# A review on action recognition for accident detection in smart city transportation systems

Victor A. Adewopo[1*] , Nelly Elsayed[1], Zag ElSayed[1], Murat Ozer[1], Ahmed Abdelgawad[2] and Magdy Bayoumi[3]

*Correspondence:
Adewopva@mail.uc.edu

[1] University of Cincinnati,
Cincinnati, OH, USA
[2] Central Michigan University,
Mount Pleasant, MI, USA
[3] University of Louisiana
at Lafayette, Lafayette, LA, USA

## Abstract

Accident detection and public traffic safety is a crucial aspect of safe and better community. Monitoring traffic flow in smart cities using different surveillance cameras plays a crucial role in recognizing accidents and alerting first responders. In computer vision tasks, utilizing action recognition (AR) has contributed to high-precision video surveillance, medical imaging, and digital signal processing applications. This paper presents an intensive review focusing on action recognition in accident detection and autonomous transportation systems for smart city. This paper focused on AR systems that use diverse sources of traffic video, such as static surveillance cameras on traffic intersections, highway monitoring cameras, drone cameras, and dash-cams. Through this review, we identified the primary techniques, taxonomies, and algorithms used in AR for autonomous transportation and accident detection. We also examined datasets utilized in the AR tasks, identifying the primary sources of datasets and features of the datasets. This paper provides a potential research direction to develop and integrate accident detection systems for autonomous cars and public traffic safety systems by alerting emergency personnel and law enforcement in the event of road traffic accidents to minimize the human error in accident reporting and provide a spontaneous response to victims.

**Keywords:** Accident detection, Action recognition, Computer vision, Smart city, Smart transportation

## Introduction

In the field of computer vision, action recognition is a domain that has gained much attention since the advancement of convolution neural networks (CNNs) as a tool for solving complex computer vision tasks and has gained more traction in the research community over the past few years [1]. Action recognition has been used in several real-life applications such as safety and security AI [2, 3], healthcare AI [2, 4, 5], and media AI [6, 7]. Developing algorithms that can intuitively detect actions in video stream presents an opportunity to advance research frontier using AI for human action recognition. Action recognition cuts across three major activities: feature detection, action representation, and action classification. Detecting actions in sequence of images or video streams presents a unique challenge based on cluttered background, occlusion, and difficulties labeling human actions that are distinct from one person to another [8].

Action recognition with a single action class in a video stream lags in practical applications. However, in a multi-class AR task, action localization in the untrimmed video is tedious as it involves developing architectures that can accurately set boundaries and train end-to-end algorithms to recognize different action classes [9]. Pose estimation algorithms have been widely proposed in action recognition problems to help recognize and understand how each action happens [10]. Pose estimation has been successful in multiple human action recognition tasks such as [11–14]. However, the algorithm is not beneficial for vehicle accident detection due to the specificity of the problem and the differences between humans and vehicles in terms of physical behaviors and constructions.

Transfer learning is a commonly used technique in extracting features of deep neural networks that have been trained on a specific domain with robust dataset to a new domain/area of application with reduced computational resources. Previous research has leveraged transfer learning approach to improve video stream action localization. Iqbal et al. [15] experimented with action localization on pre-selected frames by leveraging transfer learning from the existing model. The overarching goal was to simplify the complex architectures, expensive computation costs, and inefficient inferencing in existing methodologies.

The current research trend in action recognition is focused on deep neural networks with two stream architectures Optical flow and RGB (red, green, and blue) [16]. Transferring features from the pre-trained model on small action classes significantly improves AR models' performance, while other areas of focus have been on temporal localization and segmentation of actions in untrimmed video. Hidden Markov's model has been used to capture long-range dependencies in frame-wise action recognition [17]. In contrast, the spatiotemporal convolution and semi-hidden Markov model were used in capturing multiple action transitions in untrimmed video [18]. Iqbal et al. [15] utilized transfer learning technique with the I3D network on temporarily untrimmed video to localize all action class instances in a video stream. Their experimental research using deep vanilla temporal convolutions network on features extracted from I3D yielded state-of-the-art results with a lightweight model and simple convolution network to extract features from the existing model without multiple layers and gated convolutions [15].

The main focus of this paper is the investigation of seminary articles on accident detection and the review of methods that have been explored by researchers who use computer vision and action recognition techniques for detecting traffic accidents. Based on the previous background information on action recognition, accident detection requires pattern matching and capturing spatial–temporal information along with other road structure artifacts in order to detect traffic accidents. Based on the previous background information on action recognition, accident detection requires pattern matching and capturing spatial–temporal information along with other road structure artifacts in order to detect traffic accidents. The research presented in [19, 20] combined sensors and AR methods in detecting traffic accidents anomalies in traffic flows. Recent deep learning techniques employed transformer models, graph-based and attention mechanisms for accident detection [21, 22]. Traffic accident detection is beneficial for managing urban traffic and providing adequate information to motorists regarding alternate routes while aiding emergency responders in taking quick action. Yu et al. [23] proposed an accident detection method at the road level that integrates internal factors

(road type, road structure, environment) with external factors such as driver behavior, weather, and road congestion. We reviewed an extensive number of seminary articles on accident detection, including but not limited to [21, 24–26]. Road traffic accident is one of the leading causes of non-natural death. Artificial intelligence plays a significant role in detecting accidents and recognizing scene activity in autonomous transportation. Many research advancements focused on developing algorithms for detecting accidents and modeling spatiotemporal information found in road structures. However, previous studies have not extensively addressed different techniques and criteria for establishing new benchmarks datasets for accident detection in smart cities. In an effort to develop a consistent benchmark, our study examined seminary articles on accident detection that have been published in the past ten years. This approach aims to provide a more comprehensive understanding of the performance of each model. This paper provides a comprehensive review of action recognition focusing on accident detection and autonomous transportation in smart city transportation systems. This review includes the state-of-the-art techniques researchers have proposed, accident detection algorithms, the application of AR/accident detection in smart cities, and transfer learning approach from complex architectures. Furthermore, we identified gaps in the existing literature on accident detection and formulated research questions to stimulate further research on public traffic safety using an accident detection model integrated into automated smart city traffic monitoring and safety technologies. The main contribution of this paper is summarized below:

- Provided a comprehensive comparison of different action recognition techniques used in smart city transportation systems and synthesize the state-of the-art research findings within the past ten years on autonomous transportation and accident detection.
- Interpreted and analyze benchmark datasets, algorithms, and metrics used by relevant research work on traffic control and accident detection domain.
- Explored literature gaps in existing methodology that can be addressed by current technological advancement.
- Identified potential future research questions that leverage existing methodology with reduced model complexities and computation resources.

