Specification
3	Traffic incident clearance time prediction using AI	Specification, generalization
4	Traffic accident duration prediction	Synonym
5	Deep learning models for traffic incident analysis	Specification, focus shift
6	Random forest for traffic incident duration prediction	Deeper specification of the first term
7	Traffic incident description analysis in congestion mitigation	Double focus shift
8	PRISMA method in traffic incident duration prediction research	Contextualization
9	Literature review on traffic incident duration prediction	Contextualization, generalization
10	AI in traffic congestion reduction	Generalization, focus shift
11	Traffic accident duration prediction using methods of AI	Specification

Making search queries using these variations maximizes the chances of finding relevant literature. Combining different techniques uncovers a wider range of articles, ensuring a comprehensive understanding of the research area.

The following databases were used for the literature identification:

- ScienceDirect
- Google Scholar
- Research Gate

The alternative source of information on the relevant literature is the Connected Papers service, which builds a graph of studies based on their semantic similarity (e.g., papers on a similar topic will have a stronger connection than less similar papers). This service requires a sample paper as a starting point to search for similar ones. The search using this approach was performed after the identification of relevant literature using common databases.

The databases were accessed through the University of Technology Sydney, and the publications were limited between 1980 and 2022. In total, 1378 sources were collected, 1285 were found using conventional databases, and 81 using Connected Papers.

The PRISMA process for this literature review is detailed in the following:

1. A database of records has been collected using previously described databases and keywords, resulting in 1285 resources.
2. Resources were screened for duplicates, which resulted in 18 resources being removed. Then, records were filtered by paper title, resulting in the removal of 948 entries, mainly due to interference with topics of "internet traffic" (which also relies on the use of ML methods and network analysis) and "molecular traffic," as well as the presence of many studies related to injury statistics and safety analysis related to traffic accidents.

3. Filtered resources (319 in total) were screened by abstract, which resulted in the removal of 83 records due to the mention of unrelated methodologies and findings.
4. Eligible records (236 in total) were then screened by content (reading of methodology and conclusion sections), resulting in the removal of 54 records.
5. In total, we obtained 174 records from the database search.
6. The most relevant review (Ruimin Li, F. Pereira, M. Ben-Akiva, "Overview of Traffic Incident Duration Analysis and Prediction") from the previous search was selected, and a graph of related papers was built (see Figure 2). A similar process was performed for a semantic similarity search using the Connected Papers service, which resulted in 26 additional papers being selected.
7. In total, we obtained 200 relevant resources for the literature review. The highest number of papers is dated between 2010 and 2022, with peaks in 2002, 2013, 2016, and 2018–2021 (see Figure 3). Peaks during these years can be attributed to the introduction of novel ML methods (e.g., RF in 2002 [10] and XGBoost in 2016 [11]).

2. Overview of Literature Reviews on Traffic Incident Duration Prediction

Efficient traffic management in ITSs relies heavily on accurately predicting the duration, impact, and severity of traffic incidents. This review examines current research on predicting traffic incident duration, focusing on statistical and ML methodologies.

The reviewed studies share common objectives: to systematically evaluate and compare existing statistical and ML methodologies for traffic incident duration prediction; to identify key factors influencing incident duration and understand their impact on model performance; to analyze data requirements, limitations, and potential integration opportunities for improving prediction accuracy; and to provide recommendations for future research and development in the field.

Tang et al. [12] undertook a comparative analysis of traditional statistical methods, namely, accelerated failure time (AFT), quantile regression (QR), finite mixture (FM), and Random Parameters Hazard-Based Duration (RPHD), with modern ML techniques such as K-nearest neighbors (K-NN), support vector machine (SVM), back propagation neural network (BPNN), and RF. The study highlighted the strengths of statistical models in handling unobserved heterogeneity, while acknowledging the robust performance offered by ML approaches. The authors suggested exploring hybrid models, which combine both statistical and ML methodologies, as a promising avenue for achieving optimal accuracy. The key takeaway from this review was the realization that hybrid approaches might offer improved performance. Statistical models like AFT and RPHD were better at handling unobserved heterogeneity, whereas ML models such as SVM and RF demonstrated robust predictive performance.

Al-Salman et al. [13] provided a more extensive overview of traffic incident duration prediction methodologies. Their review covered a wide variety of techniques, including regression models, stochastic models, decision trees (DTs), neural networks, hazard-based duration models, SVM, Bayesian networks, and hybrid models. This comprehensive approach allowed them to conclude that no single method is universally superior, emphasizing the importance of context-specific selection and hybrid models. They observed that different methods excel in specific contexts, and integrating these methods through hybrid models holds significant potential for improved performance.

Valenti et al. [14] narrowed their focus to comparing the performance of five specific models, namely, multiple linear regression, DT, artificial neural network (ANN), relevance vector machine (RVM), and K-NN, across various incident duration ranges. They found that each model displayed strengths within specific duration ranges, advocating the use of a preliminary classification scheme based on anticipated incident duration to select the most suitable model. Specifically, RVM achieved the highest overall accuracy, suggesting that model selection should take into account the anticipated incident duration. Based on this observation, the authors recommended implementing a preliminary classification system that would guide the selection of the optimal model for a given incident.

Li et al. [7] traced the evolution of methods, highlighting data limitations and future research directions. They discussed various statistical, HBDM, DT, ANN, SVM, and hybrid methods. The authors found that data integration, handling outliers and unobserved factors, and developing time-sequential models are crucial. A significant finding was the necessity for integrating diverse data sources and handling outliers and unobserved factors. Li et al. highlighted the importance of developing time-sequential models, which can dynamically update predictions as new data become available. They recommend integrating data sources, exploring advanced ML, addressing unobserved heterogeneity, and developing time-dependent models.

Hamad et al. [15] provided a detailed comparative analysis of eleven ML classifiers for predicting freeway incident durations. Their evaluation focused on accuracy, computational complexity, and the impact of feature selection on model performance. The classifiers evaluated included SVM, K-NN, Gaussian process classification (GPC), DT, RF, logistic regression (LR), Naive Bayes (NB), discriminant analysis (DA), multilayer perceptron (MLP), stochastic gradient descent (SGD), and bagging/boosting ensembles. Their findings revealed that SVM, K-NN, and GPC achieved the highest prediction accuracy at 97%. Notably, SVM exhibited a favorable balance between accuracy and computational complexity. The authors recommended exploring hyperparameter optimization techniques, hybrid models, and real-time incident detection using mobile data as promising directions for future research. Authors recommended further exploration of hyperparameter optimization, hybrid models, and real-time incident detection using mobile traffic data to enhance prediction capabilities.

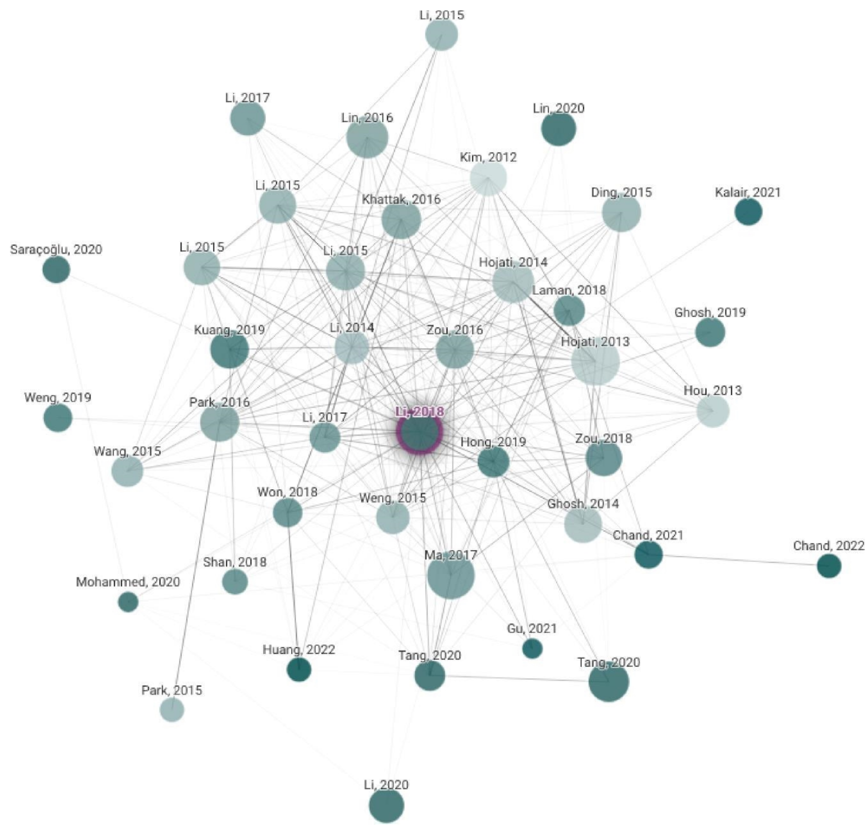

Figure 2: Connected Papers graph example.

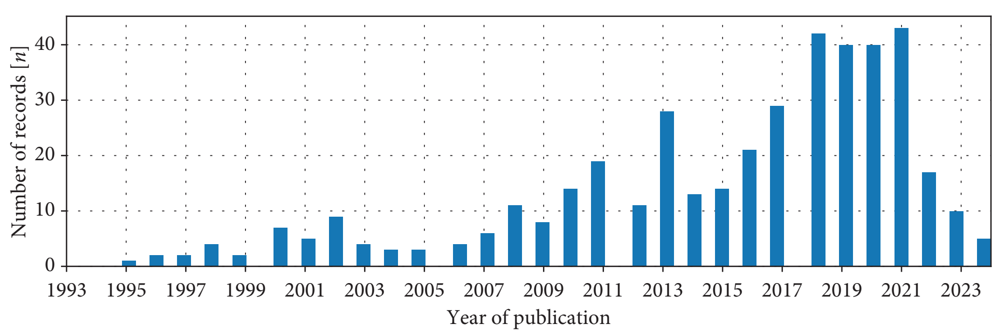

Figure 3: Reviewed publications grouped by year.

Future research directions include the development of hybrid models that combine statistical and ML methodologies, the integration of real-time data from diverse sources, the development of models capable of capturing unobserved heterogeneity and factors influencing duration, and the outlier analysis and incorporation of explainable AI to interpret results of ML models (including feature importance estimation methods and dimensionality reduction with the purpose of visualization).

Tang et al. [12] and Al-Salman et al. [13] both emphasized the potential of hybrid models. Valenti et al. [14] and Li et al. [7] underscored the importance of model selection based on expected incident duration and highlighted the need for advanced data integration techniques. Hamad et al. [15] provided a comprehensive evaluation of ML classifiers, recommending real-time data integration and consideration of hyperparameters for model optimization.

The selection and utilization of the most influential features as well as model parameters are critical for enhancing model accuracy and managing computational complexity. This adaptation makes traffic incident duration models more accurate, but makes them also highly specific to the dataset, which may hinder the transferability of the prediction models across different traffic networks.

3. Concepts and Definitions in Traffic Incident Analysis

3.1. Traffic Incident Duration Definition and Phases. Traffic congestion can be recurrent and nonrecurrent [16]. Nonrecurrent traffic congestion is unexpected congestion, caused by random events affecting traffic flow such as traffic incidents, weather phenomena, vehicle breakdowns, and hazards. Recurrent traffic congestion is a predictable regularly occurring congestion, which observed in places where traffic flow regularly exceeds road capacity [17].

Traffic incident duration, as defined in the Highway Capacity Manual [18], comprises four distinct phases:

- Incident Detection: the time interval between the occurrence of an incident and its subsequent reporting.
- Incident Response: This refers to the span of time between the reporting of the incident and the moment the first investigator arrives at the scene of the incident.
- Incident Clearance: This is the amount of time from the arrival of the primary investigator at the scene until the incident area is completely cleared.
- Incident Recovery: This is the duration between the clearing of the incident scene and the point when traffic flow returns to its normal condition.

Determining the precise point in time when a traffic incident occurs can be challenging due to the availability and accuracy of the data. Two potential options to consider are as follows: 1) the time when the car crash happened or 2) the time when traffic state decline began being registered. Each of these options has its drawbacks:

- Time when the car crash happened: This option refers to the exact moment when the vehicles involved in the accident collided. Choosing this point in time provides a clear starting point for measuring the incident's impact and the time it takes for the entire process, from detection to recovery. However, determining the exact time of the crash can be difficult, as it often relies on witnesses or participants reporting the time accurately. Additionally, in some cases, there might be a delay between the time of the crash and the time when traffic accident starts to be registered (some authorities record the time of accident notification as an accident start time).
- Time when traffic speed decline started being registered: This option focuses on the time when the incident's impact on traffic flow becomes apparent through a noticeable decline in traffic speed. This approach can be more easily determined using traffic sensors or cameras, which continuously monitor traffic conditions. Furthermore, it may better represent the actual impact of the incident on traffic flow. However, one potential drawback is that this approach may not account for the time it takes to detect the incident and initiate the response process.

Response time (RT) is defined as a time interval comprised of both response team preparation time and travel time to the incident site. RT was modeled in [19]. In another study, recovery time was analyzed on freeway segments in Southeast Queensland (Australia) [20] and was derived using historical loop detector data and traffic incident characteristics at the time and location of the incident. The event of nonrecurrent traffic congestion was detected based on the allowable percentage of speed decrease. The time interval of the incident was determined by forward and backward searches in time for time intervals, when traffic speed was unaffected, which appear to be bounding for the traffic incident.

Research published in [21] includes the calculation of recovery time from the time required for the restoration of travel time during the affected traffic state to travel time during the normal traffic state. That research points to the possibility to use traffic flow data to uncover more precise traffic incident duration instead of relying on the definition given by the response team. As pointed out in Figure 4, we can derive the duration of selected phases of traffic incidents from traffic speed data (or possibly traffic flow data). We also want to highlight that guidelines of incident timeline reporting should be taken into consideration when modeling traffic incidents. The reported incident start and end time can mismatch the actual impact of the accident on the traffic speed. Dataset providers may not take into consideration time intervals of the incident detection and recovery simply providing the time interval between the report received and the incident cleared message as the total incident duration. This difference in reporting may affect the accuracy of traffic accident impact prediction models. Some incidents may be reported much later than they occurred resulting in very short (0–5 min) reported accident durations.

Understanding these distinct phases is critical for effectively managing and mitigating the impact of traffic incidents. Each phase may be influenced by various factors, and accurately predicting the duration of each phase can help reduce the overall negative impact of incidents on traffic flow, road users, and the environment. By thoroughly examining these phases and incorporating them into prediction models, transportation authorities and incident management systems can better allocate resources, prioritize responses, and minimize the consequences of traffic incidents.

In conclusion, traffic incident duration consists of multiple phases. Nevertheless, data availability for the duration of such phases is rare. Duration of phases and total traffic incident duration (assumable, even more correct than recorded by response teams) can be extracted using traffic flow data assuming the data streams are reliable and free of anomalies or outliers that might affect precision and analysis.

3.2. Characteristics of Traffic Incidents. Features used by research studies on traffic incident duration are very diverse. Some researchers did not perform possible feature manipulations despite data availability (e.g., AM/PM peak hour

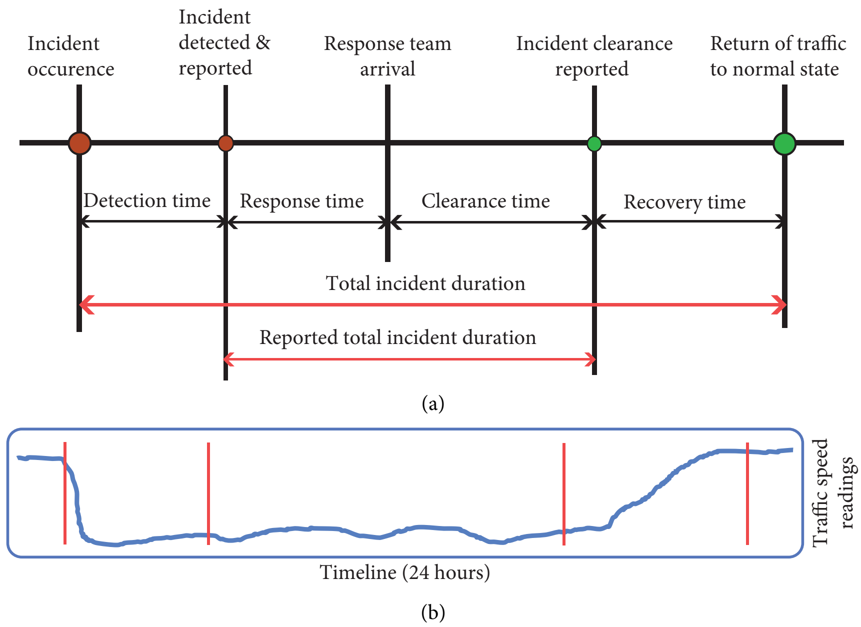

FIGURE 4: Traffic incident duration representation: (a) Timeline segmentation. (b) Example of speed readings on the day of traffic accident.

binary feature or nighttime). Traffic incident research has the possibility to benefit from comprehensive feature derivation based on time, weather, and road network data. By tracking the use of features and assessment of their significance, we can make decisions on the concentration of future efforts in data processing. Also, we can develop strategies for feature extraction considering techniques used by other researchers.

For example, in [22], authors use the feature "season" which is represented as a set of four discrete values (summer, winter, spring, and autumn). The season can highly affect safety on the road due to weather effects: winter storms, ice on the road, rain showers, and other environmental effects (which affect visibility, control, and safety) linked to the time of the year. But in some papers [23], authors do not use these features. Rainfall intensity was found highly relevant for the prediction of accident severity in Seoul, South Korea [24].

Roadway geometry is one of the critical factors which can affect the capability of a road system to withstand the incident impact [4]. Various features are used to define the road structure and use ML models: road segment length, road segment centroids, gradient, curvature, and general road density surrounding the event area [25]. A comprehensive review of spatial network (i.e. road network) theory applications provides an in-depth analysis of graph theory indices including betweenness centrality, ringness, route factor, detour index, and alpha index [26]. All of these indices can be calculated for road networks and incorporated into ML pipeline. Accident risk is found to be increasing when traffic speed slows down while traffic density goes up for Yingtian Expressway [27] which highlights the importance of speed-density diagrams for incident-related traffic state visualization. In that research, no correlation was observed between traffic flow and crash risk. Various road alignment factors can include curve, tangent, vertical grade, superelevation, horizontal alignment, and curve length [24].

Identification of road geometries is important for the analysis of incident occurrence. The impact of variables associated with crash frequency was found to be varying across parts of Tennessee, USA [28]. The spatial analysis demonstrated that segment length and median segment width had the highest impact on crash frequency in eastern regions, while commercial land use had the highest connection to crash frequency in southern regions. Multiple studies have indicated that patterns and dependencies in the spatial and temporal dimensions are likely to exist, often represented as clusters or hot spots [29].

Tables 2 and 3 include incident characteristics used among different research studies, with citing specific ones, unobserved or rarely observed in papers. For the general set of characteristics, readers can refer to [7, 23].

4. Datasets and Data Availability

In recent years, the availability of traffic accident data has significantly enhanced our understanding of the factors that contribute to these incidents and their outcomes [34–36]. These data include details such as the location and time of the accident, the types of vehicles involved, and the extent of the incident effects, providing insights into both the causes and consequences of traffic accidents. By applying ML techniques, researchers can predict the likelihood of future crashes, classify accidents by their severity, and estimate the duration of incidents. These predictive capabilities are essential for developing strategies to reduce the risk and impact of accidents.

4.1. Traffic Incident Duration Datasets. The availability of comprehensive, publicly accessible datasets on traffic incidents plays a significant role in the field of traffic safety

Table 2: Table of temporal features used to describe traffic incident.

Feature	Values	References
Peak hour	{0, 1}	—
Weekday	{0, 1}	—
Weekend	{0, 1}	[30]
Season	{winter, autumn, summer, spring}	[22]
Time of day	{0..23}	[23]
Peak hours	{Offpeak/AMpeak (6 – 9AM)/PMpeak (3 – 6PM)}	[31]
Daytime	{Evening, night – time}	[31]

Table 3: Table of features used to describe traffic incident across different studies.