The structure of this paper is organized as follows: "Action recognition applications" section presents background and existing literature review on the domain mentioned above. The literature search, methodology, inclusion, and exclusion criteria are discussed in "Literature search" section. The research findings and detailed analysis are discussed in "Results" section. Finally, "Limitation" and "Conclusion" sections elaborated on the limitations and conclusion of the study.

## Action recognition applications

Action recognition is a revolutionary topic in machine learning and computer vision that has been utilized in intelligent systems such as human-assisted AI (e.g., surgery [27, 28], sports [29, 30], education [6]), smart cities [31], safety and security [32, 33], smart home [34], crisis informatics [35], medical imaging [36, 37], and robotics [38, 39].

Considering the wide application area of AR, in this research, we limit our scope to the application of action recognition addressing accident detection in smart city autonomous transportation.

### Action recognition in smart city

A futuristic direction in computer vision is the application of intelligent systems in autonomously performing human activities that are repetitive in nature and capital intensive.

In a smart city surveillance system, violence can be efficiently spotted to alert appropriate enforcement agencies with automated analysis of video contents in surveillance cameras [40]. For example, SenSquare is a mobile crowd-sensing framework for smart cities that involves users' participation in large data gathering [41]. The SenSquare system was implemented using crowd-sensing heterogeneous data sources for gathering data and developing classification algorithms in order to detect potential hazardous behavior in the environment [41–43]. The community-based monitoring paradigm focuses on tracking users, monitoring emergencies, and responding to them. Law enforcement agencies continuously face an uphill battle in controlling the increase in crime rates and gun violence. Deploying intelligent surveillance cameras can assist in the automatic detection of firearms and alert security agencies in near real time when firearm has been detected. Romero et al. [44] developed an object detection model that can detect firearms and crime scenes in dangerous situations based on Yolo's object detection framework using surveillance cameras. Jamil et al. [45] proposed human action recognition system utilizing spatial–temporal weighted BILSTM-CNN framework for accurate firefighter's activity recognition during hazardous scenarios, integrating 1D-CNN and a contextaware-enhanced BILSTM in three-stream architecture.

Human behavior and specific actions can be analyzed and classified using imaging and AI technologies. Patil et al. [46] research demonstrated the feasibility of using visual-based methods for facial emotion recognition (FER), leveraging both visual and physiological biosignals, which has potential applications in areas such as lie detectors and human–machine interfaces on portable hardware. Similarly, the application of AR models in understanding human behavior offers possibilities for smart city safety, especially in tracking drivers' behavior. The National Highway Traffic Safety Administration (NHTSA) report highlighted an increase in the number of fatalities caused by distracted drivers between 2019 and 2020, which is higher than the number of fatalities caused by total accidents in 2017. The number of fatalities caused by distracted drivers rose to more than 8.5% of total fatalities during 2017 [47]. Celaya et al. [48] proposed a deep convolution neural network for detecting texting and driving behavior using a car-mounted wide-angle camera with a pre-trained Inception v3 model. Emerging technologies like the AR model can be integrated with CCTV cameras to reduce fire accidents in smart cities. As described in [49] on fire detection method in smart city environments using the Yolo4 algorithm, a robust model based on augmented data (different weather environments) and a reduced network structure demonstrated excellent performance and is highly effective for detecting fire disasters. In this paper, we focus on accident detection using data obtained from different types of surveillance cameras in smart city transportation safety and monitoring systems.

### Action recognition in autonomous transportation and accident detection

Robotics and auto navigation systems have also benefited from using AR for autopilot, specifically in obstacle detection, accident prevention, and lane departure assistance [40]. Accident detection in autonomous transportation systems is essential for tracking vehicles and identifying anomalies in traffic patterns. Cai et al. [50] discussed the detection of abnormal traffic flow using clustering techniques on main flow direction vectors and a k-means clustering algorithm to identify outliers that deviates from normal trajectory pattern or motion flow in highways. Previous research explored intelligent visual descriptions of scenes with connected image points using spatiotemporal dynamics in the hidden Markov model [51]. Recent research approached this challenge using machine learning algorithms and deep learning techniques [52–54]. Robles-Serrano et al. [54] combined convolution layers and long short-term memory LSTM architectures in capturing spatiotemporal features from a sequence of images in video streams that have proven to achieve better performance [55, 56]. Due to the capability of convolutional layers to extract features from each image in video stream and the capability of LSTM to learn temporal information between images in video sequence [57–59]. Obstacle detection is an integral part of intelligent transportation systems, the research of Liang et al. [60] presented a refined multi-object detection algorithm by combining DarkNet-53 with the enhanced features of DenseNet. The proposed system evaluated on benchmark datasets (KITTI and Pascal VOC) showed notable improvements in model adaptability, especially in addressing the challenges of occlusion, underscoring its value in intelligent transportation obstacle detection [60]. Accident detection task includes the detection of spatiotemporal dependencies in multiple frames from video surveillance. Hence, correctly classifying video input as an accident is more challenging in developing accident detection models and requires highly voluminous data. Carreira et al. [61] introduced a new two-stream inflated 3D ConvNet (I3D) based on a 2D ConvNet inflation. The authors seek to unravel the correlation between increase in performance and complex networks by inflating the pooling kernel image classification architectures to an inflated two-stream ConvNets (I3D). The results of their proposed framework suggested that there is always a boost in performance by pre-training a model. However, the extent of the boost varies significantly with the type of architecture.

### Accident detection methods

Most researchers proposed their own datasets and evaluation criteria in action recognition tasks, making it challenging to identify the most appropriate datasets and compare results. Performance metrics also vary across multiple research works; developing a standardized evaluation technique will lead to more robust research in the application of AR for accident detection tasks. Current methods allow some data samples to be repeated/duplicated in train/test data which directly causes bias in actual performance when evaluating a new research work [62]. Stisen et al. [63] examined the effects of heterogeneous devices on human activity recognition (variations in training and test device hardware) on model performance using hand-crafted features and employed popular classifiers such as nearest-neighbor, support vector machines, and random forest. They noticed sampling instabilities occurred across various devices. Dataset source also plays

Adewopo *et al. Journal of Electrical Systems and Inf Technol*      (2023) 10:57

Page 6 of 27

a crucial role in designing accident detection models because Videos captured by dash-cams hold different video trajectories and street vision than highway or traffic lights surveillance cameras. The dashcams capture the traffic video from a horizontal view. In such captured videos, both the camera and surrounding objects are moving. These increase the problem's complexity, especially when classifying objects approaching the dashcam and the objects the car is moving toward with a stationed dashcam. Traffic light and highway cameras record the scene in a vertical view, with the camera in a fixed position, while moving objects are recorded at a fixed point. Therefore, addressing each type of video content plays a significant role in calculating the trajectories, the acceleration of objects, and the moving directions. Sayed et al. [64] highlighted challenges in AI-based traffic flow prediction, such as the scarcity of high-quality training data and computationally effective methods. These issues, coupled with underutilized spatiotemporal correlations in deep learning methods restrict advancements in traffic flow predictions [64].

### Machine learning and statistical models

Most machine-learning algorithms focus on vehicle trajectory, motion, acceleration, and car position in detecting car accidents. Singh et al. [64] combined two algorithms using object detection and anomaly algorithm detection to identify accidents. Singh et al. [64] proposed a framework that extracts deep representation using autoencoders and an unsupervised model (SVM) to detect the possibility of an accident. The vehicle's trajectories at the intersection points were used to increase the proposed architecture's precision and reliability. Joshua et al. [19] proposed mathematical relationships obtained through multiple linear and Poisson regression analyses to identify factors contributing to significant truck accidents on the highway using accident dataset from Virginia highway traffic in combination with other geometric variables to model the percentage of trucks involved in road accident. Arvin et al. [20] leveraged the availability of extensive data from interconnected devices in making correlations between erratic driving volatility and historical crash datasets from intersections in Michigan. Statistical variables such as fixed parameters, random parameters, geographically weighted Poisson regressions, and longitudinal and lateral acceleration were used to identify road accident crash hotspots.

### Deep machine learning models

Most deep learning algorithms focus on vehicle trajectory, motion/acceleration, and car position for detecting car accidents. Chan et al. [65] proposed a dynamic-spatial attention (DSA) recurrent neural network (RNN) for anticipating accidents in dashcam videos based on vehicle trajectory and motion. The developed algorithm contains an object detector to dynamically gather subtle cues and temporal dependencies of all cues to predict accidents two seconds before they occur with a recall of 80% and low precision of 56.14%. The model's generalizability in detecting accidents in varying weather conditions was not measured based on limited videos with rain, snow, and day/night, among other weather conditions. Robles-Serrano et al. [54] explored deep neural networks for accident detection using a three-stage approach by firstly segmenting visual characteristics of objects in the dataset, building on Inception V4 model architecture to extract temporal components of the dataset used in detecting accidents followed by temporal

video segmentation. A structural similarity index was applied to the dataset at preprocessing time to accurately select image frames within the data representing an accident or no accident as part of the temporal video segmentation to eliminate frames that do not contain event occurrence or repetition of a selected event. During preprocessing, pixel-to-pixel comparisons were made to select a certain number of consecutive frames that contained features to train the model based on a specified threshold. Finally, the framework was designed to detect accidents automatically using Convolution LSTM (ConvLSTM) layers to capture spatial and temporal dependencies in input data [66, 67]. This type of neural network has proven to perform better than LSTM and CNN architectures when dealing with datasets that have both spatial and temporal structures. One of the potential limitations of model bias is based on vehicle types and other environmental conditions, such as vehicle variety and the absence of pedestrians and cyclists.

### *Social network and geosocial media data*

The enormous amount of information being constantly shared daily across various social media platforms contains artifacts that can be analyzed to generate meaningful insights for traffic events [68]. However, the ability to manually monitor and analyze exploding information seems impossible based on the high volume and unstructured formats of information being presented [69]. Monitoring traffic-related information on social media has been proven to be beneficial in detecting traffic events. Xu et al. [70] provided a synthesis of research work that explored the usage of geosocial media data for detecting traffic events. Events such as road accidents, road closures, and traffic conditions are typically shared among a network of people through social media platforms. Such events can be tracked with the aid of GPS in getting first responders to the event location and often contain information that triggered such events. Xu et al. [71] utilized Twitter data for mining and filtering noisy data by association rules among words related to traffic events. The proposed framework achieved 81% accuracy in classifying data into non-traffic events, traffic accidents, roadwork, and severe weather conditions. Similarly, Salas et al. [72] developed a framework leveraging social media data to crawl, process, and filter social media data for implying traffic incidents and real-time detection of traffic events with text classification algorithm [73].

### Literature search

The literature search process consists of four steps, including (i) selecting eligibility criteria (Inclusion and Exclusion criteria), (ii) formulating research objectives, (iii) identifying search strategy, and (iv) data extraction [74, 75]. This study employed systematic review methodology to address the research questions posited through a systematic and replicable process [76]. Specifically, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement was used as a model for this review [77, 78]. The papers selected were analyzed and synthesized based on established eligibility criteria to address the research questions postulated in the following subsection.

### Research questions and objectives

Developing an AR model for specific tasks will enhance the use of AI systems in automating human actions and autonomously detecting actions in live feeds. Once the

Adewopo *et al. Journal of Electrical Systems and Inf Technol*        (2023) 10:57

Page 8 of 27

inclusion selection process has been carried out, based on pre-established criteria, the main results of selected works are codified, extracted, and synthesized to guide this research. The following research questions were addressed in this comprehensive review:

- *RQ1*: What are the main action recognition techniques/applications in accident detection and autonomous transportation?
- *RQ2*: What are the main taxonomies and algorithms used in action recognition for accident detection and autonomous transportation?
- *RQ3*: What are the main datasets, features, and metrics used in action

Recognition for accident detection tasks?

### Selecting eligibility criteria

This review includes research articles related to action recognition. The topics include (autonomous transportation, traffic control, and accident detection using computer vision, published in peer-reviewed journals. Based on continuously evolving advancements in the technical field, we limited our scope to research articles published in the last ten years before this review (from 2012 to 2022). Only research articles published in English language were used. The inclusion and exclusion criteria were detailed in "Inclusion criteria" and "Exclusion Criteria" sections, respectively. This systematic review is based primarily on computer vision tasks using the AR model in autonomous transportation and smart city accident detection.

### Inclusion criteria

In our inclusion criteria, the publications needed to meet the following characteristics in order to be included:

1. Articles should be in action recognition and computer vision research domain.
2. Studies that include validation of the proposed techniques.
3. Published within the last ten years (i.e., between 2012 and 2022).
4. Peer-reviewed full research papers.
5. Contain analysis of spatial/temporal information in the paper.

### Exclusion criteria

In our exclusion CRITERIA, the following exclusions were implemented:

1. Does not contain video/motion analysis.
2. Published before 2012.
3. Not peer-reviewed research paper.
4. Paper does not provide clear findings and analysis of results.
5. Written in other languages excluding english.
6. Duplicated studies.