Feature	Values	References
Incident type	{Vehiclefire, outofgas, breakdown, etc.}	[17]
Number of vehicles involved	$\{1, \ldots, N\}$	—
Multiple vehicles involved	{0, 1}	[20]
Type of vehicle involved #1	{Motorcycle, Van, Pickup}	[17]
Type of vehicle involved #2	{Largevehicle}	[19]
Type of vehicle involved #3	{Truck}	[32]
Location of incident on the road	{Forfreeways: ramp, left/rightshoulder}	[17]
Number of lanes	$\{1..N\}$	[20]
Link capacity	$\{N\}$	[20]
Average speed at the time of incident	{}	[20]
Number of affected lanes	$\{1..N\}$	[23]
All lines affected	{0, 1}	[19, 30]
Incident severity	$\{1..N\}$	[23, 33]
Lighting condition	{day, night}	[33]
Secondary crash	{0, 1}	[33]
Fire, injury	{0, 1}	[19]
Fatality	{0, 1}	[20]
Traffic disrupted	{0, 1}	[20]
Traffic flow on adjacent lanes	$\{N\}$	[23]
Medical required	{0, 1}	[20]
Rollover	{0, 1}	[32]
Weather #1	{Windy, Clear, Rain}	[18]
Weather #2	{Sunny, Cloudy, Storm}	[32]
Weather #3	{Rain, Snow, Wind, Fog}	[31]
Position within road	{Inner, Outer, Middlelane}	[32]
Lane number	$\{1..N\}$	[23]
Other features	**Values**	**References**
Distance from the city center	{Rkm}	[20, 23]
Traffic condition	{congested, uncongested}	—

analysis. These datasets, curated and maintained by various national and international organizations, provide detailed records of traffic accidents, including information on crash locations, vehicle types, and the severity of incidents. Open access to such data enables collaboration among researchers worldwide, facilitating reproducible research. There are numerous publicly available datasets of traffic incident reports that specifically contain the incident duration variables, or where duration can be estimated from the start and end times of the incidents.

- Countrywide Traffic Accident Dataset (CTADS) is one of the biggest datasets on traffic accidents recently released in 2021 [37, 38], which contains 1.5 million accident reports collected for almost 4.5 years since March 2016, each report containing 49 features obtained from MapQuest and Bing services. This dataset was used previously to predict the accident duration [39, 40]. Key features include the severity (rated 1–4) indicating the impact on traffic, and start time and end time marking the duration of the incident. The location is detailed through GPS coordinates, along with address-related information such as street, city, state, and zipcode. Weather conditions at the time of the accident are captured through attributes like temperature, humidity, visibility, and weather condition class. The dataset also notes nearby points of interest (POIs) such as traffic signals, junctions, and railways. Additionally, twilight conditions, like sunrise or sunset and civil twilight, indicate the time of day during the incident.
- Caltrans Traffic Performance Measurement System (PeMS) [41] dataset contains incident reports with a timeline of events and description of the incident (as

a sequence of abbreviations) as it becomes available and status updates from a dispatch unit. Also, the dataset contains 5-min aggregated traffic speed, traffic flow, and traffic occupancy records as well as vehicle detector status. The availability of these data allows for analyzing traffic incidents in conjunction with vehicle detector data. A brief description of the incident includes location, area, start time, duration, and freeway ID.

- Road Traffic Incidents of Sydney Greater Metropolitan Area (RTI Sydney GMA) dataset [42] detailed in the paper titled "Data on road traffic incidents for Sydney Greater Metropolitan Area," [42], offers a comprehensive overview of road traffic incidents (RTIs) within the Sydney Greater Metropolitan Area (GMA), New South Wales, Australia, covering the period from January 1, 2017, to July 31, 2022. This dataset includes a total of 85,611 incidents, comprising crashes and vehicle breakdowns. A distinguishing feature of this dataset is the inclusion of specific data on the duration of each incident (RTI duration) and statistical data across local governmental areas, which are divided into 333 zones based on Statistical Area Level-2 (SA2) classifications. The dataset was sourced from the Open Data Hub, Transport for New South Wales (TfNSW), Australia. The dataset is linked to various independent variables such as road network characteristics, public transport network characteristics, socioeconomic data, and travel characteristics at the SA2 level.

In the field of road safety analysis, several key datasets provide comprehensive statistical and casualty information on motor vehicle traffic incidents.

- National Highway Traffic Safety Administration's (NHTSA) Fatality Analysis Reporting System (FARS): This dataset contains detailed information on fatal motor vehicle traffic crashes in the United States that occurred since 1975 [43].
- National Transportation Atlas Database (NTAD) from the United States Department of Transportation's Bureau of Transportation Statistics (BTS) [44] contains detailed information on nonfatal motor vehicle traffic crashes in the United States since 1994.
- European Commission's Road Safety Atlas [45] provides accident statistics for each European country using interactive maps and satellite images.
- UK Road Safety Statistics [46]: This dataset contains detailed information on fatal and nonfatal road traffic accidents in the UK since 1979.
- California Highway Patrol (CHP) Statewide Integrated Traffic Records System (SWITRS) [47]: A California-wide dataset containing detailed information on motor vehicle collisions reported to CHP. Accident report details contain data on the location, severity, road condition, and victim data including age and degree of injury. Due to the extensive timeline and precision of reporting, this dataset was previously used to analyze the effect of country-scale events on crash severity [48].
- Australian Road Deaths Database (ARDD) [49] provides basic details of road traffic crash fatalities in Australia as reported by the police each month to the State and Territory road safety authorities. The dataset includes information on fatal crashes: year, month, day of the week, time, location, crash type, and vehicle type involved.
- World Health Organization's Global Health Estimates [50]: This dataset contains detailed global estimates on road traffic injuries, deaths, and disability-adjusted life years from 1990 to present.

4.2. Supplementary Data Sources on Traffic Flow for Traffic Incident Duration Analysis. In the domain of traffic incident analysis, the integration of alternative data sources has become increasingly important for studying accident dynamics. Beyond conventional traffic accident reports, a variety of datasets now provide detailed information on traffic conditions such as traffic speed, occupancy, and flow across different roads and lanes over time. These datasets, characterized by high granularity and real-time data availability, allow for a comprehensive analysis of traffic flow patterns, vehicle behaviors, and incident impacts in ways that were previously not feasible.

By combining these alternative data sources with traditional accident reports, researchers can establish connections between incident characteristics and their specific impact on traffic flow. Key examples of such datasets include the PeMS dataset and connected vehicle data from CompassIoT and TomTom:

- CompassIoT [51] is a dataset and API of connected vehicle road trips across Australia which aggregates data from 64 different manufacturers across a number of global car brands. It started from 200,000 connected vehicles in 2018 to over 2.2 million trips and billions of data points in 2023. The dataset has low data latency, where real-time API can yield data every 5 s. Data on braking acceleration and steering can allow identification of dangerous road conditions and risky behavior characteristic to road accidents.
- TomTom [52] is historical traffic database with information on road speeds, travel times, and traffic density. Allows customized queries for route and area analysis providing statistics on travel time and speed.

Main factors contributing to traffic jams are high traffic flow (e.g., during peak hours), road network bottlenecks (a spatial aspect of road geometry), and local disturbances in the traffic flow (e.g., actual traffic incidents which act as a trigger of traffic jam) [53]. An emerging approach is to use computer vision methods to detect traffic incidents using CCTV cameras. A spatial–temporal near-accident condition detection system has been recently proposed, which utilizes object detection, segmentation, and tracking [54]. The use of computer vision methods with road surveillance cameras can be effectively used to estimate the traffic incident timeline at the intersection.

Advanced driver assistance systems are intended to mitigate or prevent crashes by providing vehicle drivers with the necessary information to avoid collisions. To accomplish this task, the truck driver behavior encountering vulnerable road users (e.g., cyclists at intersections) has been studied [55].

4.3. Supplementary Data Sources on Road Geometry for Traffic Incident Duration Analysis. Roadway geometry significantly influences how quickly an incident can be managed and cleared. For example, if an accident occurs on a narrow road or a section with limited shoulder space, emergency responders may have difficulty reaching the scene. Areas with a high frequency of crashes often reflect underlying issues with roadway design. By analyzing accident report data in conjunction with roadway geometry, predictive models can identify patterns that help forecast how they might affect the duration of the incident [56]. For instance, a DT model might predict that incidents on a steep curve are likely to take longer to clear than those on a straight, flat road.

Various road geometry features can be used to predict the traffic accident risk since they affect the driver's behavior. Roadway geometry also directly affects the likelihood of accidents [57]. Features such as sharp curves, narrow lanes, steep grades, and complex intersections can increase the probability of accidents by creating challenging driving conditions [58]. When incidents occur in areas with such challenging geometry, they may lead to more severe traffic disruptions and longer incident duration.

- OpenStreetMap (OSM) [59] is a collaborative project that provides a freely accessible and detailed global map, including comprehensive information on road geometry. The OSM dataset includes data on road types, lane configurations, road widths, intersections, roundabouts, and other critical geometric features. This dataset is widely used for mapping, route planning, and integrating road geometry into traffic analysis and modeling [60]. The data are available in various formats such as XML and PBF (Protocol Buffer Binary Format) and can be converted into GIS-compatible formats like shapefiles or GeoJSON, making it a versatile resource for a range of applications in traffic and urban studies.
- The OS National Geographic Database (NGD) [61] Transport Theme provides a comprehensive network dataset and topographic depiction of Great Britain's roads, railways, tracks, and paths. This dataset integrates Ordnance Survey's large-scale road and path content with authoritative routing information from the National Street Gazetteers (NSGs) and the Trunk Road Street Gazetteer (TRSG). The OS NGD Transport Theme is organized into three main collections: Routing and Asset Management Information (RAMI), Transport Features, and Transport Network. These collections encompass 35 different feature types, providing detailed and structured data for transportation analysis.

While recent studies have largely focused on reported incident parameters as noted in [23, 62], road topology is an essential factor in estimating incident probability, with poorly designed junctions and inappropriate speed limits contributing to increased risk. In fact, according to [63], approximately 5% of road junctions account for 50% of accidents within the city of London. Therefore, it is crucial to analyze incident duration and probability by incorporating road topology into consideration. Geographically weighted Poisson regression (GWPR) models were used to model frequencies of harsh driving behavior events, which were found to be positively correlated with segment length and presence of traffic lights and negatively with neighborhood complexity (which is a density area in proximity of the event) [25].

5. Predictive Incident Duration Modeling

In the context of traffic incident analysis, the incident duration modeling task is formulated as a regression problem, where the goal is to predict the actual duration of a traffic incident based on its features.

Given a dataset where each traffic incident is represented by a feature vector $\boldsymbol{x}_i$, which includes N_f features, the task is to predict the incident duration y_i, a continuous variable. This can be modeled as

$$\hat{y}_i = f(\boldsymbol{x}_i; \theta), \tag{1}$$

where

- $\hat{y}_i$ is the predicted duration of the traffic incident.
- $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{iN_f}]$ is the feature vector for the i-th incident.
- θ represents the parameters of the predictive model f (e.g., coefficients in a linear regression model, weights in a neural network).

The objective of the regression task is to minimize the difference between the predicted duration $\hat{y}_i$ and the actual duration y_i. This difference is typically measured using a loss function $L(y_i, \hat{y}_i)$, such as mean squared error (MSE):

$$L(\theta) = \frac{1}{N_i} \sum_{i=1}^{N_i} \left(y_i - \hat{y}_i\right)^2, \tag{2}$$

where

- N_i is the total number of traffic incidents in the dataset.
- y_i is the actual duration of the i-th traffic incident.
- $\hat{y}_i$ is the predicted duration of the i-th traffic incident.

The model $f(\boldsymbol{x}_i; \theta)$ is trained to find the optimal parameters θ that minimize the loss function $L(\theta)$, thereby enhancing the accuracy of the duration predictions.

In addition to predicting the point estimate of the incident duration, it is often useful to model the distribution of traffic incident duration. This approach provides a probabilistic understanding of the expected duration, which can be critical for risk assessment and decision-making in traffic management.

One common method to model the distribution is to assume that the incident duration follows a certain probability distribution, such as a Gaussian (normal) distribution, a log-normal distribution, or a Weibull distribution. The choice of distribution depends on the observed characteristics of the data.

For instance, if the incident duration is positively skewed, a log-normal distribution might be appropriate as follows:

$$y_i \sim \text{LogNormal}\left(\mu_i, \sigma_i^2\right), \tag{3}$$

where

- μ_i and σ_i^2 are the mean and variance of the logarithm of the duration, respectively.

Alternatively, if the incident durations exhibit a heavy-tailed distribution, a Weibull distribution might be used as follows:

$$y_i \sim \text{Weibull}\left(\lambda_i, k_i\right), \tag{4}$$

where

- λ_i is the scale parameter.
- k_i is the shape parameter.

In this case, the modeling task involves estimating the parameters λ_i and k_i based on the incident features $\boldsymbol{x}_i$. These parameters can be linked to the features through regression models, allowing for the prediction of the entire distribution of incident durations rather than just a single-point estimate.

Modeling the distribution of durations provides more comprehensive information than point estimates alone, enabling traffic management systems to assess the likelihood of various duration outcomes and plan accordingly. This probabilistic approach is especially useful for dealing with the inherent uncertainty in predicting traffic incident durations.

Duration of detection, response, and clearance phases can modeled separately in the literature by using hazard-based duration modeling [31]. Researches in [22] focused on the use of multiple types of distributions (Log-normal, Gamma, etc.) for the four-time intervals within the incident duration structure corresponding to: response team preparation time, response team travel time, incident clearance time, and total incident duration time. They found the importance of different distributions to approximate different incident duration stages.

5.1. Traditional Approaches for Traffic Incident Duration Modeling. Traffic incident duration modeling has been explored using a variety of statistical models. Early studies such as Sullivan (1997) [64] used a log-normal distribution for their models. However, recent research trends [65, 66] indicate a shift toward using the log-logistic distribution. This preference is primarily due to the greater degree of goodness of fit provided by the log-logistic model in comparison with the log-normal distribution.

Several recent studies have experimented with multi-component log-logistic models. For instance, a g-component log-logistic model was proposed by [67], and a competing risks mixture model that incorporates a multinomial log-logistic model was presented in [68].

The challenge lies in the fact that actual incident duration distributions only offer approximate information. Moreover, it is significantly influenced by various parameters like the conditions of the day or night [69], the chosen method of incident clearance [68], weather conditions, traffic density, time periods, and incident location [70]. These parameters contribute differently to various types of accidents and uniquely impact traffic incidents such as rear-end collisions, sideswipes, collisions with fixtures, and rollovers. Consequently, estimating the distribution of incident duration is a complex task, heavily reliant on numerous traffic flow and incident variables.

Researchers have also used hazard-based models for traffic incident duration analysis [20, 31]. These models employ a hazard function to estimate the conditional probability of an incident ending within next time interval, assuming it lasted until the beginning of that interval. For example, the probability that an incident may end may be very high considering it lasted multiple hours on a highway during sunny day. Multicomponent log-logistic models have also been explored, such as the g-component log-logistic model [67] or the competing risks mixture model incorporating a multinomial log-logistic model [68].

A recent study by [33] employed Weibull, log-normal, and log-logistic distribution for modeling clearance and impact duration. The incident impact duration was modeled on historical speed data from BlueTOAD devices rather than durations reported by response teams, offering potentially greater accuracy. The study also revealed the effects of certain parameters (nighttime, severity, EMS involvement, etc.) on both durations.

Given the varied modeling accuracy for impact and clearance duration [18] and the observed differences between distinct incident types, it seems prudent to model durations separately for each incident type. This would allow for a more nuanced understanding of how different models perform in each context. Furthermore, given that different distributions can approximate different phases of traffic incident [22], it might be beneficial to fit distinct distributions to the dataset divided by specific variables (peak hour, etc.). This could result in a finer understanding of feature importance across different datasets.

5.2. ML for Incident Duration Prediction. Various ML methods have been utilized to predict incident duration. These techniques encompass tree-based methods as discussed in [71], fuzzy logic systems explored in [72], Bayesian networks analyzed in [73], and ANNs as presented in [74, 75]. More recently, [76] investigated the gradient boosting decision tree (GBDT) as an improved method for incident duration prediction. Additionally, it has been found that Gaussian process regression and ANNs surpass tree methods and SVM in predicting incident duration, as demonstrated by [62]. By utilizing recent advancements in the field of accident analysis, the examination of road topology, spatial relations, and advanced ML techniques can lead to better insights into incident duration prediction and risk management.

Incident factors could include a range of variables such as the location of the incident, the time it occurred, weather conditions, traffic density, road topology, vehicle types involved, and the nature of the incident (e.g., collision, breakdown). By understanding which of these factors are most important, it is possible to improve the model's performance, making predictions of incident duration more accurate.

Overall, understanding the factors that influence incident duration and contribute to the probability of the accident can help improve traffic management systems and enhance road safety. Therefore, exploring different ML methods to predict incident duration is crucial to reduce traffic congestion and improve overall traffic flow. For instance, if it is known that incidents occurring at poorly designed junctions typically last longer due to complex traffic dynamics, this factor can be given more weight in the ML model. Similarly, if certain weather conditions are found to exacerbate incident durations, these conditions can be detailed in higher amount of attributes (e.g., a cloud coverage from a few levels can be extended into actual percentage, and presence of strong wind can be detailed into wind direction and wind speed).

Classification of incident duration: The estimation of incident duration can also be approached through classification methods [23]. To implement this, a specific threshold for duration is established, and predictions are made to determine whether an incident will exceed the specified time.

Incidents on different road types: The majority of previous research has primarily focused on predicting incident duration for specific types of roads, such as freeways or motorways, where data accuracy is generally higher than on arterial roads. As of 2018, only a limited number of studies have applied prediction strategies to normal arterial roads due to the increased modeling complexity and issues related to location mismatching. Most traffic incident duration analysis research concentrates solely on one type of road network, like freeways or highways, as evidenced by [65, 77, 78], and [79]. This observation is further supported by a recent state-of-the-art review published in [7], which highlights the challenges in addressing this problem for arterial roads and the scarcity of studies in this particular domain.

Model transfer between different road types is an important aspect to consider in incident duration prediction research. By studying this transferability, researchers can determine whether a model developed for one type of road network can be effectively applied to another, potentially saving time and resources in the development of new prediction models.

Model transfer: A technique that leverages knowledge gained from one domain or dataset to improve performance in another can be particularly useful in studying model transfer between road types. By examining the similarities and differences in road features, traffic patterns, and incident characteristics, researchers can gain insights into the factors that contribute to the transferability of prediction models across different road networks. Exploring model transfer between road types offers several benefits. First, it allows researchers to identify generalizable features and prediction patterns that can be applied across various road networks. This can lead to the development of more robust and versatile prediction models. Second, understanding model transferability can help identify potential limitations and areas of improvement for existing models, ultimately enhancing their accuracy and performance when applied to new road types.

Interpretable models: Bayesian networks can predict incident injury severity and are interpretable models [80]. They found to surpass various simple regression models in predicting incident severity, considering three severity indicators: number of fatalities, injuries, and property damage [81]. Interpretability goes beyond tree models. Tree rules can be extracted from other prediction models like Bayesian networks through knowledge distillation [82]. It not only represents the model as an interpretable DT but also estimates feature importance. It is possible to gain a deeper understanding of incident duration estimation through classification methods, ANNs, and Bayesian networks, and their potential for providing interpretable models and feature importance insights.

Feature importance: On top of any regression or classification model, we can perform feature importance estimation using SHapley Additive exPlanations (SHAP)—a unified measure for interpreting the output of any ML model. It connects the game theory concept of Shapley values with local explanations, improving the interpretability of the model and helping to understand the contribution of each feature to the prediction. In the context of incident duration estimation, SHAP can be particularly useful for feature importance analysis. In short, by using SHAP, we can not only create highly accurate models for predicting incident duration but also gain insights into the underlying factors that impact these predictions, and create explanations that can be easily interpreted by nonexperts. This approach can be computationally expensive since it performs model training on multiple subsets of data (specific selection of rows and columns from the feature matrix) [83, 84].

5.3. Evaluation of Prediction Accuracy. To evaluate the regression or classification model performance, the most commonly used metrics are as follows (see Table 4).

The variables in the formulas are defined as follows:

- n: The number of samples or instances in the dataset.
- A_t: The actual value or true duration of the traffic accident at time t.
- F_t: The predicted value or forecasted duration of the traffic accident at time t.
- tp: The number of true positives, which are cases where the model correctly predicts the positive class (e.g., accidents with long duration).

Table 4: Summary of metrics for evaluating traffic accident duration prediction models.

Metric	(R) Regression or (C) classification	Formula	Description	Application to traffic accident duration prediction
MAPE	R	$(1/n)\sum_{t=1}^{n} \lvert (A_t - F_t)/A_t \rvert$	Regression metric. Measures the average percentage error of regression predictions	Evaluates the average percentage difference between the predicted and actual accident durations. Lower values indicating better performance. Previous studies indicate MAPE values around 20%–65% [85]. A_t should not be zero value.
RMSE	R	$\sqrt{(1/n)\sum_{i=1}^{n} (A_t - F_t)^2}$	Regression metric. Quantifies the average difference between regression predictions and actual values, sensitive to large errors	Assesses the average difference between the predicted and actual accident durations. Lower values indicating better performance
Accuracy	C	(tp + tn)/(tp + tn + fp + fn)	Classification metric. Indicates the proportion of correct classification predictions	Measures the overall proportion of correct predictions for all accident duration categories. Higher values indicating better performance
Precision	C	tp/(tp + fp)	Classification metric. Reflects the accuracy of positive predictions made by the model	Indicates the fraction of accurately predicted accident durations out of all predictions for a specific category. Higher values indicating better performance
Recall	C	tp/(tp + fn)	Classification metric. Measures the proportion of true positives out of all actual positive instances	Represents the model's ability to identify all the relevant cases for a specific accident duration category. Higher values indicating better performance
F1-score	C	2 * (precision * recall)/(precision + recall)	Classification metric. Provides a balanced measure of classification performance considering precision and recall	Gives a balanced performance assessment for the model, considering both precision and recall for accident duration categories. Higher values indicating better performance
AUC	C	Area under the receiver operating characteristic (ROC) curve	Classification metric. Measures the overall classification performance across all classification thresholds	Assesses the model's ability to distinguish between different accident duration categories. Higher values indicating better performance

- *tn*: The number of true negatives, which are cases where the model correctly predicts the negative class (e.g., accidents with short duration).
- *fp*: The number of false positives, which are cases where the model incorrectly predicts the positive class (e.g., predicting an accident as long duration when it is actually short duration).
- *fn*: The number of false negatives, which are cases where the model incorrectly predicts the negative class (e.g., predicting an accident as short duration when it is actually long duration).

Various metrics used for the regression tasks, RMSE, and MAPE being the most common for the regression task (see Table 5).

Table 5: Metrics used across reviewed papers.

Metric	Studies
MAE	[14, 22, 62, 82, 86–97]
RMSE	[14, 22, 40, 62, 68, 86–91, 94–104]
MAPE	[14, 22, 23, 33, 40, 67, 68, 86, 88–93, 95, 97, 98, 100–114]
MSE	[96, 100]
AUC	[115–117]
Recall	[23, 40, 118]
Precision	[23, 118]
F1	[23, 40, 118]

6. ML Techniques for Traffic Incident Duration Prediction

Figure 5 illustrates the ML pipeline used for predicting the duration of traffic accidents. The pipeline consists of (1) data preprocessing (cleaning, data imputation, label encoding, and outlier detection); (2) data manipulation that may include the feature transformation (principal component analysis (PCA) and latent Dirichlet allocation (LDA), log-transformation of target variable) and feature selection (e.g., using correlation-based feature selection, univariate feature selection, and recursive feature elimination [RFE]), model training (e.g., using LR, SVMs, K-NN, DTs, RFs, and neural networks), and feature importance estimation (using Gini importance, permutation importance, or SHAP); (3) model training and evaluation (including cross-validation, confusion matrix, and ROC curve), and model deployment (accuracy, precision, recall, F1-score, RMSE, and MAPE). The model is then validated to ensure its accuracy and reliability. The pipeline shows a general way of predicting the duration of stochastic events (like accidents) using ML methods and can be used for other similar tasks (e.g., prediction of accident severity).

A ML workflow for traffic accident duration prediction includes various stages regarding preparing the data, engineering and selecting features, training and evaluating models, and ultimately deploying the model:

- Data Preparation: The first step in the process involves preparing the data for analysis. This includes imputing missing values, standardizing the data, and converting various data types. These preprocessing steps are crucial to ensure that the ML algorithms can effectively learn from the data.
- Feature Engineering: The next step involves transforming the raw data into a set of features that can be used as input for the ML algorithms. This includes type conversion, encoding categorical variables, and functional conversion, such as the vectorization of text data.
- Feature Transformation: This stage involves applying techniques like PCA and LDA to reduce the dimensionality of the feature set. This can help improve the performance of the ML algorithms by reducing noise and computational complexity.
- Feature Selection: In this step, various techniques like correlation-based feature selection, univariate feature selection, and RFE are used to select the most relevant features for the prediction task. This helps to further reduce the complexity of the model and improve its generalization capability.
- Model Training: At this stage, several ML algorithms such as LR, SVMs, K-NN, DTs, RFs, and neural networks are trained on the prepared data.
- Model Evaluation: The trained models are evaluated using cross-validation, with metrics like accuracy, precision, recall, F1 score, RMSE, and MAPE. The evaluation process helps in selecting the best performing model for the prediction task.
- Feature Importance Estimation: Techniques like permutation importance and SHAP are used to estimate the importance of different features in the model. This can provide insights into which features have the most significant impact on the prediction of traffic accident duration.
- Iterative Decision Point: A decision is made based on the model's performance. If the performance is sufficient, the process moves to the next step of model deployment. Otherwise, the model and data are further refined using techniques like hyperparameter variation, outlier removal, model selection, or model modification.
- Model Deployment: Once a satisfactory model and data organization are obtained, the resulting system is deployed for traffic accident duration prediction. Performance metrics like accuracy, precision, recall, F1 score, RMSE, and MAPE are reported as model performance estimations.

As a typical ML workflow for traffic accident duration prediction, it includes decision points and iterative steps for refining the model and data, and it highlights the importance of data preprocessing, feature engineering, model training and evaluation, and iterative refinement to achieve the best possible model for predicting traffic accident durations.

The task of predicting the duration of an incident can be solved by using ML methods. Among these methods are as follows: tree-based classification methods [66, 71], fuzzy logic [72], Bayesian networks [73], LR analysis [119], ANN

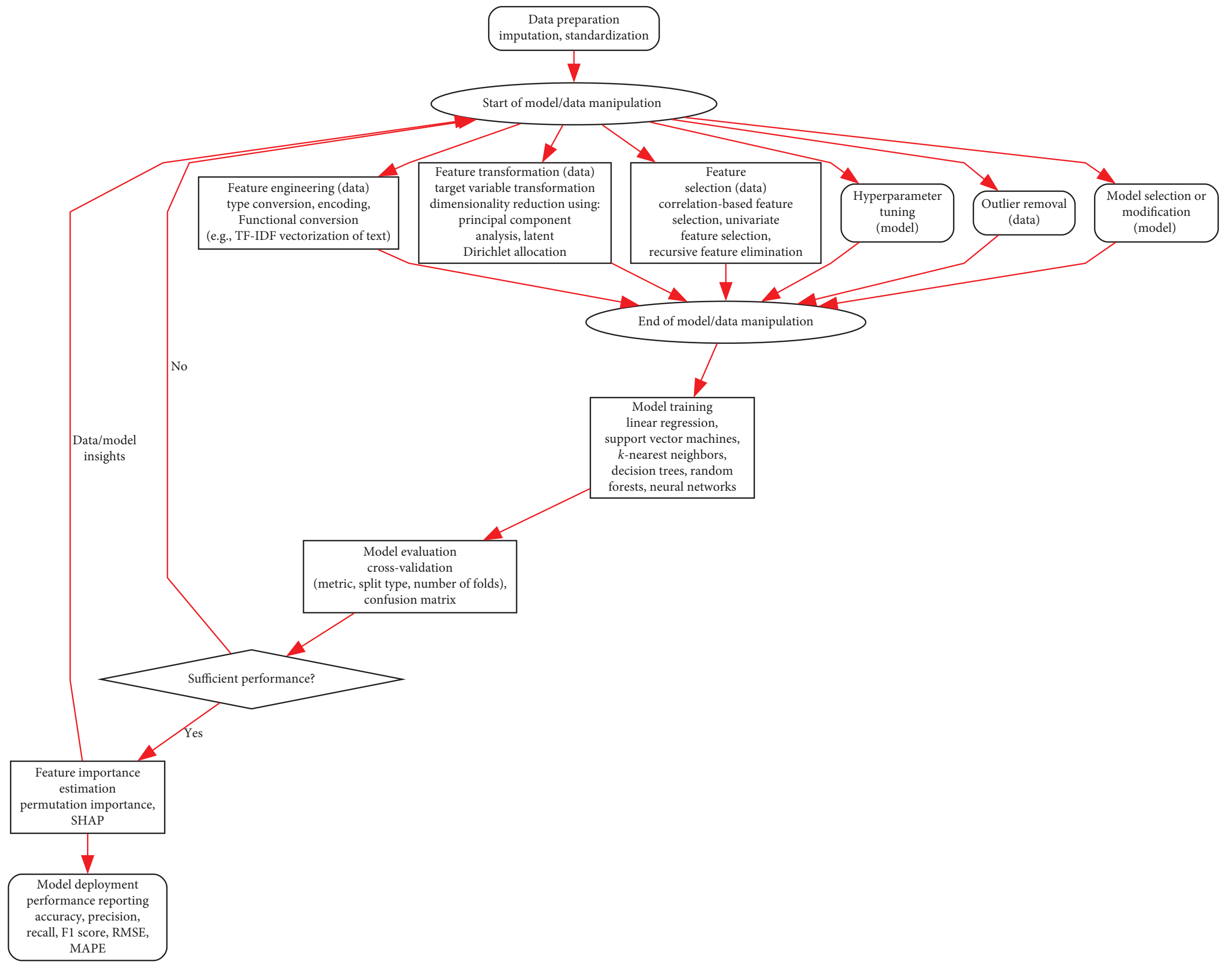

Figure 5: Machine learning pipeline for traffic accident duration prediction using table data.

[74, 75, 120], and support vector regression (SVR) [121]. Recently, GBDTs have been revealed to be a better performing method for incident duration prediction [76]. Gaussian process regression and ANNs were found to outperform tree methods and SVRs in incident duration prediction [62].

Extreme learning machine (ELM) [122] is a ML method, which incorporates a feed-forward neural network initialized with random weights and consequent training step based on produced random feature mapping, designed to avoid overfitting of neural network.

Several classic ML methods are widely utilized for traffic incident duration modeling represented in Table 6. More advanced methods are represented in Table 7.

These models can be applied to both classification and regression problems, with the exception of LR (used only for classification tasks) and linear regression (used only for regression tasks).

Recently, a two-step approach for predicting the duration of traffic incidents has been proposed [114]. The first step was binary classification, which used a cost-sensitive Bayesian network to predict whether an incident would last for more or less than 30 min. The second step was regression, which used the K-NN approach to predict the duration of the incident. Although the approach was effective, it had one significant limitation: Manually selecting the threshold for binary classification may not yield the optimal performance when using data-driven approaches or ML methods (e.g., it could result in class imbalance). Therefore, a fixed and varying incident duration threshold needs to be compared to find the best class balance for classification models; even more, a comparison with a multiclass classification approach may also be employed. Also, more advanced regression models together with outlier removal procedures can be used to model groups of incidents separately providing a better and more precise prediction of the incident duration [40]. Alternative to the incident duration threshold split, clustering methods can also be used to detect groups of traffic incident reports to model them separately. Overall, we want to highlight that within the accident report dataset, there can exist multiple report groups and when using separate modeling may provide better accident duration prediction performance [39].

Recently, research has been conducted on the classification of driving states using multiple hyperoptimized ML models, with the entire feature space visualized through t-distributed stochastic neighbor embedding (t-SNE) [137]. RF emerged as the top-performing model in terms of

Table 6: Classic methods.

Method	Description	Traffic accident duration prediction	Advantages	Limitations
K-Nearest neighbors (K-NN) [123]	Predicts based on the majority vote or average of the *k* closest neighboring data points using a distance metric.	K-NN predicts traffic accident duration using data on similar accidents in the training data. By considering the *k* closest neighbors, the algorithm accounts for the local structure of the data.	Simple to implement Can handle nonlinear patterns	Sensitive to the choice of k Poor scalability: computationally expensive for large datasets
Linear regression	Models the relationship between features and the dependent variable in regression tasks using linear equations.	LR predicts traffic accident duration by modeling the linear relationship between accident features and duration. However, it may underperform if the relationship is not linear.	Easy to interpret Computationally efficient	Assumes linear relationship Simplicity: May not capture complex patterns
Random forests (RF) [124]	Combines the average or majority votes from multiple decision trees trained on randomly selected data subsets using bootstrap aggregation.	RF can predict traffic accident duration by using multiple decision trees which reduces the sensitivity of the model to noise in the data, providing a more robust prediction.	Robust to noise Can handle missing values Less prone to overfitting Allows feature importance estimation	May require a large number of trees for robustness against outliers Computationally expensive
Support vector machines (SVM) [125]	Effective method for both classification and regression tasks.	SVMs can be used to classify accident severity or predict the exact duration of the incident. Support vector regression (SVR) can predict continuous values, while SVM can categorize accidents based on certain criteria (e.g., short/long duration, severity).	Can handle nonlinear patterns Robust to noise Effective with high-dimensional data	Computationally expensive for large datasets and high-dimensional data May require tuning of hyperparameters
Bayesian models [126]	Bayesian techniques encompass a wide range of methods suitable for regression and classification tasks.	For traffic accident duration prediction, Gaussian process regression (GPR) [127] and Bayesian networks [128] can be utilized. Bayesian models account for uncertainty in their predictions and are particularly beneficial when dealing with limited data, measurement error, and/or missing observations. GPR is a powerful nonlinear method and can be used to interpret nonlinear systems without prior knowledge	Handles uncertainty Resilient to noise Efficient for low amounts of data Flexible modeling of relationships	Computationally demanding for large datasets May need expert knowledge to define model structure
Neural networks (NN) [129]	Neural networks, particularly deep learning models, can be used for regression and classification tasks.	Feed-forward neural networks (FFNN) and recurrent neural networks (RNN), including long short-term memory (LSTM) [130], can predict traffic accident durations, considering the temporal aspect of the data	Can capture complex patterns Effective with large datasets	Requires large datasets Computationally expensive May require tuning of hyperparameters

Table 7: Advanced methods used in traffic accident duration prediction.

Method	Description	Traffic Accident Duration Prediction	Advantages	Limitations
Light gradient-boosted machines (LightGBM) [131]	Enhances gradient boosting in tree-based models, employing gradient-based one-side sampling (GOSS) to exclude data points with small residuals	LightGBM predicts traffic accident duration by using a more efficient and scalable approach to gradient boosting. By focusing on data points with large residuals through GOSS, it achieves faster training while maintaining accuracy	Faster training than GBDT and XGBoost Scalable to large datasets Can handle missing values and categorical features	May require tuning of hyperparameters Sensitive to noisy data Could perform worse than XGBoost due to nongreedy tree split search
CatBoost [132]	A powerful gradient boosting library developed for efficiently handling categorical features without extensive preprocessing, using ordered boosting and oblivious trees	CatBoost is fit for traffic accident duration prediction as it efficiently processes typical categorical features of accident reports like road types, weather conditions, and vehicle types, and can be applied to both regression and classification tasks	Designed for handling categorical features Robust to overfitting Applicable to regression and classification tasks Can handle missing values	Computationally expensive for large datasets Hyperparameter tuning may be necessary
Extreme learning machines (ELM) [133]	A single-hidden-layer feed-forward neural network (SLFN) initialized with random parameters, where optimization performed only for the output layer's weights	Extreme learning machines are suitable for traffic accident duration prediction on large datasets, as they provide faster training compared to traditional feed-forward neural networks while maintaining good accuracy. The method randomly assigns input weights and biases, focusing solely on the optimization of weights of the output layer	Faster training than traditional neural networks Can handle nonlinear patterns	May struggle with complex patterns or categorical data (may require one-hot encoding) Less interpretable than other methods
Adaptive boosting (AdaBoost) [134]	This algorithm combines multiple weak learners into a strong learner, adjusting training instance weights adaptively	AdaBoost can be employed for predicting traffic accident duration, forming a strong predictive model by combining multiple weak learners. By adaptively adjusting the weights of training instances, the model focuses on harder-to-predict instances, enhancing overall accuracy	Robust to overfitting Resistant to noise Metamodel: can use various weak learners (e.g., decision trees)	May be computationally expensive based on the weak learner type and depth of boosting May necessitate weak learner hyperparameter tuning
Gradient boosting decision trees (GBDT) [135]	A technique that trains a sequence of models, where each subsequent model is added to minimize the residuals of the previous models	GBDT is suitable for predicting traffic accident durations by progressively enhancing prediction performance based on the residuals of prior models. The final prediction is an ensemble of all weak learners	Resistant to overfitting Handles missing values effectively Accommodates mixed feature types	Sensitive to noisy data Longer training time (typical for boosting methods)
eXtreme Gradient Boosting (XGBoost) [136]	Determines split values by performing an exhaustive search over all possible splits for each feature, incorporating a regularization parameter	XGBoost can predict traffic accident durations with high accuracy due to its ability to exhaustively search for optimal split values and control model complexity with regularization	High prediction accuracy Parallelizable for faster training Regularization prevents overfitting	Longer training time (common for boosting methods) May require tuning of hyperparameters

prediction accuracy. However, there are more advanced tree-based models available that employ gradient boosting, such as gradient-boosted DTs.

Notably, a significant portion of the literature on traffic incident duration prediction focuses on a single specific road type like freeways, highways, or motorways. The data accuracy on these roads is typically higher compared to arterial roads. Consequently, relatively fewer studies have been carried out on arterial roads due to the complex modeling and location-matching challenges associated with them, up to the year 2018 [65, 77–79]. This observation is highlighted in a recent state-of-the-art review [7], which underscores the difficulty of addressing this problem in arterial roads and points out the scarcity of research in this domain. This specialized approach might reduce the applicability and generalizability of the developed models, as they may not capture the distinctive characteristics and challenges associated with varied road networks.

6.1. Classification and Regression Tasks. In traffic incident analysis using ML methods, incident features are generally prepared and represented as a numerical matrix:

$$X = \left[x_{ij}\right]_{i=1\ldots N_i, j=1\ldots N_f}, \tag{5}$$

where N_i represents the total number of traffic incidents and N_f represents each incident's characteristics.

The task of estimating traffic incident duration can be framed as a classification task. Here, an arbitrary threshold for the duration is set and the model's task is to predict whether the incident extends beyond that limit. For such classification problems, the incident duration classification vector can be expressed as follows:

$$Y_c = \begin{cases} 0, & \text{if } y_i \leq T_c \,(\text{short} - \text{term incidents}), \\ 1, & \text{if } y_i > T_c \,(\text{long} - \text{term incidents}), \end{cases} \tag{6}$$

where

- N is the duration of the traffic incident in minutes.
- Y_c is the binary classification output.
- T_c is the incident duration threshold.