**Table 1** Article data source

| ID | Database | Link | Number of articles |
|----|----------|------|--------------------|
| D1 | IEEE Xplore | http://ieeexplore.ieee.org/ | 299 |
| D2 | ACM | http://dl.acm.org/ | 181 |
| D3 | Web of Science | http://www.webofscience.com/ | 445 |
| D4 | Springer Link | http://link.springer.com/ | 572 |
| D5 | Science Direct | http://sciencedirect.com/ | 533 |



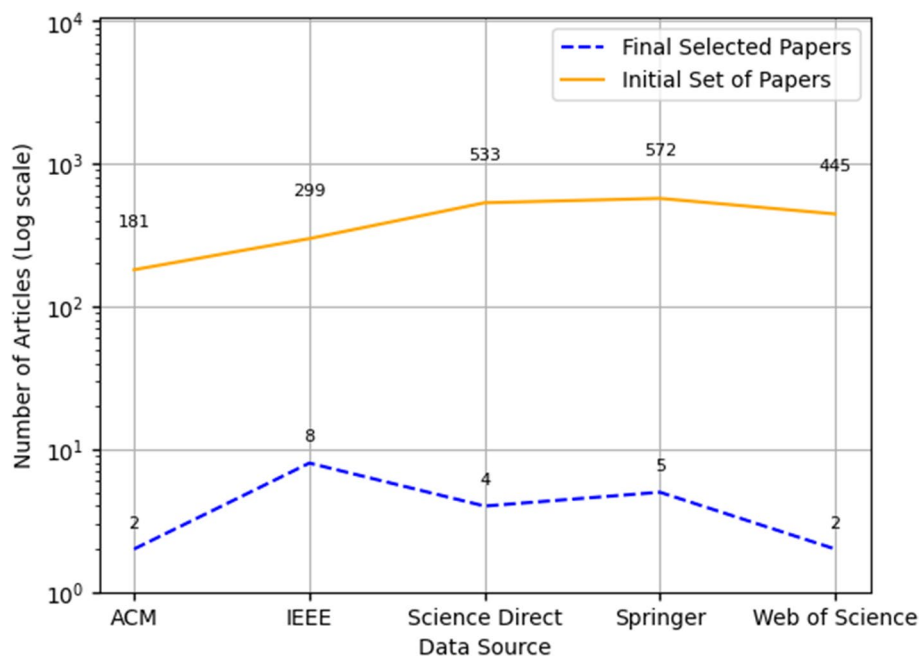**Fig. 1** Proportion of selected studies

### Information sources and search strategy

The papers included in our review were identified by searching electronic databases that were published in English. The databases in Table 1 were used as the primary source of articles for this review. These databases provide impactful articles from full-text journals and conferences relevant to Action Recognition tasks in smart city automation, autonomous transportation, and accident detection. The first phase includes searching the databases in Table 1 with advanced search and filtering techniques to limit search results to only relevant studies. The number of articles retrieved from each database and the final number of papers selected is showcased in Fig. 1. Combining the following keywords with conjunctions "AND" and disjunctions "OR" resulted in a total of 2,030 papers in an automated search, as shown in Table 1. The most common terms used for our search were:

1. Action Recognition.
2. Transportation.

Adewopo *et al. Journal of Electrical Systems and Inf Technol*  (2023) 10:57

Page 10 of 27

3. Traffic control.
4. Accident Detection.

The results of our search and the corresponding query that has been used are as follows:

- *IEEE Xplore*: We received *299 papers* from IEEE using the search string:[(("All Metadata":Action Recognition) AND ("All Metadata": Transportation) OR ("All Metadata":Action Recognition) AND ("All Meta-data":Traffic) OR ("All Metadata":Action Recognition) AND ("All Meta-data":Accident Detection))] between 2013 and 2022
- *ACM*: We received *181 papers* from ACM using the search string: [All-Field:( "Action Recognition") AND AllField:("Transportation") OR All-Field:( "Action Recognition") AND AllField:("Traffic") OR AllField:( "Action Recognition") AND AllField:("Accident Detection")]
- *Web of Science*: We received 445 papers from Web of Science using the search string:[((ALL = (Action Recognition) AND ALL = (Transportation OR Traffic OR Accident Detection))) AND (PY = = ("2022" OR"2021" OR"2020" OR"2019" OR"2018" OR"2017" OR"2016" OR"2015" OR"2014" OR"2013"))]
- *Springer Link*: We received *572 papers* from Springer Link using the search string: [("Action Recognition") AND (("Transportation") OR ("Traffic") OR ("Accident Detection"))] between 2013 and 2022
- *Science Direct*: We received *533 papers* from Science Direct using the search string: [("Action Recognition" AND'Transportation') OR ("Action Recognition" AND'Traffic') OR ("Action Recognition" AND'Accident Detection')] between 2013 and 2022

**Study selection**
The articles were evaluated and selected according to the mentioned criteria in "Literature search" section. After a preliminary database search using the approved search strategy conducted by student researchers and eliminating duplicates, a total of 1829 articles were screened by two faculty researchers and one student researcher independently who are domain experts. The abstracts, titles, and keywords from selected articles were reviewed for relevance based on the inclusion and exclusion criteria. Articles that did not meet the eligibility criteria or were not relevant to address the research question were removed. The independent researchers rated each article based on the inclusion criteria and eligibility criteria. The painstaking protocol observed in the selection process ensures that all articles included are relevant to this study. A total of 1650 papers were excluded because they do not contain video analysis or employ AR techniques in detecting accidents. Thirty-three papers were excluded because they lacked validation techniques for the proposed methodology, 108 papers identified as review papers were excluded, and 17 papers contained only abstracts. Duplicate studies that cover the same issues are excluded from the study.

Figure 1 showcases the proportion of initial articles and final articles selected from each of the five online data sources listed in Table 1. Finally, only 21 papers were selected for analysis, as shown in Fig. 2.

## Results

Following PRISMA guidelines, 2030 publications were identified through five databases included, and the results of the 21 papers selected for review are presented in this section. Figure 3 showcases the publication year for selected papers. It is noteworthy that the majority of the selected papers were published between 2019 and 2021. Taking advantage of the advancement in technology and smart city automation, recent research employs deep learning techniques to model traffic-related activities in the smart city utilizing computers equipped with high-performance GPU processors.