The purpose here is to predict whether an incident will last longer than threshold T_c based on its features.

6.2. Feature Engineering in Traffic Accident Modeling. In traffic incident duration regression analysis, striking a balance between the number of features considered and the dataset size is essential. Using an excessive number of features with a limited dataset size can lead to overfitting. Some features may be beneficial, irrelevant, or have varying degrees of significance, while others might not impact prediction results at all. By conducting a feature importance analysis, we can guide traffic management facilities to record the most essential data and exclude unnecessary information related to traffic incidents. Additionally, we can improve the precision of specific observations, such as weather and lighting conditions, which have been shown to play a substantial role in some research studies. By focusing on the unique requirements of each task, researchers can effectively identify the most relevant features and data sources, which in turn can enhance the performance of the predictive models.

6.2.1. Essential Factors in Traffic Accident Modeling. By exploring the effects of various features and their significance on incident duration, researchers can develop more accurate models that take into account critical factors while discarding less relevant ones. This can lead to improved prediction performance and better decision-making in traffic management and incident response strategies. Table 8 for aspects of incidents found relevant across various studies.

According to research conducted in 2009 on Washington State freeways [19], response team preparation times increased noticeably during the summer and autumn. Although the weather did lengthen preparation times, it had no significant effect on clearance and response team travel times. The study found that "peak hours" had the most profound impact on response team preparation delay. Conversely, the RT was lower during summer and winter. This lower RT during peak hours was attributed to the response procedures aiming to resolve incidents as quickly as possible during these hours. Furthermore, weekend and nighttime incidents were associated with longer clearance and impact durations due to fewer staff on duty [19, 33].

A study analyzing Beijing traffic data from 2008 [22] revealed that "peak hours" significantly affected response team travel time and clearance time but did not impact response team preparation time. Interestingly, during the summer and autumn months, response team preparation time was higher, which did not affect the clearance and travel time. The type of incident, such as an overturned vehicle, significantly influenced both the response team preparation and incident clearance times but did not affect travel time. The study also identified the involvement of bikes, which implied human injuries, and the night shift (10 pm–6 am) as the most important factors for the total incident duration.

Understanding the impact of features such as peak hours, response preparation time during different seasons, and the type of incidents can significantly enhance the accuracy of incident duration prediction models. By focusing on relevant data and discarding the irrelevant, a more accurate prediction and efficient response can be achieved, thereby reducing the overall impact of traffic incidents on freeways [19, 22, 33].

One important conclusion is that rather than using the number of affected lanes [23] (which is only part of information related to road segment), we can also add a resulting value called "all lines affected" [30] or even more informative—ratio of affected lanes (e.g., 50%) indirectly incorporating the number of lanes of a section, where incidents occurred into the model. Also, we can include the number of lanes for the each involved section as a feature.

The average traffic speed during a 60-min time interval was found to be statistically significant for the traffic incident duration modeling [20]. One can use aggregated traffic flow

Table 8: Summary of key findings on factors influencing traffic incident duration.

Key Finding	Description
Increased preparation times during summer and autumn	According to research conducted in 2009 on Washington state freeways, response team preparation times increased noticeably during the summer and autumn [19]
Impact of weather on preparation times	Although the weather lengthened preparation times, it had no significant effect on clearance and response team travel times [19]
Peak hours effect on preparation delay	"Peak hours" found to have the most profound impact on response team preparation delay [19]
Improved response during peak hours in summer and winter	Response times were lower during summer and winter in peak hours due to response procedures aimed at resolving incidents quickly during those hours [19]
Weekend and nighttime incidents	Weekend and nighttime incidents led to longer clearance and impact durations due to fewer staff on duty [19, 33]
Peak hours effect on travel and clearance time	A study analyzing Beijing traffic data from 2008 showed that "peak hours" significantly affected response team travel time and clearance time but did not impact response team preparation time [22]
Higher preparation time in summer and autumn	During summer and autumn months, response team preparation time was higher, which did not affect clearance and travel time [22]
Impact of incident type on preparation and clearance times	The type of incident, such as an overturned vehicle, significantly influenced both response team preparation and incident clearance times but did not affect travel time [22]
Influence of night shift and human injuries	The involvement of bikes, indicating human injuries, and the night shift (10 pm–6 am) were identified as critical factors for the total incident duration [22]
"All lines affected" and ratios	Rather than only using the number of affected lanes [23], it may be useful to include the resulting value called "all lines affected" [30] or the ratio of affected lanes (e.g., 50%), indirectly incorporating the number of lanes in the section where incidents occurred or include the number of lanes for each involved section as a feature
Significance of average traffic speed	The average traffic speed during a 60-min time interval is statistically significant for traffic incident duration modeling [20]. One can use aggregated traffic flow data (e.g., traffic speed or flow count) within specific time intervals as a feature
Precise lighting conditions calculation	Lighting conditions can be calculated more precisely than using binary day/night values [33]. Using longitude, latitude, and time, determine the angle between the sun and traffic direction and calculate precise lighting conditions based on the elevation of the sun above the horizon during the incident
Effects of lighting on driver behavior	Research using an interactive driving simulator showed that driver perception is limited during nighttime. Drivers also limit their travel speed due to impaired visibility [138]. Improvements like road markers and rumble strips were proposed to enhance visibility. This highlights the potential to derive visibility of the road structure relative to traffic incidents and corresponding lighting conditions

data (e.g., represented as a speed of traffic or possibly traffic flow count) within specific time intervals as a feature.

Lighting conditions can be calculated much more precisely than just using binary day/night values [33]. Using longitude, latitude, and time, we can determine the angle between the sun and traffic direction at the time and place of the incident. Also, we can calculate precise lighting conditions based on the elevation of the sun above the horizon during the time of the incident.

Effects of lighting conditions on driver behavior were assessed using an interactive driving simulator [138]. According to the research, driver perception was found to be limited during nighttime; drivers were also found to be limiting travel speed due to impaired visibility (which includes an incorrect and untimely perception of the road). Installation of road markers (to improve the perception of the road curve) and rumble strips were proposed. The impaired visibility during nighttime and the proposed measures point that it is possible to derive visibility of the road structure from the point of traffic incidents and corresponding lighting conditions (distance of driver's eyesight in relation to road segments).

6.2.2. Log-Transformation of the Traffic Incident Duration Variable. In predictive ML tasks, it is crucial to examine the statistical properties of the target variable. Particularly, the target variable's distributional characteristics can significantly impact the accuracy and performance of the predictive models. Often, the incident duration distribution is not normally distributed; instead, it follows a skewed distribution such as a log-normal or log-logistic (see Figure 6). This skewness can be due to numerous extreme values, possibly caused by a few incidents lasting extraordinarily longer than the majority. The logarithmic transformation is a monotonic transformation known for its effectiveness in stabilizing variance, reducing skewness, and making the target predicted variable (incident duration) resemble a normal distribution more closely. This is especially critical for linear models and other algorithms that have an implicit or explicit assumption of normality or symmetry in the target variable distribution. The log-transformation can help reducing the skewness and decreasing the influence of extreme values. For log-normally distributed data, a log-transformation will result in a normal distribution. For log-logistically distributed data, a log-transformation may not lead to a normal distribution but can still significantly reduce skewness. Therefore, for AI-based traffic incident duration prediction tasks, if preliminary data analysis indicates that incident duration follows a skewed distribution (log-normal or log-logistic), a log-transformation is a beneficial preprocessing step.

6.2.3. Textual Accident Description Analysis. Traffic accident reports usually contain a textual description of the accident [37, 38]. In recent years, multiple systems were presented to detect traffic accidents using text analysis of social network contents [139–141]. Various methods were also proposed for traffic-related sentiment analysis of social networks: sentiment classification using ontology and LDA [142] and the use of gated recurrent unit (GRU) model and generative adversarial networks to estimate the traffic information sentiment [143]. Overall, sentiment analysis has been performed for various traffic rules including "yellow light rule" using social network analysis [143, 144].

Also, accident reports can contain a category and subcategory definition of the accident (e.g., types of vehicles involved and multiple or single-vehicle crash). The unique property of text description of an accident is that it can contain information regarding event categories not predefined in the accident reporting form [139].

A typical pipeline for textual description preprocessing includes the following [141]: 1) tokenization—text being split into a list of words called tokens; 2) stop-word removal—the removal of pronouns, prepositions, symbols, and articles not providing any valuable information for accident description; 3) lemmatization and stemming—words are reduced to their base form (e.g., involved -> involve, injuries -> injury, reported -> report) or to their root form (e.g., injuries, injury -> injur); 4) case conversion—text is converted into lower case, where the difference between uppercase and lowercase words is not relevant; 5) part-of-speech (POS) tagging—each word gets its type associated with it (e.g., traffic -> noun, stop -> verb); and 6) text representation conversion, which relies on a bag-of-words (BOW) representation (each word is represented as a one-hot encoded vector) or on a neural network–based word embedding method like Word2Vec or FastText, which capture semantic similarities between words. The Word2Vec approach has a significant limitation—the inability to represent a new word which was absent in the training dataset with a vector. The FastText resolves this issue by representing each word as a sum of related n-gram vectors.

After the data preparation and representation conversion, various recurrent models are then used to perform tasks related to text analysis.

Incident description features were used in topical text modeling [145]. Previously, the LSTM architecture has been used for the task of detection of incidents from social media data [146]. LSTM was also successfully used for stock price time-series prediction [147], making it applicable for the modeling of traffic flow/speed time-series data.

Text analysis of accident reports is vital for understanding the underlying causes of traffic accidents. By providing insight into the information provided to describe accidents, text analysis can help to identify dangerous conditions that lead to accidents. Traffic accident descriptions can contain inaccuracies due to human factor (e.g., inaccurate accident timeline), which highlights the importance of automated accident detection and timeline estimation from traffic state data [148]. Ultimately, text analysis is a powerful tool for gaining even deeper insight into the causes of traffic accidents unconstrained by accident reporting forms and developing strategies to reduce them.

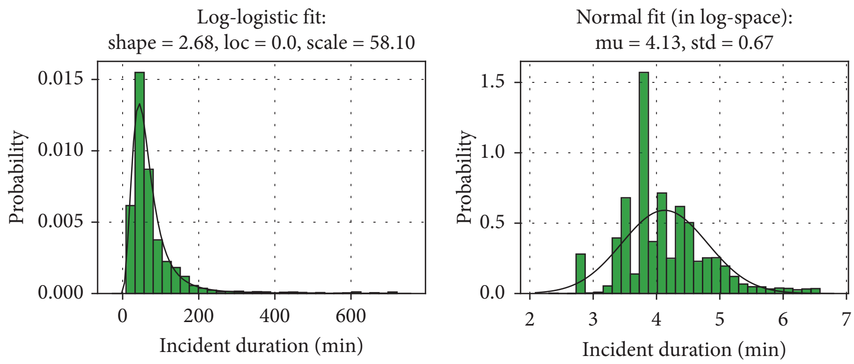

Figure 6: Example of distribution fit for the sample data of San Francisco dataset [40].

6.2.4. Feature Importance and Feature Selection. By using a combination of data-driven methods and ensemble techniques, researchers can more accurately assess the importance of various features in the context of traffic incident duration prediction. This can lead to the development of more effective models that take into account critical factors while discarding less relevant ones, ultimately improving prediction performance and enhancing decision-making in traffic management and incident response strategies. Furthermore, adopting a data-driven approach tailored to specific objectives can emphasize the types of data that need to be collected for particular models and prediction tasks, such as accident risk, duration, or severity prediction.

A promising approach for analyzing feature importance is using DTs from tree-ensemble models, as proposed in [149]. A data-driven technique can be utilized to fuse information from multiple sources, where the Gini index, extracted from RFs, was utilized to estimate feature importance [150]. However, it is crucial to recognize that a random model might exhibit significant variability in data mapping when the relationship between features and the target variable is weak. This can result in the feature importance value being heavily influenced by the randomness intrinsic to the model evaluation. In order to address this challenge, it is essential to consider model ensembles or use methods that provide a more stable and robust estimation of feature importance. Techniques such as the SHAP [151] offer more advanced approaches for estimating feature importance, as they fuse estimations from multiple models trained on various subsets of the dataset.

There are models of different complexities used to approximate traffic incident duration and duration of its phases. Khattak [119] used simple linear model to approximate clearance time, which can be defined as a function of incident report variables (feature_*) and coefficients (A_*):

$$\begin{aligned}\text{clearance time} = {} & A_1 * \text{feature}_1 + A_2 * \text{feature}_2 \\ & + \cdots + A_2 * \text{feature}_N.\end{aligned} \tag{7}$$

By using this simple model, we can easily determine the most important features by consequently removing them and estimating the resulting error (e.g., estimate change in MSE in relation to the removed feature). This method for determining feature importance is called RFE [152] and is a general method which is applicable to different approximation models.

The SHAP [151] offers an advanced methodology for estimating feature importance, as it combines estimations derived from multiple models trained on various subsets of the dataset. These subsets are selected based on both feature-scale and index-scale considerations. A typical SHAP plot (Figure 7) visually represents the SHAP values and their impact on model predictions. In this plot, each point corresponds to a specific feature and is positioned along the vertical axis (Oy-axis) according to its SHAP value score. The points are color-coded based on their respective values, ranging from low to high. The horizontal axis (Ox-axis) represents the influence of each feature on the overall prediction output. The SHAP approach offers several benefits when estimating feature importance. By aggregating information from multiple models and dataset subsets, it provides a more robust and comprehensive representation of the importance of each feature. Additionally, the feature importance plot allows for easy interpretation of the relationships between features and their impact on model predictions, making it an invaluable tool for understanding the model performance on complex datasets.

6.3. Advanced ML Techniques. Effective traffic incident duration prediction and classification are vital for traffic management. While ML aids these tasks, challenges specific to the data-driven approach like outliers, imbalanced data, and high dimensionality can impair model performance. This section covers techniques to address these issues.

6.3.1. Anomaly and Outlier Detection. While the literature extensively explores ML for traffic incident severity classification (e.g., [153]) and duration prediction (e.g., [7]), it often overlooks the critical challenges posed by outliers and imbalanced datasets, which can significantly hinder model performance.

Several studies, including [74, 121] and [154], have utilized classification or clustering methods for detecting traffic incidents through the analysis of traffic conditions. Anomaly detection methods, such as one-class SVM and Isolation Forest (IF), offer a distinct advantage in identifying deviations from established traffic patterns. By learning the

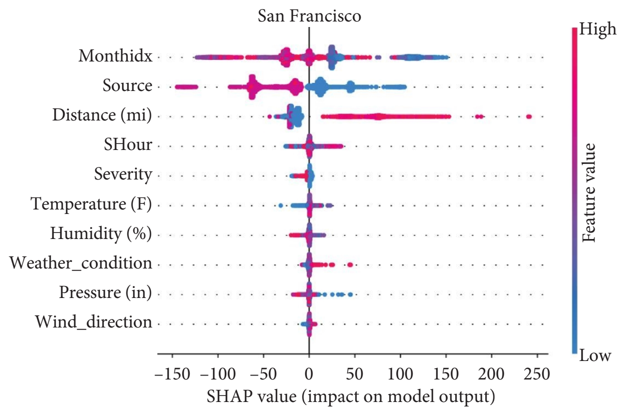

Figure 7: Feature importance for all-to-all regression using XGBoost for San Francisco, USA [40].

"normal" traffic behavior, these methods can effectively flag unusual events, including previously unseen incident types. This adaptability makes them particularly suitable for detecting nonrecurring incidents, potentially surpassing the limitations of traditional classification and regression techniques. Nonrecurring traffic incidents are inherently rare and atypical, making their detection a prime candidate for anomaly detection within traffic data. Anomaly detection excels in identifying outliers, making it ideal for detecting infrequent, nonrecurring incidents that might be missed by conventional methods. In the context of traffic incident duration prediction, the application of anomaly detection techniques can be particularly beneficial to enhance incident duration prediction model accuracy [15].

Anomaly detection methods such as one-class SVM and IF provide a measure of an anomaly for each data point. Their effectiveness can be analyzed alongside traditionally employed regression (e.g., GBDT, ANN, Gaussian process regression) and classification (GBDT, ANN) methods for estimating incident probability and duration. Anomaly detection can be particularly effective for modeling durations that follow a long-tail distribution where tree models underperform, as suggested by Ma et al. [76]. Outliers in the long tail could be identified as anomalies corresponding to rare incident durations.

IF [155] is an efficient anomaly detection technique that uses a forest of random split trees to isolate anomalies. It randomly selects a feature and a feature value, then partitions the dataset recursively until each data point is isolated. A shorter average tree depth across all trees implies a data point is an anomaly, as outliers can be isolated quicker. On contrast, the local outlier factor (LOF) method [156] computes anomaly scores grounded on the local deviation of data point density within the K-NN. The method calculates a local reachability density (LRD) for each point, an inverse measure of the average reachability distance (RD) from that point to its neighbors in a k-filtered neighborhood. The LOF score is the ratio of the LRDs of a point's neighbors to its own. Sorting the LOF scores helps identify outliers residing in low-density zones, while nonoutliers are situated in higher-density regions.

Both IF and LOF offer distinctive advantages in isolating anomalous traffic incidents. IF identifies incidents significantly deviating from the majority, unveiling underlying patterns or contributing elements. For instance, a cluster of high-scoring incidents can hint at problematic road designs needing improvement. LOF, meanwhile, finds incidents unusual for their local context, offering targeted information for prioritizing road safety efforts. LOF and IF provide invaluable analytical tools for traffic accident data, making them robust techniques to augment the data analysis of traffic incidents. One-class SVM, covariance estimator, LOF method, and IF have been applied for outlier detection in ML pipeline to predict rail-road accidents [157]. Anomalies detected in historical traffic state data (traffic flow, traffic speed, and occupancy) and vehicle trajectory [158], which can be associated with disruptions produced by traffic accidents. Three main approaches for anomaly detection include the use of statistical models, similarity-based models (which rely on data difference measures and neighborhood estimation methods to find outliers) and pattern-mining methods (which resolve the correlation problem of similarity-based models but are very time-consuming).

In conclusion, anomaly detection can be an effective tool for improving the estimation of incident probability and duration, as it can identify and isolate rare events (or anomalous accident reports) and also can eliminate data records which can contribute to the bias of the prediction model (which can negatively affect model's performance [159]).

6.3.2. Imbalanced Dataset Classification. In traffic incident duration prediction, we may encounter rare incidents with durations exceeding 60 min, representing extreme values. Traffic incident duration is often modeled using log-normal, Gamma, and Weibull distributions, as they exhibit asymmetric and long-tailed distributions. Thus, classification methods for imbalanced datasets can be beneficial for classifying these rare incidents.

Three primary approaches for handling imbalanced classes include the following:

- Undersampling—Reducing the size of the majority class to achieve class balance.
- Oversampling—Increasing the size of the minority class (e.g., by generating synthetic samples) to achieve class balance.
- Combined approach—Utilizing both under- and oversampling techniques to balance class proportions [160].

Situations in which these approaches can be used can be attributed to dataset sizes and computational resource availability (see Table 9). For example, in the case of limited computational resources, the data may be subsampled (e.g., 20% of data points can be extracted from the total amount of data) to perform experiments and establish and test the prediction pipeline before deploying the model and using it on complete data. A combined approach may be used to perform a undersampling (to attribute to computational resource limitations) and then oversampling of minority class to achieve the class balance. However, it is necessary to state that these approaches are used to enhance the model performance during model training and should not be used on the validation dataset. It is important to maintain the original distribution of the validation set, so it accurately represents the real-world data that the model will encounter. It is important to split your data into training, validation, and test sets before applying these techniques. If you were to perform oversampling or use a combined approach before splitting, information from the validation/test sets could inadvertently leak into the training set, leading to overly optimistic performance estimates.

Various methods implement each approach for addressing imbalanced datasets:

- Random undersampling [161]—This method involves randomly selecting samples from the majority class to achieve better balance with the minority class.
- Tomek links [162]—This technique identifies pairs of instances in which one is from the majority class and the other is from the minority class, and they are nearest neighbors. The majority class instance in each pair is removed to balance the dataset.
- Synthetic minority oversampling technique (SMOTE) [163]—This method generates synthetic examples for the minority class by selecting random points on line segments between k neighbors in a multidimensional space.
- Borderline-SMOTE [164]—An improvement over SMOTE that focuses on generating synthetic samples near the decision boundary, where the minority class is at a higher risk of being misclassified.
- Adaptive synthetic sampling approach for imbalanced learning (ADASYN) [165]—This technique uses the distance distribution among k neighbors to determine the number of synthetic samples to generate. The sample generation process is similar to SMOTE, but the number of samples generated is adaptive based on the frequency of the distances among k neighbors.
- Granular repetitive SVM undersampling [166]—This method leverages the SVM algorithm for undersampling since only a few support vectors are crucial for classification in SVM. By selecting instances near the decision boundary, this technique reduces the majority class while maintaining informative samples.

These techniques can be combined or modified to create tailored solutions for specific imbalanced dataset classification problems (see Table 9). The choice of method should be guided by the dataset's characteristics, the problem domain, and the desired classification performance.

6.3.3. Boosted Models and Ensembles. Different ML techniques exist in the literature, and each demonstrates different extrapolation performances; AdaBoostSVM and Extreme, for example, are ML approaches which has to property to make predictions with reduced overfitting.

AdaBoost is a metaestimator, a ML method which implies a training of a set of weak classifiers with adaptive change to weights samples depending on the correctness of classification (after each boosting operation, weights of samples are adapted, so next trained weak estimator gives higher priority to miss-classified samples) [167]. Classifiers are then combined into weighted voting (linear) combinations.

AdaBoost and SVM were combined to solve the classification problem of imbalanced datasets [121] using spatial–temporal traffic data for the case of automatic incident classification.

ELM [154] is a ML method, which incorporates a feedforward neural network initialized with random weights and consequent training step based on produced random feature mapping, designed to avoid overfitting of neural network. The method is two-step: 1) Neural network initialized using random weights (this way we perform feature mapping into ELM feature space); 2) then, Moore–Penrose generalized inverse performed on the hidden layer output matrix and solving feature classification problem using Gaussian estimation. ELM is very fast to train and provides better generalization than ANN, also demonstrated remarkable efficiency in comparison with SVM, NB, and ANN on the traffic incident detection task on I-880 freeway in California [154].

Multiple models across many research studies failed to predict extreme values for the traffic incident duration [22, 168]. ML method, GBDT, demonstrates superior performance on a wide variety of tasks also known as failing at predicting very long incident duration, which is represented as extreme values within part of long-tailed distribution [76].

6.3.4. Techniques for Dimensionality Reduction and Model Interpretability. Usually, traffic flow data are represented as traffic state readings across multiple (up to hundreds) vehicle detectors in the transport network. Options to use such a high-dimensional dataset include the following: (1) use the closest to the POI readings (e.g. traffic flow on five closest road segments) [23], (2) use all the data available on traffic flow in the network [23], and (3) perform dimensionality reduction before modeling [121].

PCA is a statistical method that uses an orthogonal transformation to transform multidimensional data into a sequence of linearly uncorrelated variables (components). Each resulting component is a linear combination of input features. The very first component has the greatest variance. The method was proposed in 1901 by Pearson [169]. PCA was used to reduce 23 dimensions of spatial–temporal signals to two dimensions for the task of automatic incident classification proposed in [121].

Uniform Manifold Approximation and Projection (UMAP) [170] is a new dimension reduction method (2018), based on the use of Riemann geometry and algebraic topology. UMAP optimizes the placement of data in a small dimension space to minimize cross-entropy between two topological representations. The disadvantage of the method is its noninterpretability and noninvertibility (one cannot

Table 9: Considerations for using over/undersampling methods.

Criteria	Undersampling	Oversampling	Combined Approach
Large dataset	Suitable	—	May be used
Small dataset	—	Suitable	—
Extreme imbalance	—	May be used	Suitable
Limited computational resources	Suitable	—	May be used

inverse mapping to feature data). UMAP can be used both for dimensional reduction and data visualization. To perform clustering and modeling of data points in reduced dimensions and to perform the data visualization, it is necessary to use different sets of options (e.g. minimal distance between points in reduced dimension needs to be set to zero when preparing data for clustering).

Dimensionality reduction methods can help when we have multidimensional data with a lot of dimensions. A lesser number of dimensions can provide simpler and faster models or allow to the process of large datasets within computational resource constraints.

Interpretable ML [171] involves extracting relevant knowledge from ML models to provide insights into the problem domain, guiding further actions and discovery. This knowledge can include feature importance, dataset visualization, or learned relationships within the data.

Various methods and techniques can be employed to create interpretable models.

- DT-based methods: These methods represent the model as an interpretable DT, illustrating the classification decisions and feature importance estimates. Examples include Classification and Regression Trees (CART) and C4.5 [172].
- Knowledge distillation: This technique is utilized to extract decision rules from different prediction models, such as Bayesian networks [82], and can be used to create interpretable tree or rule-set models.
- SHAP (SHapley Additive exPlanations): SHAP values [151] provide a unified measure of feature importance that can be used with any ML model, allowing for interpretability and comparison across models.
- UMAP: UMAP [170] is a dimensionality reduction technique that creates interpretable visualizations of high-dimensional datasets while preserving local and global structures and can be fine-tuned either for ML tasks (classification, clustering, etc.) or visualization.
- t-SNE: t-SNE [173] is another dimensionality reduction technique that excels at visualizing high-dimensional data by placing similar data points close together in a lower-dimensional space, making complex structures more interpretable.
- Bayesian networks: These models can produce interpretable representations for predicting incident injury severity [80]. They have been shown to outperform regression models in incident severity prediction, involving indicators such as the number of fatalities, injuries, and property damage [81].

In summary, interpretable ML allows to extract relevant knowledge from various models, which can then be utilized by decision-makers. DT-based methods, knowledge distillation, SHAP, UMAP, t-SNE, and Bayesian networks are particularly useful for creating interpretable models and estimating feature importance. SHAP's flexibility, being applicable to any ML model, makes it a particularly valuable tool for model interpretability.

7. Overview of 2023–2024 Studies on Traffic Incident Duration Prediction

Traffic incident duration prediction remains a critical area of research within transportation management and emergency response systems. The 2023–2024 studies showcase advancements in the utilization of various data types and ML techniques to enhance the prediction accuracy of traffic incident durations.

7.1. Studies Utilizing Unstructured Text Data. A pivotal aspect of improving prediction models is the effective utilization of both structured and unstructured data types. Traditional methods often focus on structured data, which, although valuable, neglects the intrinsic information embedded within textual descriptions. Recent literature has increasingly highlighted the potential of integrating unstructured text data to enhance prediction accuracy. Predicting traffic incident duration is crucial for efficient traffic management and emergency response systems. Traditional approaches rely heavily on structured data, such as time, location, and weather conditions, to estimate incident clearance times. However, recent studies have highlighted the limitations of solely relying on structured data and have explored the potential of incorporating unstructured text data from incident reports to improve prediction accuracy [174–177].

Chen and Tao [174] investigated the use of text mining and ensemble learning techniques to predict traffic accident durations on expressways. Their study utilized a dataset of 22,497 traffic accident samples from the Shaanxi Province expressway monitoring system, covering structured data and unstructured text data from accident reports. The authors utilized Term Frequency-Inverse Document Frequency (TF-IDF) to extract features from the text data and compared the performance of six ML models (DT, K-NN, SVR, RF, GBDT, and eXtreme Gradient Boosting) with and without the inclusion of text data. Their findings indicated that integrating text data significantly improved the accuracy of all models, with the hybrid TF-IDF-RF model achieving the best performance. This study highlighted the importance of incorporating text data and the effectiveness of ensemble learning in traffic incident duration prediction.

Zhao, Ma, Peng, and Cheng [175] focused on predicting metro incident durations using a combination of structured data and unstructured text logs. They analyzed 5 years of incident records from the Hong Kong Mass Transit Railway, using a biterm topic model (BTM) to extract latent topics from incident narratives. These topics, alongside structured data features, were integrated into various prediction models, including ridge regression, SVR, radial basis function (RBF) neural network, GBDTs, and the AFT model. The study demonstrated that incorporating the BTM significantly enhanced prediction accuracy, particularly for incidents lasting over 30 min. The authors emphasized the value of capturing response strategies from text data, which proved to be more influential than inferred causes in predicting incident duration.

Park, Lee, and Dimitrijevic [176] proposed a supervised topic modeling approach, specifically labeled LDA (L-LDA), to predict incident duration times using text-based incident reports. Their study utilized 1466 incident records from the Korea Expressway Corporation, demonstrating that L-LDA achieved promising prediction accuracies, especially for longer duration incidents, outperforming established models like SVMs, K-NN, and DTs. The authors emphasized the importance of meticulous data preprocessing and the potential of supervised topic modeling for extracting meaningful information from text data to enhance incident duration prediction.

Chen et al. [177] investigated the effectiveness of multimodal data in predicting traffic accident durations on expressways. Their study utilized a dataset of 3887 traffic accident records from Shaanxi Province, China, encompassing structured data and unstructured text data. The authors compared the performance of various ML and DL models, including a proposed BiGRU-CNN architecture. They found that DL models, particularly BiGRU-CNN, exhibited superior performance when utilizing multimodal features. Notably, the Word2Vec model proved to be more effective than more complex alternatives like BERT and RoBERTa in extracting meaningful features from text data for this specific application.

These studies collectively highlight the significant potential of incorporating unstructured text data from incident reports to enhance the accuracy of traffic incident duration prediction models.

7.2. Studies on Influential Incident Characteristics. Numerous recent studies have explored the factors influencing incident duration, using various methodologies and datasets. This literature review synthesizes recent findings on this topic, focusing on the influence of weather, traffic conditions, and accident characteristics.

Su et al. [178] investigated the impact of weather conditions on traffic incident delays using data from New York State in 2020. The authors used advanced statistical models, namely, the random parameters hazard-based duration model (RPHDHM) and the random parameters logit model (RPLHM), to analyze the relationship between weather factors (wind speed, temperature, visibility, precipitation) and traffic delay duration and severity. The study identified strong breezes, low visibility, and precipitation exceeding 1 mm as significant contributors to increased delay duration. However, model transferability limitations were acknowledged, highlighting the need for validation in different contexts.

Zeng et al. [179] focused on the impact of real-time weather conditions on freeway accident clearance times, utilizing data from the Kaiyang Freeway in China (2014). Their study uses a novel grouped RPHDHM with time-varying covariates, revealing that higher wind speeds, lower temperatures, and lower humidity were associated with longer clearance durations. The study underscored the importance of incorporating dynamic weather data into incident management strategies and highlighted the presence of unobserved heterogeneity in accident clearance times. However, data limitations (single freeway, single year) necessitate further research for generalizability.

Luo and Liu [180] examined road closure time characteristics of tunnel traffic accidents using data from Pennsylvania (1997–2020). They used a comparative approach to analyze differences between tunnel and nontunnel incidents, finding that tunnel accidents were more likely to lead to road closures despite being generally less severe. A negative binomial regression model revealed that hit-fixed-object collisions, accidents involving injuries or fatalities, DUI involvement, heavy truck involvement, and the number of vehicles involved were significant predictors of extended closure times. The study highlighted the unique challenges of tunnel environments and the need for more granular data on accident location within tunnels and tunnel geometric features.

Lyu and Lin [181] investigated accident duration using survival analysis techniques (Kaplan–Meier and Cox proportional hazards models). By combining accident records and traffic condition data from Amap, the study found that nighttime accidents and off-rush hour accidents were significantly longer on urban arterial roads. Faster TIM staff arrival was associated with reduced durations, highlighting the importance of efficient incident response. Surprisingly, higher road complexity correlated with shorter durations, suggesting quicker self-clearance by drivers.

Finally, Zhang et al. [182] analyzed factors influencing highway accident clearance time in Shandong Province, China, using data from 2016 to 2019. Utilizing both generalized linear model (GLM) and mixed-effects model, the study identified embedding congestion, weather conditions, accident type, time of day, and vehicle types as significant predictors. Surprisingly, heavier congestion was associated with faster clearance times, contradicting common expectations. Sunny days and accidents involving cars were also linked to shorter clearance times, while nighttime accidents and those involving larger vehicles tended to have longer durations.

7.3. Applications of ML Models. Recent studies have utilized extensive traffic incident datasets. For instance, [15] employed a dataset of over 110,000 incident records from Houston, encompassing 52 features related to incident characteristics, traffic, weather, and time. Similarly, [183] used a comprehensive dataset of 370,000 incidents from the same city, with 63 features including incident details, weather, and road conditions.

Yang, Corbally, and Malekjafarian [184] focused on motorway incident duration estimation, specifically on the M50 motorway in Ireland. Their study compared the performance of SVM, ANNs, and regression trees using a dataset incorporating incident, weather, and traffic flow data. Their findings highlighted the superior performance of SVM with a Gaussian kernel for regression analysis and the flexibility of ANN for predicting longer incident durations.

Building on the importance of feature engineering, Obaid et al. [183] explored the impact of six different feature engineering techniques on model performance (including log-normal transformation for skewed data, min-max normalization, and PCA for dimensionality reduction). Using a larger dataset of approximately 370,000 incident records, their study demonstrated that PCA consistently improved accuracy across all tested ML models. Additionally, RFE emerged as the most effective feature selection method, identifying an optimal feature set size of 20.

A range of ML algorithms has been utilized to predict incident duration. Reference [15] evaluated eleven classical ML classifiers, including SVM, K-NN, DT, and GPC. Their results indicated that SVM, K-NN, and GPC achieved the highest accuracy (97%); however, DT, while less accurate, excelled in computational efficiency. Reference [183] compared the performance of six ML models, including MLR, DT, SVR, K-NN, ensemble models (e.g., SVR bagging and gradient boosting), and ANN. They found that SVR, KNN, and MLR were most sensitive to feature optimization techniques.

Despite advancements, challenges remain in accurately predicting incident duration. Data limitations, including inconsistencies and missing values, pose significant challenges. Reference [184] had to remove over 60% of their initial dataset due to inconsistencies, emphasizing the need for more comprehensive and standardized data collection. The study by [185] highlights the issue of imbalanced datasets, where nonincident records significantly outnumber incident records, impacting the performance of ML models. To address this challenge, the researchers implemented a two-stage framework using XGBoost with resampling techniques (random oversampling, weighting, and SMOTE), achieving high accuracy in both incident detection and multicategory classification. While some models found to exhibit high accuracy, their computational complexity may hinder real-time implementation.

8. Summary

Traffic accident analysis is crucial for enhancing road safety, reducing traffic congestion, and improving urban infrastructure planning. However, it faces several significant challenges and gaps that need to be addressed to design models with optimal performance. This section summarizes the primary challenges and gaps identified in the reviewed studies, including methodologies and techniques that can be used to address these issues.

8.1. Challenges and Gaps. Traffic accident analysis has the following challenges which we address in this review:

- Traffic accident reports contain multiple characteristics of a traffic accident. Each characteristic can have various effects on accident modeling performance. To solve the problem of determining which data to collect and what details (features), we need to use feature importance estimation methods.
- The task of predicting traffic accident duration group (e.g., predicting if an accident will be short-term or long-term) can create a problem of imbalanced classification due to uneven number of accidents in duration groups. To solve this issue, multiple approaches for imbalanced classification exist including the use of specific models, metrics, and data processing techniques.
- Traffic accident duration distribution usually follows log-normal or log-logistic distribution which is a skewed distribution. ML models show better performance with normally distributed predicted variables. Therefore, the target variable needs to be processed to enhance predictions.
- Accident reports also may contain reporting-specific errors or anomalies in reporting. In general, the outlier removal procedure improves the performance of ML models.
- Accident reports can represent very large datasets with high dimensionality. For example, CTADS contains 1.5 million accident reports, 49 features each. When working with high-dimensionality data, it may be necessary to use dimensionality reduction to reduce the model training time and memory requirements.
- In particular cases, data availability on accidents can be low. In that case, we need to seek ways to improve the extrapolation performance of our models. The extrapolation ability and noise resistance of ML models can be improved by using model ensembles and bilevel frameworks.
- The specifics of accident report datasets are the presence of textual accident descriptions, which can contain valuable information to enhance the prediction performance of accident modeling. Various natural language processing (NLP) techniques are of high importance for the task of utilizing accident description.
- Historical traffic speed or traffic flow data can be available for accident reports. In this case, time-series modeling techniques can be utilized to extract useful information to enhance the performance of accident modeling.

- One of the main challenges in incident identification from traffic flow is to provide descriptive statistics for abrupt changes in traffic state [27].
- There are multiple novel ML and DL models, which have not been used in traffic accident modeling so far, but many hybrid or advanced frameworks can still be applied to enhance the performance of the incident duration modeling.

Traffic jam identification methods (which rely on vehicle detector data) can be further extended from the use of algorithms to the use of ML methods intended for time-series processing. Further research on the topic of incident-related traffic state identification can be performed using merged datasets of traffic accident reports and traffic states (flow, speed, and occupancy) recorded in their proximity [27].

The use of anomaly detection methods in urban traffic data was found to be seldom [158] which implies that future research can be performed in that direction. The use of ML methods in transportation was found not being used to its full advantage [186]: Seventy-four percent of papers were found to be relying on prediction methods like XGBoost, RF, LSTM, and MLP with only minimal use of sophisticated ML methods.

The main limitations of using DL models for traffic incident duration modeling are as follows: (1) data availability: DL models require large volumes of data for training, which can be difficult to access or cannot be provided by the traffic management authorities due to privacy or security concerns; (2) data quality: DL models are sensitive to data quality, including outliers, missing values, and user-input errors (like incorrect labeling, misreported incident duration); and (3) interpretability: DL models are often represented as "black-box" models (which means that relationships developed between inputs and outputs are hidden inside the model) [187], making it hard to understand how the model arriving at a particular result, which may limit the model deployment due to possible model bias toward data. The absence of interpretability is particularly critical since black-box models cannot be considered reliable in traffic safety applications.

8.2. Summary on the Use of ML Techniques. The majority of studies rely on SVMs or NB for traffic incident duration prediction (see Table 10). The use of more advanced methods like XGBoost or GBDT is rare which is surprising given their effectiveness. This can be explained by a generally slow attribution of both sophisticated models and data-driven approaches to the traffic accident research which we observe in the literature. Also, the complexity of ML pipelines has increased in recent years due to the need for the incorporation of more data science knowledge to merge with traditional transport modeling techniques.

Another finding is that various ML pipeline elements (like dimensionality reduction or feature selection) are rarely used across incident duration prediction studies (see Table 11). This reflects a lack of advanced ML techniques that can be explored for incident modeling. The lack of popularity among these pipeline approaches leaves room for more innovative ideas for data filtering and feature ranking from the beginning of the incident duration prediction modeling. The use of SHAP and feature selection is found to be lacking in studies, while it may provide a list of entries in the incident reporting form with the highest contribution to the accuracy of the traffic incident duration prediction. An additional description of the most important features may provide a further increase in prediction accuracy. Bilevel frameworks allow the separation of the task of the incident duration prediction into incident duration classification and regression tasks. Allocating different kinds of models to each task may improve the overall model performance. The use of bilevel frameworks as a technique for prediction performance improvement is also found to be rare. Dimensionality reduction had a low relevance in multiple prior studies due to the small size and low dimensionality of incident reports (see the table of dataset sizes used in years prior to 2018 [7]), but with current advancements in traffic data collection (see Section 2) and significant consequent increase in amount and variety of data collected, we see a rise in the relevance of data preprocessing methods.