### RQ1: main action recognition techniques/applications in accident detection and autonomous transportation

The first research question of our study is to examine the main AR techniques and applications within smart cities and autonomous transportation, as shown in "Research questions and objectives" section. Many researchers have proposed other methods to model traffic management and traffic prediction, including vector auto-regression,
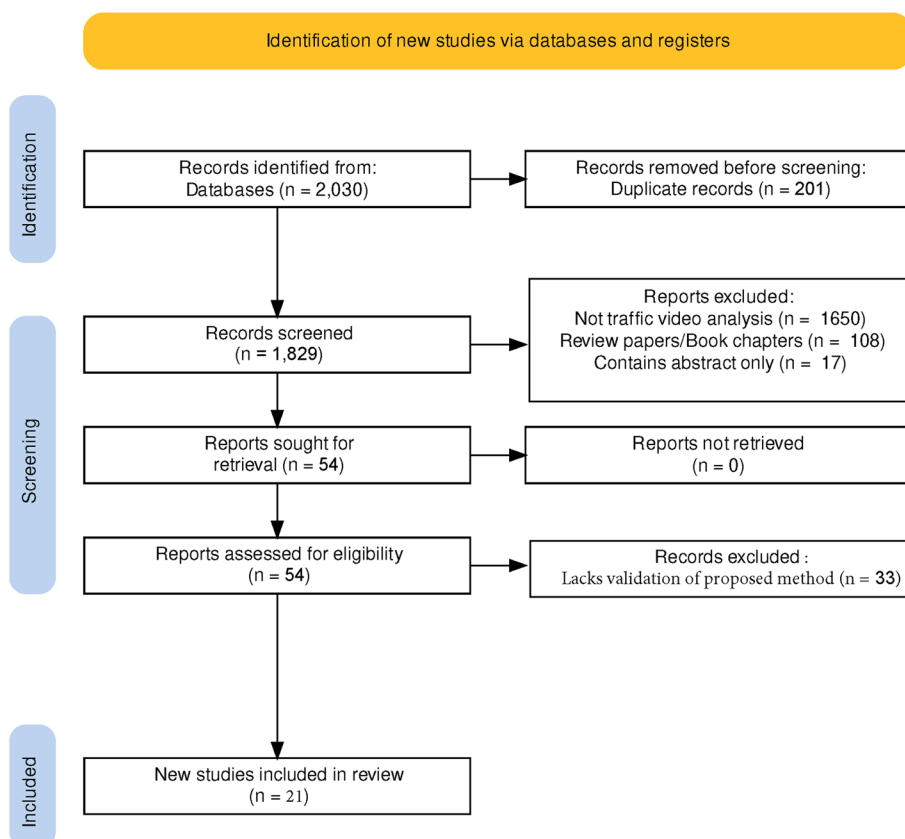


**Fig. 2** Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow chart of the systematic review
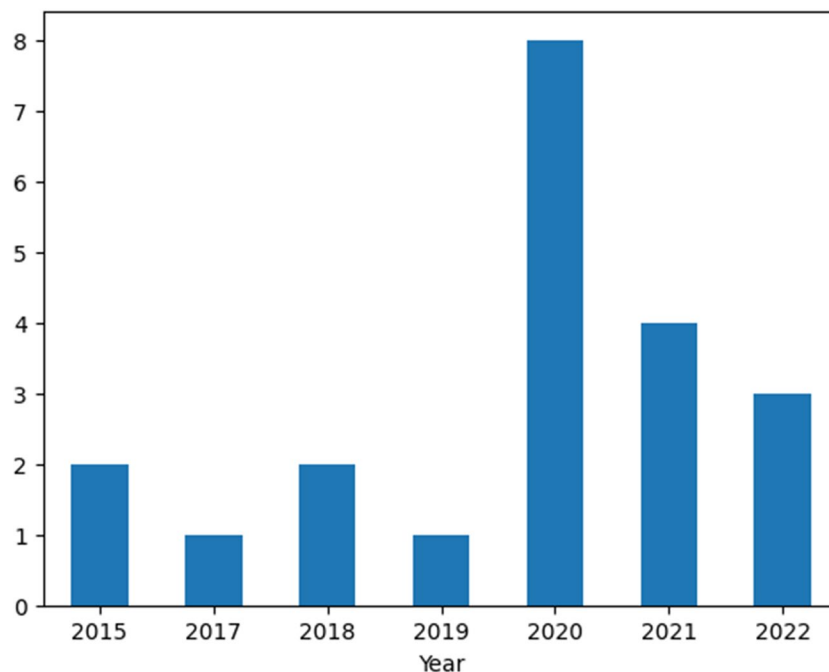
**Fig. 3** The number of papers published per year surveyed

support vector regression, auto-regressive integrated moving average (ARIMA), Kalman filter, RNN, and transformer models [79, 80]. In time series data, such as traffic control data, the approaches have not been able to capture both spatial and temporal information concisely. Recent efforts have improved the accuracy performance of GNN and GaAN [81, 82]. Ijjina et al. [83]. proposed a supervised deep learning framework to detect and identify road-side vehicular accidents by extracting feature points such as car trajectory, weather conditions (daylight variations), and velocity in detecting traffic anomalies in real time. Fernandez-Llorc et al. [84] study utilized a disjoint two-stream convolutional network and spatiotemporal multiplier network with the visual cues extracted from the camera to detect lane change or vehicle maneuvers. You et al. [85] discovered that time segmentation methods such as SS-TCN and MS-TCN were more successful at higher IoU thresholds. Their experiment also suggests that the region convolutional 3D Network (R-C3D) algorithm has a comparable result when compared to segmentation-based approaches. However, newer methods like R(2+1)D and SlowFast network have improved accuracy. Most techniques fail to capture traffic anomalies accurately on DoTA datasets, suggesting that traffic anomaly classification is a challenging task. Yao et al. [86] suggest that distant objects and occluded objects are difficult to classify because of their low visibility. Collisions with moving vehicles present a similar problem because the vehicle ahead is substantially obscured by the vehicle it impacts. There may be instances when a vehicle hits obstacles that are not detected, such as bumpers or traffic cones. Most often, anomalous vehicles are responsible for occluding the obstacles. It is hard to detect horizontal vehicle collisions due to their vertical trajectory making traffic anomalies subtle and thus hard to detect. The joint sparse modeling (JSM) method extracts motion trajectory to evaluate traffic scenes but ignores traffic events that occurred in an unusual manner [86, 87]. Srinivasan et al. [24] developed a scalable

algorithm (DETR) for high-speed object detection, with less complex architecture and higher accuracy compared to other object detection algorithms using correlation techniques between objects in video data. Tables 2 and 3 address the research question on main action recognition techniques and applications in autonomous transportation.