8.3. Implications of AI Developments on Traffic Accident Analysis. The accelerated progress in the area of AI profoundly impacts various domains, including the analysis of traffic accidents. AI's capacity to handle large datasets, forecast hazardous situations, and augment real-time responses can markedly transform our approach toward understanding, preventing, and addressing traffic accidents in the near future. There are multiple ways AI is reshaping the arena of traffic accident analysis and prevention:

- Enhanced Data Collection and Analysis, growing amounts of data: The development of computer hardware and software in recent years has greatly improved the collection, processing, and analysis of traffic data. As a consequence, the amount and dimensionality of data are rapidly growing. AI systems can benefit from large volumes of data. These methods can also process data from various sources such as videos from traffic cameras, GPS devices, high-dimensional geospatial data, and messages on social media platforms. This may help to identify accident patterns, contributing factors, and high-risk areas more efficiently, enabling targeted interventions to reduce the frequency and severity of accidents in identified areas.
- Real-time Monitoring and Response: Traffic surveillance systems (including video data from traffic cameras) can utilize AI methods for monitor traffic conditions in real-time, detecting incidents as they occur. This enables quicker RTs from emergency services, potentially saving lives and reducing the severity of injuries.
- Enhanced Accident Analysis: AI can assist in reconstructing traffic accidents more accurately by analyzing data from multiple sources simultaneously, such as

Table 10: Most popular machine learning methods used across reviewed papers.

Method	Studies
Random forest	[95, 116, 188]
XGBoost	[23, 40, 100, 111, 189]
Support vector machine (SVM)	[54, 62, 82, 86, 87, 89, 90, 95, 102, 103, 111, 112, 115, 116, 188, 190, 191]
Linear regression	[73, 82, 90, 97, 103, 111, 114–116, 188, 190, 194–197]
Naive Bayes	[92, 101, 110, 114, 192, 193]
Decision tree	[62, 73, 91, 118]
Gradient-boosted decision trees	[23, 40, 100]
K-nearest neighbors	[194, 196]

Table 11: Most popular machine learning pipeline elements used across reviewed papers.

Method	Studies
Principal component analysis (PCA)	[121, 157]
Latent discriminant analysis (LDA)	[89, 92, 106]
SHapley Additive exPlanations (SHAP)	[40, 108, 189]
Feature selection	[93, 188, 198]
Clustering	[87, 108, 112, 196]
Ensemble	[62, 116]
Bilevel frameworks	[40, 114]

vehicle sensors, traffic cameras, POIs in proximity, accident reports, traffic speed, and flow readings from nearby traffic detectors. This can help identify the highly complex patterns of related factors leading to the accident affecting the design of future safety measures.

- Preaccident Analysis: Traffic authorities can leverage ML algorithms to anticipate potential accidents before they occur. By analyzing data from diverse sources such as weather reports, historical accident data, and real-time traffic conditions, AI can predict high-risk areas and the likelihood of an accident in a specific time frame. These predictive insights could help authorities implement preventive measures and drivers adjust their routes or driving behavior.
- Accident-Related Medical Assistance Prediction: AI can also play a crucial role in predicting necessary medical assistance. By analyzing accident report data (on the moment of report), AI can assist in predicting the likelihood of certain injuries based on the nature of the accident, traffic state readings in proximity. This can help emergency responders and hospitals prepare more effectively for incoming patients, potentially improving outcomes.
- Enhanced Emergency Services Dispatch: AI systems can aid in the swift and efficient dispatch of emergency services in the event of an accident. By analyzing data from the accident, like the severity of the crash, the number of vehicles involved, and the potential number of injuries, AI can make informed recommendations about the resources required at the scene.
- Enhanced Road Design and Planning: Data-driven insights from AI can help urban planners and engineers design safer roads (e.g., by manipulating the road design in software and estimating possible accident occurrence rate and predicted accident severity outcomes using AI model). By analyzing accident data, AI can identify problematic road designs that lead to higher accident rates.
- Economic Implications: Reducing the number of traffic accidents through AI could have significant economic implications. Fewer accidents mean less money spent on healthcare, vehicle repairs, and insurance claims.

In conclusion, AI has the potential to enhance accident-related data collection and analysis, real-time response (of both clearance and medical assistance teams), accident pattern analysis, and road design assessment, and result in economic savings.

8.4. Future Research Directions in Incident Modeling. The application of traditional clustering methods can provide insights on spatial–temporal patterns and hot spots of traffic accidents [29]. The evolution of cluster size over time can provide valuable insight into contributing factors, which lead to accident hot-spot appearance, disappearance, growth, and decline. The procedure of accident hot-spot detection and their evolution prediction can provide important information to traffic management authorities. Runtime performance of the online algorithms to find subtrajectory outliers (which may help to detect accidents in real-time) was found to be low and may require the development of more efficient methods [158].

The following topics of future research can be addressed:

- Dataset integration and fusion models: There are various datasets which exist adjacent in time and space to incident reports like PeMS, which contains data on traffic flow, speed, and occupancy in the proximity of traffic accident reports from the CTADS dataset, which may be used to enhance traffic incident duration modeling. The availability of multiple datasets of different types that describe the accident may require the use of data fusion models [199] and/or feature embedding methods [100].
- The use of textual data: The textual incident report can contain information unconstrained to reporting form, which may be used to enhance the incident duration prediction accuracy. This approach requires knowledge and techniques from the field of NLP.

- The use of sophisticated ML and DL models: Multiple studies on traffic crashes indicate nonlinear relationships and threshold effects between independent variables and dependent variables [189, 200]. Also, it was highlighted by one of the previous reviews that this study area relies mostly on classical ML models.
- The use of sophisticated ML pipeline elements like anomaly detection, hyperparameter optimization, and dimensionality reduction: Oversampling and undersampling found to be rarely used and may enhance the incident duration prediction performance.
- The effect of combined use of traffic incident duration prediction with VMSs on traffic flow can be studied further to assess the study the potential of VMS use for traffic incident impact mitigation [112].
- Further study on real-time incident reporting data can be performed using PeMS dataset, which includes a timeline of textual incident description availability over time.
- The use of feature importance estimation techniques is to assess the impact of specific reported values on the incident duration prediction accuracy. In particular, the effect of weather conditions on incident duration can be studied [94].
- The requirement of extrapolation tests: The deployment of the incident duration prediction model requires multiple considerations like model bias for data or time/space extrapolation performance [94].
- The rise in relevance of advanced data preprocessing methods: In particular, dimensionality reduction techniques become more relevant due to the significant increase in the amount and variety of data collected across traffic networks.

Traffic incident duration prediction is a complex and important task, which may benefit from further research with the use of sophisticated models of AI. Intelligent models, such as those based on ML, can provide predictions with high accuracy. The use of these models can help traffic management authorities to improve traffic flow and reduce the impact of traffic incidents. Further research is needed to improve the accuracy of these models, such as dataset integration, complex and hybrid ML and DL models, the use of textual data, anomaly detection, and hyperparameter optimization. This research has the potential to enhance traffic incident duration prediction performance, ultimately leading to improved traffic flow and reduced impact from traffic incidents.

9. Conclusion

Today, we can use several data-driven approaches for solving the incident duration prediction task. However, each of the above topics presents several challenging aspects which can range from data collection, preparation, anomaly detection, up to strategies of choosing the best performing model or variables, techniques to model, and estimate the total duration of traffic incidents across the network and regions. This motivated our current literature review which was organized to provide the reader with understanding of the complexities of each modeling step.

Data Availability Statement

This research is a literature review and did not involve the generation or analysis of any datasets. Therefore, no data are associated with this study.

Disclosure

This manuscript is extension of the research presented in the thesis “Advanced AI Techniques for Comprehensive Traffic Incident Analysis: Enhancing Incident Duration Prediction and Accident Risk Forecasting,” [201] available via Open Publications of University of Technology Sydney Scholars (OPUS). The foundational work in the thesis has been further developed in this study.

Conflicts of Interest

Adriana-Simona Mihaita reports financial support that was provided by Australian Research Council. Artur Grigorev reports a relationship with iMOVE CRC that includes funding grants. Fang Chen declares no conflicts of interest.

Funding

This work has been done as part of the ARC Linkage Project LP180100114. This research is funded by iMOVE CRC (project 5-013) and supported by the Cooperative Research Centres program, an Australian Government initiative.

Acknowledgments

This work has been done as part of the ARC Linkage Project LP180100114. This research is funded by iMOVE CRC (project 5-013) and supported by the Cooperative Research Centres program, an Australian Government initiative.

References

[1] R. Abduljabbar, H. Dia, S. Liyanage, and S. A. Bagloee, “Applications of Artificial Intelligence in Transport: An Overview,” *Sustainability* 11, no. 1 (2019): 189, https://doi.org/10.3390/su11010189.

[2] M. Machin, J. Sanguesa, P. Garrido, and F. Martinez, “On the Use of Artificial Intelligence Techniques in Intelligent Transportation Systems,” *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)* (2018): 332–337, https://doi.org/10.1109/WCNCW.2018.8369029.

[3] D. Schrank and T. Lomax, *The 2002 Urban Mobility Report (College Station, Tx: Texas Transportation Institute, texas A&m University, June)* (2002).

[4] M. A. M. A. Al-Bordiny, “Function, Requirements and Applications of Intelligent Transportation Systems (ITS),” (Faculty of Engineering, Tanta University, 2014), Ph.D. thesis.

[5] Y. Ma, Z. Wang, H. Yang, and L. Yang, “Artificial Intelligence Applications in the Development of Autonomous Vehicles: A Survey,” *IEEE/CAA Journal of Automatica Sinica*

7, no. 2 (2020): 315–329, https://doi.org/10.1109/jas.2020.1003021.

[6] A. Benterki, M. Boukhnifer, V. Judalet, and C. Maaoui, "Artificial Intelligence for Vehicle Behavior Anticipation: Hybrid Approach Based on Maneuver Classification and Trajectory Prediction," *IEEE Access* 8 (2020): 56992–57002, https://doi.org/10.1109/access.2020.2982170.

[7] R. Li, F. C. Pereira, and M. E. Ben-Akiva, "Overview of Traffic Incident Duration Analysis and Prediction," *European Transport Research Review* 10, no. 2 (2018): 22, https://doi.org/10.1186/s12544-018-0300-1.

[8] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and t. Prisma Group, "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the Prisma Statement," *Annals of Internal Medicine* 151, no. 4 (2009): 264–269, https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

[9] A. Liberati, D. G. Altman, J. Tetzlaff, et al., "The Prisma Statement for Reporting Systematic Reviews and Meta-Analyses of Studies that Evaluate Health Care Interventions: Explanation and Elaboration," *Annals of Internal Medicine* 151, no. 4 (2009): W65–W94, https://doi.org/10.7326/0003-4819-151-4-200908180-00136.

[10] A. Liaw and M. Wiener, et al., "Classification and Regression by Randomforest," *R News* 2 (2002): 18–22.

[11] T. Chen and C. Guestrin, "Xgboost: A Scalable Tree Boosting System," in *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining* (2016), 785–794, https://doi.org/10.1145/2939672.2939785.

[12] J. Tang, L. Zheng, C. Han, et al., "Statistical and Machine-Learning Methods for Clearance Time Prediction of Road Incidents: A Methodology Review," *Analytic Methods in Accident Research* 27 (2020): 100123, https://doi.org/10.1016/j.amar.2020.100123.

[13] Z. A. Mohammed, M. N. Abdullah, and I. H. Al-Hussaini, "Review of Incident Duration Prediction Methods," *International Journal of Science and Research* 9 (2020): https://doi.org/10.21275/ART20203779.

[14] G. Valenti, M. Lelli, and D. Cucina, "A Comparative Study of Models for the Incident Duration Prediction," *European Transport Research Review* 2 (2010): 103–111, https://doi.org/10.1007/s12544-010-0031-4.

[15] K. Hamad, L. Obaid, A. B. Nassif, S. Abu Dabous, R. Al-Ruzouq, and W. Zeiada, "Comprehensive Evaluation of Multiple Machine Learning Classifiers for Predicting Freeway Incident Duration," *Innovative Infrastructure Solutions* 8, no. 6 (2023): 177, https://doi.org/10.1007/s41062-023-01138-1.

[16] M. W. Adler, J. v. Ommeren, and P. Rietveld, "Road Congestion and Incident Duration," *Economics of Transportation* 2, no. 4 (2013): 109–118, https://www.sciencedirect.com/science/article/pii/S2212012213000269, https://doi.org/10.1016/j.ecotra.2013.12.003.

[17] H. J. Kim and H. K. Choi, "This Research Was Supported by the Korean Research Institute for Human Settlements (KRIHS)," *IATSS Research* 25, no. 1 (2001): 62–72, https://www.sciencedirect.com/science/article/pii/S0386111214600078, https://doi.org/10.1016/S0386-1112(14)60007-8.

[18] A. M. S. Alkaabi, D. Dissanayake, and R. Bird, "Analyzing Clearance Time of Urban Traffic Accidents in Abu Dhabi, united arab emirates, with Hazard-Based Duration Modeling Method," *Transportation Research Record: Journal of the Transportation Research Board* 2229, no. 1 (2011): 46–54, https://doi.org/10.3141/2229-06.

[19] L. Hou, Y. Lao, Y. Wang, Z. Zhang, Y. Zhang, and Z. Li, "Modeling Freeway Incident Response Time: A Mechanism-Based Approach," *Transportation Research Part C: Emerging Technologies* 28 (2013): 87–100, https://doi.org/10.1016/j.trc.2012.12.005.

[20] A. Tavassoli Hojati, L. Ferreira, S. Washington, P. Charles, and A. Shobeirinejad, "Modelling Total Duration of Traffic Incidents Including Incident Detection and Recovery Time," *Accident Analysis & Prevention* 71 (2014): 296–305, https://www.sciencedirect.com/science/article/pii/S0001457514001791, https://doi.org/10.1016/j.aap.2014.06.006.

[21] X. Zeng and P. Songchitruksa, "Empirical Method for Estimating Traffic Incident Recovery Time," *Transportation Research Record: Journal of the Transportation Research Board* 2178, no. 1 (2010): 119–127, https://doi.org/10.3141/2178-13.

[22] R. Li and P. Shang, "Incident Duration Modeling Using Flexible Parametric Hazard-Based Models," *Computational Intelligence and Neuroscience* 2014 (2014): 1–10, https://doi.org/10.1155/2014/723427.

[23] A. S. Mihaita, Z. Liu, C. Cai, and M. A. Rizoiu, "Arterial Incident Duration Prediction Using a Bi-level Framework of Extreme Gradient-Tree Boosting," *arXiv preprint arXiv:1905.12254* (2019).

[24] J. Lee, T. Yoon, S. Kwon, and J. Lee, "Model Evaluation for Forecasting Traffic Accident Severity in Rainy Seasons Using Machine Learning Algorithms: Seoul City Study," *Applied Sciences* 10, no. 1 (2019): 129, https://doi.org/10.3390/app10010129.

[25] A. Ziakopoulos, "Spatial Analysis of Harsh Driving Behavior Events in Urban Networks Using High-Resolution Smartphone and Geometric Data," *Accident Analysis & Prevention* 157 (2021): 106189, https://doi.org/10.1016/j.aap.2021.106189.

[26] M. Barthélemy, "Spatial Networks," *Physics Reports* 499, no. 1-3 (2011): 1–101, https://doi.org/10.1016/j.physrep.2010.11.002.

[27] T. Liu, Z. Li, P. Liu, C. Xu, and D. A. Noyce, "Using Empirical Traffic Trajectory Data for Crash Risk Evaluation under Three-phase Traffic Theory Framework," *Accident Analysis & Prevention* 157 (2021): 106191, https://doi.org/10.1016/j.aap.2021.106191.

[28] A. Mohammadnazar, I. Mahdinia, N. Ahmad, A. J. Khattak, and J. Liu, "Understanding How Relationships between Crash Frequency and Correlates Vary for Multilane Rural Highways: Estimating Geographically and Temporally Weighted Regression Models," *Accident Analysis & Prevention* 157 (2021): 106146, https://doi.org/10.1016/j.aap.2021.106146.

[29] M. Al Hamami and T. Matisziw, "Measuring the Spatiotemporal Evolution of Accident Hot Spots," *Accident Analysis & Prevention* 157 (2021): 106133, https://doi.org/10.1016/j.aap.2021.106133.

[30] R. J Javid and R. Jahanbakhsh Javid, "A Framework for Travel Time Variability Analysis Using Urban Traffic Incident Data," *IATSS Research* 42, no. 1 (2018): 30–38, https://doi.org/10.1016/j.iatssr.2017.06.003.

[31] D. Nam and F. Mannering, "An Exploratory Hazard-Based Analysis of Highway Incident Duration," *Transportation Research Part A: Policy and Practice* 34, no. 2 (2000): 85–102, https://www.sciencedirect.com/science/article/pii/S0965856498000652, https://doi.org/10.1016/S0965-8564(98)00065-2.

[32] Y. S. Chung, Y. C. Chiou, and C. H. Lin, "Simultaneous Equation Modeling of Freeway Accident Duration and Lanes Blocked," *Analytic Methods in Accident Research* 7 (2015):

16–28, https://www.sciencedirect.com/science/article/pii/S2213665715000275, https://doi.org/10.1016/j.amar.2015.04.003.
[33] H. J. Haule, T. Sando, R. Lentz, C. H. Chuan, and P. Alluri, "Evaluating the Impact and Clearance Duration of Freeway Incidents," *International Journal of Transportation Science and Technology* 8, no. 1 (2019): 13–24, https://www.sciencedirect.com/science/article/pii/S2046043018300522, https://doi.org/10.1016/j.ijtst.2018.06.005.
[34] L. Eboli, C. Forciniti, and G. Mazzulla, "Factors Influencing Accident Severity: an Analysis by Road Accident Type," *Transportation Research Procedia* 47 (2020): 449–456, https://doi.org/10.1016/j.trpro.2020.03.120.
[35] H. M. Hammad, M. Ashraf, F. Abbas, et al., "RETRACTED ARTICLE: Environmental Factors Affecting the Frequency of Road Traffic Accidents: a Case Study of Sub-urban Area of Pakistan," *Environmental Science and Pollution Research* 26, no. 12 (2019): 11674–11685, https://doi.org/10.1007/s11356-019-04752-8.
[36] K. Bucsuhazy, E. Matuchova, R. Zuvala, P. Moravcova, M. Kostíková, and R. Mikulec, "Human Factors Contributing to the Road Traffic Accident Occurrence," *Transportation Research Procedia* 45 (2020): 555–561, https://doi.org/10.1016/j.trpro.2020.03.057.
[37] S. M. Moosavi, S. Gholamzadeh, A. Gandomi, and S. M. Yaghoubi, "Accident Type Detection Using Text Mining of Traffic Accident Reports," *IEEE Access* 7 (2019): 109960–109976.
[38] S. Moosavi, M. H. Samavatian, S. Parthasarathy, and R. Ramnath, *A Countrywide Traffic Accident Dataset* (2019).
[39] Y. Zhao and W. Deng, "Prediction in Traffic Accident Duration Based on Heterogeneous Ensemble Learning," *Applied Artificial Intelligence* 36, no. 1 (2022): 2018643, https://doi.org/10.1080/08839514.2021.2018643.
[40] A. Grigorev, A. S. Mihaita, S. Lee, and F. Chen, "Incident Duration Prediction Using a Bi-level Machine Learning Framework with Outlier Removal and Intra–extra Joint Optimisation," *Transportation Research Part C: Emerging Technologies* 141 (2022): 103721, https://doi.org/10.1016/j.trc.2022.103721.
[41] T. Choe, A. Skabardonis, and P. Varaiya, "Freeway Performance Measurement System: Operational Analysis Tool," *Transportation Research Record: Journal of the Transportation Research Board* 1811, no. 1 (2002): 67–75, https://doi.org/10.3141/1811-08.
[42] V. B. K. Anna, L. S. Bisht, and S. Chand, "Data on Road Traffic Incidents for Sydney Greater Metropolitan Area," *Data in Brief* 51 (2023): 109769, https://data.mendeley.com/datasets/cgnx2cs665/5, https://doi.org/10.1016/j.dib.2023.109769.
[43] National Highway Traffic Safety Administration, "Fatality Analysis Reporting System (Fars)," (2020), https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars.
[44] Bureau of Transportation Statistics, "National Transportation Atlas Database (Ntad)," (2020), https://www.bts.gov/ntad.
[45] European Commission, "Road Safety Atlas," https://ec.europa.eu/transport/road_safety/specialist/library/atlas/index_en.htm.
[46] Uk Government, "Uk Road Safety Statistics," https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.
[47] California Highway Patrol, "Statewide Integrated Traffic Records System (Switrs)," https://www.chp.ca.gov/switrs/.
[48] D. Waetjen and F. Shilling, "Rapid Reporting of Vehicle Crash Data in california to Understand Impacts from Covid-19 Pandemic on Traffic and Incidents" (2021).
[49] Department of Infrastructure Regional Development and Cities, "Australian Road Deaths Database - Ardd," https://www.bitre.gov.au/statistics/safety/fatal_road_crash_database.
[50] World Health Organization, "World Health Organization's Global Health Estimates," https://www.who.int/data/gho/data/indicators/indicator-details/GHO/estimated-number-of-road-traffic-deaths.
[51] Compass IoT, "Compass Iot," https://www.compassiot.com.au/.
[52] TomTom, "Tomtom," https://www.tomtom.com/.
[53] M. Treiber and A. Kesting, "Traffic Flow Dynamics," *Traffic Flow Dynamics: Data, Models and Simulation* (Springer-Verlag Berlin Heidelberg, 2013).
[54] X. Huang, P. He, A. Rangarajan, and S. Ranka, "Intelligent Intersection: Two-Stream Convolutional Networks for Real-Time Near-Accident Detection in Traffic Video," *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 6, no. 2 (2020): 1–28, https://doi.org/10.1145/3373647.
[55] R. Schindler and G. Bianchi Piccinini, "Truck Drivers Behavior in Encounters with Vulnerable Road Users at Intersections: Results from a Test-Track Experiment," *Accident Analysis & Prevention* 159 (2021): 106289, https://doi.org/10.1016/j.aap.2021.106289.
[56] B. Dadashova, B. A. Ramirez, J. M. McWilliams, and F. A. Izquierdo, "The Identification of Patterns of Interurban Road Accident Frequency and Severity Using Road Geometry and Traffic Indicators," *Transportation Research Procedia* 14 (2016): 4122–4129, https://doi.org/10.1016/j.trpro.2016.05.383.
[57] M. H. Islam, L. T. Hua, H. Hamid, and A. Azarkerdar, "Relationship of Accident Rates and Road Geometric Design," in *IOP Conference Series: Earth and Environmental Science* (IOP Publishing, 2019), 012040.
[58] S. Cafiso, A. Di Graziano, G. Di Silvestro, G. La Cava, and B. Persaud, "Development of Comprehensive Accident Models for Two-Lane Rural Highways Using Exposure, Geometry, Consistency and Context Variables," *Accident Analysis & Prevention* 42, no. 4 (2010): 1072–1079, https://www.sciencedirect.com/science/article/pii/S0001457509003297, https://doi.org/10.1016/j.aap.2009.12.015.
[59] O. Contributors, "Openstreetmap: The Free Wiki World Map," (2024), https://www.openstreetmap.org.
[60] M. G. Karlaftis and I. Golias, "Effects of Road Geometry and Traffic Volumes on Rural Roadway Accident Rates," *Accident Analysis & Prevention* 34, no. 3 (2002): 357–365, https://doi.org/10.1016/s0001-4575(01)00033-1.
[61] Uk Ordnance Survey, "OS National Geographic Database (NGD) OS NGD API," (2024), https://www.ordnancesurvey.co.uk/products/os-ngd-api-features.
[62] K. Hamad, M. A. Khalil, and A. R. Alozi, "Predicting Freeway Incident Duration Using Machine Learning," *International Journal of Intelligent Transportation Systems Research* 18, no. 2 (2020): 367–380, https://doi.org/10.1007/s13177-019-00205-1.
[63] R. Prieto Curiel, H. González Ramírez, and S. R. Bishop, "A Novel Rare Event Approach to Measure the Randomness and Concentration of Road Accidents," *PLoS One* 13, no. 8 (2018): e0201890, https://doi.org/10.1371/journal.pone.0201890.
[64] E. C. Sullivan, "New Model for Predicting Freeway Incidents and Incident Delays," *Journal of Transportation Engineering* 123, no. 4 (1997): 267–275, https://doi.org/10.1061/(asce)0733-947x(1997)123:4(267).
[65] Y. Chung, L. Walubita, and K. Choi, "Modeling Accident Duration and its Mitigation Strategies on South Korean

Freeway Systems," *Transportation Research Record Journal of the Transportation Research Board* 2178, no. 1 (2010): 49–57, https://doi.org/10.3141/2178-06.

[66] K. Smith and B. Smith, "Forecasting the Clearance Time of Freeway Accidents Final Report of ITS Center Project: Incident Duration Forecasting," *Technical Report* (2001).

[67] Y. Zou, K. Henrickson, D. Lord, Y. Wang, and K. Xu, "Application of Finite Mixture Models for Analysing Freeway Incident Clearance Time," *Transportmetrica: Transportation Science* 12, no. 2 (2016): 99–115, https://doi.org/10.1080/23249935.2015.1102173.

[68] R. Li, F. C. Pereira, and M. E. Ben-Akiva, "Competing Risks Mixture Model for Traffic Incident Duration Prediction," *Accident Analysis & Prevention* 75 (2015): 192–201, https://doi.org/10.1016/j.aap.2014.11.023.

[69] H. Yang, K. Ozbay, K. Xie, and Y. Ma, *Development of an Automated Approach for Quantifying Spatio-Temporal Impact of Traffic Incidents*.

[70] X. Mao, C. Yuan, J. Gan, and S. Zhang, "Risk Factors Affecting Traffic Accidents at Urban Weaving Sections: Evidence from china," *International Journal of Environmental Research and Public Health* 16, no. 9 (2019): 1542, https://doi.org/10.3390/ijerph16091542.

[71] K. Ozbay and P. Kachroo, *Incident Management in Intelligent Transportation Systems* (1999).

[72] W. Wang, H. Chen, and M. Bell, "A Study of the Characteristics of Traffic Incident Duration on Motorways," in *Traffic and Transportation Studies* (2002), 1101–1108.

[73] K. Ozbay and N. Noyan, "Estimation of Incident Clearance Times Using Bayesian Networks Approach," *Accident Analysis & Prevention* 38, no. 3 (2006): 542–555, https://doi.org/10.1016/j.aap.2005.11.012.

[74] P. Barcellos, C. Bouvié, F. L. Escouto, and J. Scharcanski, "A Novel Video Based System for Detecting and Counting Vehicles at User-Defined Virtual Loops," *Expert Systems with Applications* 42, no. 4 (2015): 1845–1856, https://doi.org/10.1016/j.eswa.2014.09.045.

[75] S. Alkheder, M. Taamneh, and S. Taamneh, "Severity Prediction of Traffic Accident Using an Artificial Neural Network," *Journal of Forecasting* 36, no. 1 (2017): 100–108, https://doi.org/10.1002/for.2425.

[76] X. Ma, C. Ding, S. Luan, Y. Wang, and Y. Wang, "Prioritizing Influential Factors for Freeway Incident Clearance Time Prediction Using the Gradient Boosting Decision Trees Method," *IEEE Transactions on Intelligent Transportation Systems* 18, no. 9 (2017): 2303–2310, https://doi.org/10.1109/tits.2016.2635719.

[77] B. Yu and Z. Xia, "A Methodology for Freeway Incident Duration Prediction Using Computerized Historical Database," *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient* (2012), 3463–3474, https://doi.org/10.1061/9780784412442.351.

[78] A. Hojati, L. Ferreira, P. Charles, and M. Kabit, "Analysing Freeway Traffic Incident Duration Using an Australian Data Set," *Road and Transport Research* 21 (2012): 16–28.

[79] C. Zhan, A. Gan, and M. Hadi, "Prediction of Lane Clearance Time of Freeway Incidents Using the M5p Tree Algorithm," *IEEE Transactions on Intelligent Transportation Systems* 12, no. 4 (2011): 1549–1557, https://doi.org/10.1109/TITS.2011.2161634.

[80] J. de Oña, R. O. Mujalli, and F. J. Calvo, "Analysis of Traffic Accident Injury Severity on Spanish Rural Highways Using Bayesian Networks," *Accident Analysis & Prevention* 43, no. 1 (2011): 402–411, https://doi.org/10.1016/j.aap.2010.09.010.

[81] F. Zong, H. Xu, and H. Zhang, "Prediction for Traffic Accident Severity: Comparing the Bayesian Network and Regression Models," *Mathematical Problems in Engineering* 2013 (2013): 1–9, https://doi.org/10.1155/2013/475194.

[82] H. Park, A. Haghani, and X. Zhang, "Interpretation of Bayesian Neural Networks for Predicting the Duration of Detected Incidents," *Journal of Intelligent Transportation Systems* 20, no. 4 (2016): 385–400, https://doi.org/10.1080/15472450.2015.1082428.

[83] A. Das and P. Rad, "Opportunities and Challenges in Explainable Artificial Intelligence (Xai): A Survey," (2020), https://deepai.org/publication/opportunities-and-challenges-in-explainable-artificial-intelligence-xai-a-survey.

[84] A. Barredo Arrieta, N. Diaz-Rodriguez, J. Del Ser, et al., "Explainable Artificial Intelligence (Xai): Concepts, Taxonomies, Opportunities and Challenges toward Responsible Ai," *Information Fusion* 58 (2020): 82–115, https://doi.org/10.1016/j.inffus.2019.12.012.

[85] J. Li, X. Zhu, C. Wang, and Q. Zhu, "An Overview of Graph Embedding: Problems, Techniques and Applications," *ACM Transactions on Knowledge Discovery from Data* 12 (2018): 1–36.

[86] B. Yu, Y. Wang, J. Yao, and J. Wang, "A Comparison of the Performance of Ann and Svm for the Prediction of Traffic Accident Duration," *Neural Network World* 26, no. 3 (2016): 271–287, https://doi.org/10.14311/nnw.2016.26.015.

[87] B. Ghosh, M. T. Asif, J. Dauwels, W. Cai, H. Guo, and U. Fastenrath, "Predicting the Duration of Non-recurring Road Incidents by Cluster-specific Models," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2016), 1522–1527.

[88] C. H. Wei and Y. Lee, "Sequential Forecast of Incident Duration Using Artificial Neural Network Models," *Accident Analysis & Prevention* 39, no. 5 (2007): 944–954, https://doi.org/10.1016/j.aap.2006.12.017.

[89] Q. Shang, T. Xie, and Y. Yu, "Prediction of Duration of Traffic Incidents by Hybrid Deep Learning Based on Multi-Source Incomplete Data," *International Journal of Environmental Research and Public Health* 19, no. 17 (2022): 10903, https://doi.org/10.3390/ijerph191710903.

[90] J. Tang, L. Zheng, C. Han, et al., "Statistical and Machine-Learning Methods for Clearance Time Prediction of Road Incidents: A Methodology Review," *Analytic Methods in Accident Research* 27 (2020): 100123, https://doi.org/10.1016/j.amar.2020.100123.

[91] S. Wang, R. Li, and M. Guo, "Application of Nonparametric Regression in Predicting Traffic Incident Duration," *Transport* 33, no. 1 (2015): 22–31, https://doi.org/10.3846/16484142.2015.1004104.

[92] F. C. Pereira, F. Rodrigues, and M. Ben-Akiva, "Text Analysis in Incident Duration Prediction," *Transportation Research Part C: Emerging Technologies* 37 (2013): 177–192, https://doi.org/10.1016/j.trc.2013.10.002.

[93] Y. Lee and C. H. Wei, "A Computerized Feature Selection Method Using Genetic Algorithms to Forecast Freeway Accident Duration Times," *Computer-Aided Civil and Infrastructure Engineering* 25, no. 2 (2010): 132–148, https://doi.org/10.1111/j.1467-8667.2009.00626.x.

[94] K. Hamad, R. Al-Ruzouq, W. Zeiada, S. Abu Dabous, and M. A. Khalil, "Predicting Incident Duration Using Random Forests," *Transportmetrica: Transport Science* 16, no. 3 (2020): 1269–1293, https://doi.org/10.1080/23249935.2020.1733132.

[95] Z. A. Mohammed, M. N. Abdullah, and I. H. Al Hussaini, "Predicting Incident Duration Based on Machine Learning Methods. Iraqi Journal of Computers, Communications," *Control and Systems Engineering* 21 (2021): 1–15.

[96] H. Al-Najada and I. Mahgoub, "Real-time Incident Clearance Time Prediction Using Traffic Data from Internet of Mobility Sensors," in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing* (IEEE, 2017), 728–735.

[97] Y. Zou, B. Lin, X. Yang, L. Wu, M. Muneeb Abid, and J. Tang, "Application of the Bayesian Model Averaging in Analyzing Freeway Traffic Incident Clearance Time for Emergency Management," *Journal of Advanced Transportation* 2021 (2021): 1–9, https://doi.org/10.1155/2021/6671983.

[98] R. Li, "Traffic Incident Duration Analysis and Prediction Models Based on the Survival Analysis Approach," *IET Intelligent Transport Systems* 9, no. 4 (2015): 351–358, https://doi.org/10.1049/iet-its.2014.0036.

[99] X. Li, J. Liu, A. Khattak, and S. Nambisan, "Sequential Prediction for Large-Scale Traffic Incident Duration: Application and Comparison of Survival Models," *Transportation Research Record: Journal of the Transportation Research Board* 2674, no. 1 (2020): 79–93, https://doi.org/10.1177/0361198119899041.

[100] A. Grigorev, A. S. Mihăiţă, K. Saleh, and M. Piccardi, "Traffic Incident Duration Prediction via a Deep Learning Framework for Text Description Encoding," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)* (IEEE, 2022), 1770–1777.

[101] H. Ozen, et al., "Multi-step Approach to Improving Accuracy of Incident Duration Estimation: Case Study of Istanbul," *Tehnički Vjesnik* 26 (2019): 1777–1783.

[102] Y. Lin and R. Li, "Real-time Traffic Accidents Post-impact Prediction: Based on Crowdsourcing Data," *Accident Analysis & Prevention* 145 (2020): 105696, https://doi.org/10.1016/j.aap.2020.105696.

[103] B. Ghosh and J. Dauwels, "Comparison of Different Bayesian Methods for Estimating Error Bars with Incident Duration Prediction," *Journal of Intelligent Transportation Systems* 26, no. 4 (2022): 420–431, https://doi.org/10.1080/15472450.2021.1894936.

[104] B. N. Araghi, S. Hu, R. Krishnan, M. Bell, and W. Ochieng, "A Comparative Study of K-Nn and Hazard-Based Models for Incident Duration Prediction," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (2014), 1608–1613, https://doi.org/10.1109/itsc.2014.6957923.

[105] B. Ghosh, M. T. Asif, J. Dauwels, U. Fastenrath, and H. Guo, "Dynamic Prediction of the Incident Duration Using Adaptive Feature Set," *IEEE Transactions on Intelligent Transportation Systems* 20, no. 11 (2019): 4019–4031, https://doi.org/10.1109/tits.2018.2878637.

[106] R. Li, F. C. Pereira, and M. E. Ben-Akiva, "Competing Risk Mixture Model and Text Analysis for Sequential Incident Duration Prediction," *Transportation Research Part C: Emerging Technologies* 54 (2015): 74–85, https://doi.org/10.1016/j.trc.2015.03.009.

[107] Y. Zou, X. Ye, K. Henrickson, J. Tang, and Y. Wang, "Jointly Analyzing Freeway Traffic Incident Clearance and Response Time Using a Copula-Based Approach," *Transportation Research Part C: Emerging Technologies* 86 (2018): 171–182, https://www.sciencedirect.com/science/article/pii/S0968090X17303108, https://doi.org/10.1016/j.trc.2017.11.004.

[108] K. Kalair and C. Connaughton, "Dynamic and Interpretable Hazard-Based Models of Traffic Incident Durations," *Frontiers in future transportation* 2 (2021): 669015, https://doi.org/10.3389/ffutr.2021.669015.

[109] R. Zahid Reza and S. S. Pulugurtha, "Forecasting Short-Term Relative Changes in Travel Time on a Freeway," *Case Studies on Transport Policy* 7, no. 2 (2019): 205–217, https://doi.org/10.1016/j.cstp.2019.03.008.

[110] L. Lin, Q. Wang, and A. W. Sadek, "A Combined M5p Tree and Hazard-Based Duration Model for Predicting Urban Freeway Traffic Accident Durations," *Accident Analysis & Prevention* 91 (2016): 114–126, https://doi.org/10.1016/j.aap.2016.03.001.

[111] J. Tang, L. Zheng, C. Han, F. Liu, and J. Cai, "Traffic Incident Clearance Time Prediction and Influencing Factor Analysis Using Extreme Gradient Boosting Model," *Journal of Advanced Transportation* 2020 (2020): 1–12, https://doi.org/10.1155/2020/6401082.

[112] B. Ghosh, "Predicting the Duration and Impact of the Non-recurring Road Incidents on the Transportation Network," (Singapore: Nanyang Technological University, 2019), Ph.D. Thesis.

[113] Y. Chung, "Development of an Accident Duration Prediction Model on the Korean Freeway Systems," *Accident Analysis & Prevention* 42, no. 1 (2010): 282–289, https://doi.org/10.1016/j.aap.2009.08.005.

[114] L. Kuang, H. Yan, Y. Zhu, S. Tu, and X. Fan, "Predicting Duration of Traffic Accidents Based on Cost-Sensitive Bayesian Network and Weighted K-Nearest Neighbor," *Journal of Intelligent Transportation Systems* 23, no. 2 (2019): 161–174, https://doi.org/10.1080/15472450.2018.1536978.

[115] Q. Zheng, C. Xu, P. Liu, and Y. Wang, "Investigating the Predictability of Crashes on Different Freeway Segments Using the Real-Time Crash Risk Models," *Accident Analysis & Prevention* 159 (2021): 106213, https://doi.org/10.1016/j.aap.2021.106213.

[116] M. Motamed, "Developing a Real-Time Freeway Incident Detection Model Using Machine Learning Techniques," (2016), https://hdl.handle.net/2152/39746.

[117] W. Zhu, J. Wu, T. Fu, J. Wang, J. Zhang, and Q. Shangguan, "Dynamic Prediction of Traffic Incident Duration on Urban Expressways: A Deep Learning Approach Based on Lstm and Mlp," *Journal of intelligent and connected vehicles* 4, no. 2 (2021): 80–91, https://doi.org/10.1108/jicv-03-2021-0004.

[118] A. Saracoglu and H. Ozen, "Estimation of Traffic Incident Duration: A Comparative Study of Decision Tree Models," *Arabian Journal for Science and Engineering* 45, no. 10 (2020): 8099–8110, https://doi.org/10.1007/s13369-020-04615-2.

[119] A. J. Khattak, J. L. Schofer, and M. H. Wang, "A Simple Time Sequential Procedure for Predicting Freeway Incident Duration," *Journal of Intelligent Transportation Systems* 2 (1995): 113–138, https://doi.org/10.1080/10248079508903820.

[120] W. Wang, H. Chen, and M. C. Bell, "Vehicle Breakdown Duration Modeling," *Journal of Transportation and Statistics* 8 (2005): 75.

[121] L. L. Wang, H. Y. Ngan, and N. H. Yung, "Automatic Incident Classification for Large-Scale Traffic Data by Adaptive Boosting Svm," *Information Sciences* 467 (2018): 59–73, https://www.sciencedirect.com/science/article/pii/S0020025518305681, https://doi.org/10.1016/j.ins.2018.07.044.