**Table 2** Studies were used to address the research question on main action recognition techniques and applications in autonomous transportation

| Article | Actions considered | Learning method | Category | RQ1 | RQ2 | RQ3 |
|---|---|---|---|---|---|---|
| Yao et al. [86] | Vehicle collision Car trajectory | Unsupervised learning | Traffic anomaly detection | ✓ | ✓ | ✓ |
| Yu et al. [23] | Vehicle collision Road condition Traffic speed | Deep learning | Accident prediction | ✓ | ✓ | ✓ |
| Wang et al. [80] | Traffic flow Traffic speed | Graph neural network | Traffic flow pattern | - | ✓ | ✓ |
| Bao et al. [88] | Car trajectory Inconsistent motion | Graph convolution network bayesian neural network | Traffic anomaly detection | ✓ | ✓ | ✓ |
| Reddy et al. [26] | Traffic signals Car distance limit Traffic speed | Deep learning | Accident detection | ✓ | ✓ | – |
| Fernandez et al. [84] | Driver intention detection Lane maneuver | Deep learning | Traffic anomaly detection | ✓ | ✓ | ✓ |
| Ali et al. [22] | Traffic flow | Graph convolution network | Traffic flow pattern | ✓ | ✓ | – |
| Wang et al. [21] | Car detection Accident detection | Graph convolution network | Accident detection | ✓ | ✓ | – |
| Alkandari et al. [89] | Traffic flow Traffic pattern | Fuzzy logic technique | Accident detection | – | ✓ | – |
| Riaz et al. [90] | Traffic speed Car distance Accident detection | Deep learning | Traffic anomaly detection | ✓ | ✓ | – |
| Tang et al. [91] | Car trajectory Car position | Probabilistic | Traffic flow pattern | ✓ | – | ✓ |
| Huang et al. [25] | Traffic volume Traffic speed | Deep learning | Accident detection | ✓ | – | – |
| Bortnikov et al. [92] | Car collision Accident detection | Deep learning | Accident detection | ✓ | – | ✓ |
| Gupta et al. [93] | Car collision Accident detection | Deep learning | Accident detection | ✓ | ✓ | ✓ |
| Yang et al. [94] | Accident detection | Deep learning | Accident detection | ✓ | ✓ | ✓ |
| Ijjina et al. [83] | Traffic speed Traffic trajectory Car distance | Deep learning | Traffic anomaly detection Accident detection | ✓ | ✓ | ✓ |
| You et al. [85] | Car trajectory Vehicle detection | Deep learning | Accident detection | ✓ | ✓ | ✓ |
| Srinivasan et al. [24] | Vehicle detection | Deep learning | Accident detection accident classification | ✓ | ✓ | ✓ |
| Hui et al. [95] | Vehicle trajectory Traffic speed | Deep learning | Accident detection | ✓ | – | – |
| Min et al. [87] | Traffic pattern Traffic trajectory | Probabilistic | Traffic flow pattern | ✓ | ✓ | ✓ |
| Vatti et al. [96] | Car collision Lane maneuver | Statistical model | Traffic flow pattern | – | ✓ | – |

The notation "–" means the research paper did not address our research question

**Table 3** Keynote of studies that were used to address the research question on main Action Recognition techniques and applications in autonomous transportation

| Article | Key notes |
| --- | --- |
| Yao et al. [86] | The researchers developed a benchmark dataset to assess the quality of traffic accident detection and anomaly detection for nine action classes |
| Yu et al. [23] | Traffic accidents can be caused by many factors, including driver behavior, weather conditions, traffic flow, and road structures. The authors investigated spatial–temporal relationships on heterogeneous data to develop a road-level accident prediction system |
| Wang et al. [80] | The goal of this project is to develop a framework for analyzing stationary time series traffic data. In addition, it is able to predict traffic information with a 14.1% improvement in MAPE compared to other baselines |
| Bao et al. [88] | Using GCN and BNN, the developed model can handle the challenges of relational feature learning and uncertainty anticipation from video data to anticipate an accident in 3.53 s with an average precision of 72.22% |
| Reddy et al. [26] | The research investigates how to extract road features relevant to the trajectory of an autonomous vehicle from real-world road conditions using Deep Q-Learning in a real-world environment setting |
| Fernandez et al. [84] | This study utilized a visual cue derived from a camera to detect lane change/vehicle maneuvers by utilizing a disjoint two-stream convolutional network and a spatiotemporal multiplier network |
| Ali et al. [22] | The authors propose a hybrid model combining GCN and DHSTNet that is effective in forecasting short-term traffic patterns in urban areas in order to improve traffic management |
| Wang et al. [21] | The researchers developed a new dataset and proposed a method for safety prediction |
| Alkandari et al. [89] | The main aim of this study is to develop a methodology for controlling the length of time that a vehicle stays in traffic based on the flow of traffic and congestion |
| Riaz et al. [90] | This study implemented the FWPredNet framework for accident and anomaly prediction, which outperformed the previous state-of-the-art framework |
| Tang et al. [91] | Researchers developed a framework for conceptually describing components of surveillance video, separating them into smaller components, and detecting activities from some short clips of two seconds |
| Huang et al. [25] | The authors present a comparative analysis of different statistical and deep learning models for solving traffic safety problems through the detection of collisions and estimating crash risk in urban Interstate highways |
| Bortnikov et al. [92] | Through the use of video games under different weather conditions and scene conditions, the study generated traffic data that was then processed and trained with a 3D CNN. This model yielded comparable results to real-life traffic videos from YouTube |
| Gupta et al. [93] | Using time-dependent frames in a video, the developed model was able to evaluate the effectiveness of the model on trimmed unlabeled video |
| Yang et al. [94] | In this paper, researchers propose developing a feature-fused SSD in order to improve detection accuracy of vehicles from the ImageNet video database |
| Ijjina et al. [83] | The proposed supervised deep learning framework detects and identifies road-side vehicular accidents by extracting feature points based on local features such as trajectory intersection and velocity, and by detecting anomalies in real-time accident conditions such as daylight variations |
| You et al. [85] | In this study, the authors discovered that time segmentation methods such as SS-TCN and MS-TCN were more successful on the dataset at higher IoU thresholds. In addition, the R-C3D algorithm has a comparable result when compared to segmentation-based approaches |
| Srinivasan et al. [24] | The authors developed a scalable algorithm for high-speed object detection (DETR), with a less complex architecture and a higher level of accuracy compared to other object detection algorithms that are based on correlations between all objects in the video data |
| Hui et al. [95] | The authors proposed using a Gaussian Mixture Model to extract foreground and background information from video streams in order to create a vision-based accident detection model |
| Min et al. [87] | The authors applied the SIFT flow method to improve dense trajectories and generate visual words that can be utilized in detecting traffic flow. The data from the experiments demonstrate that the SIFT method is effective |
| Vatti et al. [96] | The authors developed an electronic notification system that can alert relatives when a vehicle accident is detected based on the vehicle's trajectory, position, and acceleration |

The notation "−" indicates that the corresponding research paper did not address our research question.