[122] G. B. Huang, D. H. Wang, and Y. Lan, "Extreme Learning Machines: a Survey," *International journal of machine learning and cybernetics* 2 (2011): 107–122, https://doi.org/10.1007/s13042-011-0019-y.
[123] E. Fix and J. Hodges, *Discriminatory Analysis, Nonparametric Discrimination* (1951).
[124] L. Breiman, "Random Forests," *Machine Learning* 45, no. 1 (2001): 5–32, https://doi.org/10.1023/A:1010933404324.
[125] S. R. Gunn, et al., "Support Vector Machines for Classification and Regression," *ISIS technical report* 14 (1998): 5–16.
[126] C. M. Bishop and M. E. Tipping, et al., "Bayesian Regression and Classification," *Nato Science Series sub Series III Computer And Systems Sciences* 190 (2003): 267–288.
[127] M. N. Gibbs, *Bayesian Gaussian Processes for Regression and Classification* (Citeseer, 1998).
[128] C. J. Needham, J. R. Bradford, A. J. Bulpitt, and D. R. Westhead, "A Primer on Learning in Bayesian Networks for Computational Biology," *PLoS Computational Biology* 3, no. 8 (2007): e129, https://doi.org/10.1371/journal.pcbi.0030129.
[129] B. Warner and M. Misra, "Understanding Neural Networks as Statistical Tools," *The American Statistician* 50, no. 4 (1996): 284–293, https://doi.org/10.1080/00031305.1996.10473554.
[130] S. Hochreiter, "The Vanishing Gradient Problem during Learning Recurrent Neural Nets and Problem Solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 06, no. 02 (1998): 107–116, https://doi.org/10.1142/s0218488598000094.
[131] G. Ke, Q. Meng, T. Finley, et al., "Lightgbm: A Highly Efficient Gradient Boosting Decision Tree," *Advances in Neural Information Processing Systems* 30 (2017): 3146–3154.
[132] A. V. Dorogush, V. Ershov, and A. Gulin, "Catboost: Gradient Boosting with Categorical Features Support," (2018), https://arxiv.org/abs/1810.11363.
[133] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme Learning Machine: Theory and Applications," *Neurocomputing* 70, no. 1-3 (2006): 489–501, https://doi.org/10.1016/j.neucom.2005.12.126.
[134] D. D. Margineantu and T. G. Dietterich, "Pruning Adaptive Boosting," in *ICML* (Citeseer, 1997), 211–218.
[135] J. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," *Annals of Statistics* 29, no. 5 (2001): https://doi.org/10.1214/aos/1013203451.
[136] T. Chen, T. He, M. Benesty, et al., "Xgboost: Extreme Gradient Boosting," *R package Version* (2015).
[137] D. Yi, J. Su, C. Liu, M. Quddus, and W. H. Chen, "A Machine Learning Based Personalized System for Driving State Recognition," *Transportation Research Part C: Emerging Technologies* 105 (2019): 241–261, https://doi.org/10.1016/j.trc.2019.05.042.
[138] S. Hong, E. Mondes, C. oh, and G. Ulfarsson, "The Effect of Road Environment Factors on Freeway Traffic Crash Frequency during Daylight, Twilight, and Night Conditions" (2014).
[139] S. Vallejos, D. Alonso, B. Caimmi, L. Berdun, M. G. Armentano, and A. Soria, "Mining Social Networks to Detect Traffic Incidents," *Information Systems Frontiers* 23, no. 1 (2021): 115–134, https://doi.org/10.1007/s10796-020-09994-3.
[140] A. Salas, P. Georgakis, and Y. Petalas, "Incident Detection Using Data from Social Media," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (2017), 751–755, https://doi.org/10.1109/ITSC.2017.8317967.
[141] F. Ali, A. Ali, M. Imran, R. A. Naqvi, M. H. Siddiqi, and K. S. Kwak, "Traffic Accident Detection and Condition Analysis Based on Social Networking Data," *Accident Analysis & Prevention* 151 (2021): 105973, https://www.sciencedirect.com/science/article/pii/S000145752100004X, https://doi.org/10.1016/j.aap.2021.105973.
[142] F. Ali, D. Kwak, P. Khan, et al., "Transportation Sentiment Analysis Using Word Embedding and Ontology-Based Topic Modeling," *Knowledge-Based Systems* 174 (2019): 27–42, https://www.sciencedirect.com/science/article/pii/S0950705119300942, https://doi.org/10.1016/j.knosys.2019.02.033.
[143] D. Cao, S. Wang, and D. Lin, "Chinese Microblog Users' Sentiment-Based Traffic Condition Analysis," *Soft Computing* 22, no. 21 (2018): 7005–7014, https://doi.org/10.1007/s00500-018-3293-8.
[144] H. Lu, Y. Zhu, Y. Yuan, et al., "Social Signal-Driven Knowledge Automation: a Focus on Social Transportation," *IEEE Transactions on Computational Social Systems* 8, no. 3 (2021): 737–753, https://doi.org/10.1109/tcss.2021.3057332.
[145] S. Das, S. Mohanty, and P. Bhattacharyya, "Topical Text Modeling for Incident Detection on Twitter," *Expert Systems with Applications* 122 (2019): 182–195.
[146] Z. Zhang, Z. Chen, and X. Zhu, "Deep Learning Based Incident Detection from Social Media Data," in *2018 IEEE International Conference on Big Data (Big Data)* (IEEE, 2018), 1059–1064.
[147] P. K. Sen and S. K. Dhar, "Stock Price Prediction Using Lstm, Rnn and Cnn-Sliding Window Model," *International Journal of Scientific Engineering and Research* 9 (2018): 2046–2054.
[148] H. Taghipour, A. B. Parsa, R. S. Chauhan, S. Derrible, and A. K. Mohammadian, "A Novel Deep Ensemble Based Approach to Detect Crashes Using Sequential Traffic Data," *IATSS Research* 46, no. 1 (2022): 122–129, https://www.sciencedirect.com/science/article/pii/S0386111221000455, https://doi.org/10.1016/j.iatssr.2021.10.004.
[149] T. Chen, X. Shi, Y. D. Wong, and X. Yu, "Predicting Lane-Changing Risk Level Based on Vehicles' Space-Series Features: A Pre-emptive Learning Approach," *Transportation Research Part C: Emerging Technologies* 116 (2020): 102646, https://doi.org/10.1016/j.trc.2020.102646.
[150] Z. Elamrani Abou Elassad, H. Mousannif, and H. Al Moatassime, "A Real-Time Crash Prediction Fusion Framework: An Imbalance-Aware Strategy for Collision Avoidance Systems," *Transportation Research Part C: Emerging Technologies* 118 (2020): 102708, https://doi.org/10.1016/j.trc.2020.102708.
[151] S. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions," *arXiv preprint arXiv:1705.07874* (2017).
[152] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research* 3 (2003): 1157–1182.
[153] H. Nguyen, C. Cai, and F. Chen, "Automatic Classification of Traffic Incident's Severity Using Machine Learning Approaches," *IET Intelligent Transport Systems* 11, no. 10 (2017): 615–623, https://doi.org/10.1049/iet-its.2017.0051.
[154] L. Li, X. Qu, J. Zhang, and B. Ran, "Traffic Incident Detection Based on Extreme Machine Learning," *Journal of Applied Science and Engineering* 20 (2017): 409–416.
[155] F. T. Liu, K. M. Ting, and Z. H. Zhou, "Isolation Forest," in *2008 Eighth Ieee International Conference on Data Mining, IEEE* (2008), 413–422, https://doi.org/10.1109/icdm.2008.17.

[156] M. Breunig, H. P. Kriegel, R. Ng, and J. Sander, "Lof: Identifying Density-Based Local Outliers," *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (2000): 93–104, https://doi.org/10.1145/342009.335388.

[157] R. Bridgelall and D. D. Tolliver, "Railroad Accident Analysis Using Extreme Gradient Boosting," *Accident Analysis & Prevention* 156 (2021): 106126, https://doi.org/10.1016/j.aap.2021.106126.

[158] Y. Djenouri, A. Belhadi, J. C. W. Lin, D. Djenouri, and A. Cano, "A Survey on Urban Traffic Anomalies Detection Algorithms," *IEEE Access* 7 (2019): 12192–12205, https://doi.org/10.1109/access.2019.2893124.

[159] M. Won, "Outlier Analysis to Improve the Performance of an Incident Duration Estimation and Incident Management System," *Transportation Research Record: Journal of the Transportation Research Board* 2674, no. 5 (2020): 486–497, https://doi.org/10.1177/0361198120916472.

[160] R. Prati, G. Batista, and M. C. Monard, *Data Mining With Imbalanced Class Distributions: Concepts and Methods* (2009).

[161] M. A. Tahir, J. Kittler, and F. Yan, "Inverse Random under Sampling for Class Imbalance Problem and its Application to Multi-Label Classification," *Pattern Recognition* 45, no. 10 (2012): 3738–3750, https://doi.org/10.1016/j.patcog.2012.03.014.

[162] M. Zeng, B. Zou, F. Wei, X. Liu, and L. Wang, "Effective Prediction of Three Common Diseases by Combining Smote with Tomek Links Technique for Imbalanced Medical Data," in *2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS)* (IEEE, 2016), 225–228.

[163] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic Minority Oversampling Technique," *Journal of Artificial Intelligence Research* 16 (2002): 321–357, https://doi.org/10.1613/jair.953.

[164] H. Han, W. Y. Wang, and B. H. Mao, "Borderline-smote: a New Over-sampling Method in Imbalanced Data Sets Learning," in *Advances in Intelligent Computing: International Conference on Intelligent Computing, ICIC 2005, Hefei, China, August 23-26, 2005, Proceedings, Part I 1* (Springer, 2005), 878–887.

[165] H. He, Y. Bai, E. Garcia, and S. Li, "Adasyn: Adaptive Synthetic Sampling Approach for Imbalanced Learning" (2008), 1322–1328, https://doi.org/10.1109/IJCNN.2008.4633969.

[166] Y. Tang, Y. Q. Zhang, N. Chawla, and S. Krasser, "Svms Modeling for Highly Imbalanced Classification. Systems, Man, and Cybernetics, Part B: Cybernetics," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics: A Publication of the IEEE Systems, Man, and Cybernetics Society* 39, no. 1 (2009): 281–288, https://doi.org/10.1109/TSMCB.2008.2002909.

[167] R. E. Schapire, "Explaining Adaboost," in *Empirical Inference* (Springer, 2013), 37–52.

[168] L. Shen and M. Huang, "Data Mining Method for Incident Duration Prediction," in *International Conference on Applied Informatics and Communication* (Springer, 2011), 484–492.

[169] S. Wold, K. Esbensen, and P. Geladi, "Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems* 2, no. 1-3 (1987): 37–52, https://www.sciencedirect.com/science/article/pii/0169743987800849, https://doi.org/10.1016/0169-7439(87)80084-9.

[170] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform Manifold Approximation and Projection for Dimension Reduction," *arXiv preprint arXiv:1802.03426* (2018).

[171] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, Methods, and Applications in Interpretable Machine Learning," *Proceedings of the National Academy of Sciences* 116, no. 44 (2019): 22071–22080, https://www.pnas.org/content/116/44/22071, https://doi.org/10.1073/pnas.1900654116.

[172] R. J. Lewis, "An Introduction to Classification and Regression Tree (Cart) Analysis," in *Annual Meeting of the Society for Academic Emergency Medicine in San Francisco, California* (Citeseer, 2000).

[173] L. Van der Maaten and G. Hinton, "Visualizing Data Using T-Sne," *Journal of Machine Learning Research* 9 (2008).

[174] J. Chen and W. Tao, "Traffic Accident Duration Prediction Using Text Mining and Ensemble Learning on Expressways," *Scientific Reports* 12, no. 1 (2022): 21478, https://www.nature.com/articles/s41598-022-25988-4, https://doi.org/10.1038/s41598-022-25988-4.

[175] Y. Zhao, Z. Ma, H. Peng, and Z. Cheng, "Predicting Metro Incident Duration Using Structured Data and Unstructured Text Logs," *Transportmetrica: Transportation Science* (2024): 1–29, https://doi.org/10.1080/23249935.2024.2396951.

[176] J. Park, J. Lee, and B. Dimitrijevic, "Incident Duration Time Prediction Using Supervised Topic Modeling Method," *Transportation Research Record: Journal of the Transportation Research Board* 2677, no. 2 (2023): 418–430, https://doi.org/10.1177/03611981221106786.

[177] J. Chen, W. Tao, Z. Jing, P. Wang, and Y. Jin, "Traffic Accident Duration Prediction Using Multi-Mode Data and Ensemble Deep Learning," *Heliyon* 10, no. 4 (2024): e25957, https://www.cell.com/heliyon, https://doi.org/10.1016/j.heliyon.2024.e25957.

[178] X. Su, D. Zhi, D. Song, L. Tian, and Y. Yang, "Exploring Weather-Related Factors Affecting the Delay Caused by Traffic Incidents: Mitigating the Negative Effect of Traffic Incidents," *Science of the Total Environment* 877 (2023): 162938, https://www.sciencedirect.com/science/article/pii/S0048969723003120, https://doi.org/10.1016/j.scitotenv.2023.162938.

[179] Q. Zeng, F. Wang, T. Chen, and N. N. Sze, "Incorporating Real-Time Weather Conditions into Analyzing Clearance Time of Freeway Accidents: A Grouped Random Parameters Hazard-Based Duration Model with Time-Varying Covariates," *Analytic Methods in Accident Research* 38 (2023): 100267, https://doi.org/10.1016/j.amar.2023.100267.

[180] Q. Luo and C. Liu, "Exploration of Road Closure Time Characteristics of Tunnel Traffic Accidents: A Case Study in pennsylvania, usa," *Tunnelling and Underground Space Technology* 132 (2023): 104894, https://www.sciencedirect.com/science/article/pii/S088677982200548X, https://doi.org/10.1016/j.tust.2022.104894.

[181] D. Lyu and Y. Lin, "Impact Estimation of Traffic Accident Duration Based on Survival Analysis by Using Field Urban Traffic Condition," in *Proceedings of the GBCESC 2022* (Singapore: Springer Nature Singapore Pte Ltd, 2023), 1247–1254, https://doi.org/10.1007/978-981-19-5217-3_128.

[182] A. Zhang, F. Meng, W. Gong, Y. Zeng, L. Yang, and D. Yuan, "Clearance Time Prediction of Traffic Accidents: A Case Study in Shandong, china," *Australasian Journal of Disaster and Trauma Studies* 26 (2022): 185–186, https://trauma.massey.ac.nz/issues/2022-IS/AJDTS_26_IS_Zhang.pdf.

[183] L. Obaid, K. Hamad, M. A. Khalil, and A. B. Nassif, "Effect of Feature Optimization on Performance of Machine Learning Models for Predicting Traffic Incident Duration," *Engineering Applications of Artificial Intelligence* 131 (2024): 107845, https://www.sciencedirect.com/science/article/pii/S0952197624000017, https://doi.org/10.1016/j.engappai.2024.107845.

[184] L. Yang, R. Corbally, and A. Malekjafarian, "Estimating the Duration of Motorway Incidents Using Machine Learning Approaches," *Transportation Research Procedia* 72 (2023): 4119–4126, https://www.sciencedirect.com/science/article/pii/S235214652300364X, https://doi.org/10.1016/j.trpro.2023.11.364.

[185] M. Xu, H. Liu, and H. Yang, "Ensemble Learning Based Approach for Traffic Incident Detection and Multi-Category Classification," *Engineering Applications of Artificial Intelligence* 132 (2024): 107933, https://doi.org/10.1016/j.engappai.2024.107933.

[186] H. Behrooz and Y. M. Hayeri, "Machine Learning Applications in Surface Transportation Systems: A Literature Review," *Applied Sciences* 12, no. 18 (2022): 9156, https://doi.org/10.3390/app12189156.

[187] J. D. Olden and D. A. Jackson, "Illuminating the "Black Box": a Randomization Approach for Understanding Variable Contributions in Artificial Neural Networks," *Ecological Modelling* 154, no. 1-2 (2002): 135–150, https://www.sciencedirect.com/science/article/pii/S0304380002000649, https://doi.org/10.1016/S0304-3800(02)00064-9.

[188] Q. Shang, D. Tan, S. Gao, and L. Feng, "A Hybrid Method for Traffic Incident Duration Prediction Using Boa-Optimized Random Forest Combined with Neighborhood Components Analysis," *Journal of Advanced Transportation* 2019 (2019): 1–11, https://doi.org/10.1155/2019/4202735.

[189] A. B. Parsa, A. Movahedi, H. Taghipour, S. Derrible, and A. K. Mohammadian, "Toward Safer Highways, Application of Xgboost and Shap for Real-Time Accident Detection and Feature Analysis," *Accident Analysis & Prevention* 136 (2020): 105405, https://doi.org/10.1016/j.aap.2019.105405.

[190] M. Hossain, M. Abdel-Aty, M. A. Quddus, Y. Muromachi, and S. N. Sadeek, "Real-time Crash Prediction Models: State-Of-The-Art, Design Pathways and Ubiquitous Requirements," *Accident Analysis & Prevention* 124 (2019): 66–84, https://doi.org/10.1016/j.aap.2018.12.022.

[191] W. . w. Wu, S. . y. Chen, and C. . j. Zheng, "Traffic Incident Duration Prediction Based on Support Vector Regression," in *ICCTP 2011: Towards Sustainable Transportation Systems* (2011), 2412–2421.

[192] M. Ghasri, M. Maghrebi, T. H. Rashidi, and S. T. Waller, "Hazard-Based Model for Concrete Pouring Duration Using Construction Site and Supply Chain Parameters," *Automation in Construction* 71 (2016): 283–293, https://doi.org/10.1016/j.autcon.2016.08.012.

[193] L. Lin, Q. Wang, and A. W. Sadek, *Duration Prediction of Urban Freeway Traffic Accidents Based on the M5p Tree and Hazard-Based Duration Model*.

[194] Y. C. Lu, "Detecting Outliers for Improving the Quality of Incident Duration Prediction," (College Park: University of Maryland, 2021), Ph.D. Thesis.

[195] W. Kim and G. L. Chang, "Development of a Hybrid Prediction Model for Freeway Incident Duration: a Case Study in maryland," *International journal of intelligent transportation systems research* 10, no. 1 (2012): 22–33, https://doi.org/10.1007/s13177-011-0039-8.

[196] L. Lin, Q. Wang, and A. W. Sadek, "A Novel Variable Selection Method Based on Frequent Pattern Tree for Real-Time Traffic Accident Risk Prediction," *Transportation Research Part C: Emerging Technologies* 55 (2015): 444–459, https://doi.org/10.1016/j.trc.2015.03.015.

[197] S. Wang, S. Zhang, T. Wu, Y. Duan, and L. Zhou, "Research on a Dynamic Full Bayesian Classifier for Time-Series Data with Insufficient Information," *Applied Intelligence* 52 (2022): 1059–1075, https://doi.org/10.1007/s10489-021-02448-6.

[198] E. I. Vlahogianni and M. G. Karlaftis, "Fuzzy-entropy Neural Network Freeway Incident Duration Modeling with Single and Competing Uncertainties," *Computer-Aided Civil and Infrastructure Engineering* 28, no. 6 (2013): 420–433, https://doi.org/10.1111/mice.12010.

[199] Y. Wang, Y. Liu, J. Zhang, Q. Ye, and Q. Zhu, "Gsnet: Graph-Structured Network for Traffic Scene Understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[200] C. Yang, M. Chen, and Q. Yuan, "The Application of Xgboost and Shap to Examining the Factors in Freight Truck-Related Crashes: An Exploratory Analysis," *Accident Analysis & Prevention* 158 (2021): 106153, https://doi.org/10.1016/j.aap.2021.106153.

[201] A. Grigorev, "Advanced AI Techniques for Comprehensive Traffic Incident Analysis: Enhancing Incident Duration Prediction and Accident Risk Forecasting," (2023), http://hdl.handle.net/10453/177936.