**RQ2: Algorithms and taxonomies in accident detection and autonomous transportation**

In order to answer our second research question, we have identified the most critical taxonomies and algorithms used in the AR systems for autonomous transportation and accident detection, respectively. Table 4 shows the models, architecture, features used by other researchers, and the evaluation metrics for evaluating the performance of proposed models. It is noteworthy that most novel research work with novel algorithms employs different metrics to evaluate their algorithm's performance. Yao et al. [86] proposed a new metric for computing traffic anomaly scores using the spatial–temporal area under the curve (STAUC) with a future object localization (FOL) method for unsupervised video anomaly detection (VAD). More than 60% of the reviewed paper evaluated their algorithms using mean absolute percentage error (MAPE), mean absolute error (MAE), mean average precision (MAP), intersection over union (IOU) [97], and detection rate (DR). Reddy et al. [26] developed a spatiotemporal graph neural network for managing and predicting traffic accidents, while RNN, LSTM, and other architectures could not fully capture both spatial and temporal information relevant to accident detection. Their study combined GNN, RNN, and a transformer layer to model complex topological and temporal relationships in traffic video data, including adjacent traffic flows. Yu et al. [23] proposed a new graph-based spatiotemporal model to predict future traffic accidents. The integration of spatial, temporal, and external features in predicting accidents achieved a performance improvement of around 5% over the spatial autoencoder (SAE). Ali et al. [22] developed a Graph Convolutional Network coupled with a dynamic deep hybrid spatiotemporal neural network (DHSTNet) called GCN-DHST-Net, which is an enhanced GCN model for learning spatial dependencies of dynamic traffic flow. The LSTM was used to capture dynamic temporal correlations with other external features. In terms of RMSE and MAPE, the proposed model is 27.2% and 11.2% better than the current state of the art (AAtt-DHSTNet). Wang et al. [80] study focused on accident prediction that considers spatiotemporal dependence and other external factors in anticipating accident occurrence. Reddy et al. [26] proposed a hybrid method for detecting stationary objects, moving vehicles, traffic lights, and road signs using deep Q-learning with YOLOv3. Bortnikov et al. [92] study developed an HRNN for detecting accidents from CCTV surveillance by exploring temporal and spatial features of video footage. Yang et al. [94] proposed tracking-based object detection (TDO) technique and feature-fused SSD. TDO significantly improved detection results over state-of-the-art and established vehicle datasets for highway scene analysis. Huang et al. [52] developed a supervised learning algorithm to detect crash patterns from historical traffic data. They examined different prediction methods to estimate crash risk/accident occurrence. You et al. [85] also created a cause-and-effect-based traffic accident benchmark dataset with temporal intervals in each traffic accident event. The dataset provides atomic cues for reasoning in a complex environment and planning future actions, including mitigating ambiguity in traffic accidents. The framework developed by Tang et al. [91] can classify traffic data into different categories, such as detecting vehicle turning directions, bicycle lanes, and pedestrians within two seconds of traffic footage. In order to correctly predict

**Table 4** Identifying the main taxonomies and algorithms used in AR for autonomous transportation based on relevant studies to our second research question

| Authors | Model | Architecture | Model Features | ACC/AUC/DR/IOU | RMSE/MAP/MAPE | Precision | F1score | Recall | Model Comparison | ACC/AUC/DR/IOU | MAPE/MAE | Citation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Yao et al. [86] | Future object localization | Two stream RNN Ego motion RNN | Object Localization Object detection | 73 | – | – | – | – | ConvAE ConvLSTMAE AnoPred | 64.3 67.5 64.8 | – | 16 |
| Yu et al. [23] | Deep spatiotemporal graph convolutional network (DSTGCN) | Graph convolution network | Weather, road network | – | 0.34 | 82 | 85 | 89 | SVM SdAE TARPML | 0.79 0.81 0.83 | – | 25 |
| Wang et al. [80] | Spatial temporal graph neural network | Spatial-GNN+GRU+Transformer | Traffic Speed Spatiotemporal Dependencies | – | 3.99 | – | – | – | FC-LSTM TGNN | – | 3.44 2.62 | 82 |
| Bao et al. [88] | Spatiotemporal GCN (graph convolution network) | Graph convolution+RNN | Accident relevant cues | – | 53.7 | – | – | – | adaLEA DSA | – – | 52.3 48.1 | 15 |
| Reddy et al. [26] | Deep Q-Learning | Deep Q-Learning YOLOv3 | Car speed Distance and position | 90 | – | 87 | – | – | Yolo 3 | 85 | – | – |
| Fernandez et al. [84] | Two-stream network | Two-stream convolutional networks Spatiotemporal Multiplier networks | Lane change | 90.3 | – | – | – | – | Disjoint two-stream convolution | 89.6 | – | 7 |
| Ali et al. [22] | Dynamic deep spatiotemporal neural network (DHSTNet) | Graph convolution network+LSTM | Weather condition Traffic flow | – | 11.08 | – | – | – | DHSTNet Aatt–DHSTNet | – | 12.80 13.72 | 4 |
| Wang et al. [21] | Spatial–temporal Mixed attention graph-based convolution model (STMAG) | GRU+mixed attention mechanism | Object detection Lane marking | – | 3.23 | – | – | – | XGBOOST SVR LSTM | – | 3.71 3.99 3.43 | 7 |
| Huang et al. [25] | CNN-traffic-incident management (TIM) | CNN | Car speed Road occupancy | 80 | – | – | 78 | – | RF | 76 | – | 31 |

**Table 4** (continued)

| Authors | Model | Architecture | Model Features | ACC/AUC/DR/IOU | RMSE/MAP/MAPE | Precision | F1score | Recall | Model Comparison | ACC/AUC/DR/IOU | MAPE/MAE | Citation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bortnikov et al. [92] | 3D convolutional neural network (CNN) | CNN | Optical flow Vehicle trajectory | – | – | – | 71 | – | – | – | – | 13 |
| Gupta et al. [93] | Time-distributed RNN | LSTM | Temporal features Hierarchical features | 94 | – | 95 | 84 | 75 | – | – | – | 1 |
| Yang et al. [94] | Feature-fused SSD detector | Single-shot multibox detector (SSD) | Detection box | – | 70.5 | – | – | – | SSD TPN | – | – | 2 |
| Ijjina et al. [83] | Mask R-CNN | Deep CNN | Car speed Vehicle trajectory | DR – 71 | – | – | – | – | Deep spatiotemporal network | DR–77 | – | 24 |
| You et al. [85] | Single-stream temporal action proposals (SST) | Temporal segment networks (TSNs) | – | IOU – 42.07 | – | – | – | – | R–C3D MS–TCN | – | – | 10 |
| Srinivasan et al. [24] | Detection transformers and random forest classifier (DETR) | DETR+CNN | Object detection | 78.7 | – | 77 | 77 | 78 | ARRS CVABTS | DR–50 DR–71 | – | 1 |
| Hui et al. [95] | Gaussian mixture model (GMM) | – | Vehicle detection Object tracking | – | – | – | – | – | – | – | – | 39 |
| Min et al. [87] | Sparse topic model | Scale-invariant feature transform (SIFT) flow | Motion Pattern | AUC – 91.2 | – | – | – | – | GPR JSM BiLSTM | 85.5 80.2 88.1 | – | 5 |
| Vatti et al. [96] | Accident detection and communication system | – | Motion pattern | – | – | – | – | – | – | – | – | 16 |

The notation "–" means the metric is not applicable

accidents and classify external factors leading to accident occurrence, Wang et al. [80] take into account spatiotemporal dependence in their proposed methodology.

### RQ3: Main dataset, features, and metrics for action recognition for accident detection

Our third research question focused on exploring the dataset used for accident detection. Table 5 showcases the dataset features, type of sensors/video data, and the link to publicly available datasets for accident detection in a smart city. Yao et al. [86] developed a benchmark dataset to access the quality of traffic accident detection and anomaly detection for nine action classes. Based on the limited annotated real life accident datasets, Bortnikov [92] utilized simulated game video data with varied weather and scene conditions. The method yielded comparable results to real-life traffic videos on YouTube, as shown in Table 5. The majority of dataset used in accident detection and autonomous vehicles are collected from dashcams, traffic surveillance cameras, drones such as HighD, InD, or Interaction datasets [98, 99] and cameras installed on buildings. For example, NGSIM HW101 and NGSIM I-80 datasets [100, 101] contain 45 min of images recorded from a building for eight synchronized cameras at 10 Hz. Fernandez-Llorca et al. [84] suggest that this dataset (NGSIM HW101) is not fully applicable for onboard detection applications even though it is beneficial for understanding and assessing the motion and behavior of vehicles and drivers under different traffic conditions. PKU dataset includes more than 5700 environmental trajectories collected using multiple horizontal 2-D LiDAR covering 360°, including vehicles trajectory data over 64 km and 19 h of footage [102]. The Prevention dataset includes data from three radars, two cameras, and one light detection and ranging (LiDAR), covering a range of 80 m around an ego-vehicle to support the development of intelligent systems for vehicle detection and tracking [103]. Similarly, the apolloscape dataset was developed to support automatic driving and navigation in smart cities. The dataset contains about 100 K image frames and 1000 km trajectories collected using four cameras and two laser scanners with 3D perception LiDAR [104]. Ijjina et al. [83] compiled surveillance videos at 30 frames per second (FPS) trimmed down to 20 s video chunks collected from CCTV videos recorded at road intersections from different parts of the world with diversified ambient conditions such as harsh sunlight, daylight hours, snow and night hours.

## Discussion

### Key summary and recommendations RQ1

From the analysis of the selected papers. The main action recognition techniques in accident detection and autonomous transportation includes supervised deep learning framework for detecting and identifying road-side vehicular accidents by extracting feature points, as proposed by Ijjina et al. [83]. Disjoint two-stream convolutional network and spatiotemporal multiplier network for detecting lane change or vehicle maneuvers, as studied by Fernandez-Llorc et al. [84]. The research of You et al. [85] demonstrates that time segmentation methods such as Single-Stream Temporal Convolutional Network (SS-TCN) and Multi-Stream Temporal Convolutional Network (MS-TCN) performed better with at a higher intersection over union (IoU) threshold indicating the effectiveness of the method in capturing fine-grained temporal and complex patterns in action recognition tasks. Region convolutional 3D network (R-C3D) algorithm shows a

**Table 5** Overview of datasets used in AR for autonomous transportation, features of the datasets, and download links to the datasets

| Authors | Model | Dataset | Fps | Dataset feature | Accessibility | Data collection approach |
|---|---|---|---|---|---|---|
| Yao et al. [86] | Future object localization | DoTA (detection of traffic anomaly) | 10 | 4,77 videos | Yes—link | Dashcam |
| Yu et al. [23] | Deep spatiotemporal graph convolutional network (DSTGCN) | Traffic Accident Data, Taxi GPS Data, Meteorological data | – | Varied weather (Cloudy, Snow, etc.) Road network and PoI (point of Interest) | NA | Sensors Traffic surveillance |
| Wang et al. [80] | Spatial temporal graph neural network | PEMS–BAY, META–LA | – | 53,116 videos from 325 sensors, 34,272 videos from 207 sensors | Yes—link | Sensors Traffic surveillance |
| Bao et al. [88] | Spatiotemporal GCN (graph convolution network) | CCD (Car Crash Dataset), DAD and A3D | – | 6.35 h, varied weather conditions (snow, day and night), 2.43 h and 3.56 h | Yes—link | Dashcam |
| Reddy et al. [26] | Deep Q-learning | Traffic driving data | – | 182 drive sequences | NA | Dashcam |
| Fernandez et al. [84] | Two-stream network | PREVENTION dataset | – | 6 h video (80 m around ego–vehicle), 3 radars 2 cameras and 1 LiDAR | Yes—link | LiDAR Radar, DashCam |
| Ali et al. [22] | Dynamic deep spatiotemporal neural network (DHSTNet) | TaxiBj, Bike NYC | – | 16 months video recordings | NA | Sensors |
| Wang et al. [21] | Spatial–temporal mixed attention graph-based convolution model (STMAG) | Curated Traffic data | – | 2000 videos | No-future release | Dashcam |
| Alkandari et al. [89] | Dynamic webster with dynamic cycle time algorithm (DWDC) | – | – | – | NA | Sensors |
| Riaz et al. [90] | FWPredNet | KITTI, HTA, D2City | 65 | 600 frames from Kitti, 286 clips, 65 frames and 678 video | Yes—link | Dashcam |
| Tang et al. [91] | Mixure LDA and expectation–maximization algorithm | 40–seconds traffic video | 15 | 600*800 frame dimension | NA | Traffic surveillance |
| Huang et al. [25] | CNN-traffic incident management (TIM) | Traffic management centers report, IOWA DOT radar sensors | – | 856 crash reports, 29 sensors | Yes—link | Sensors |
| Bortnikov et al. [92] | 3D convolutional neural network (CNN) | Video game GTA V, YouTube Car Accident Video | – | 5 h recording | NA | Traffic Surveillance, Dashcam |

**Table 5** (continued)

| Authors | Model | Dataset | Fps | Dataset feature | Accessibility | Data collection approach |
|---------|-------|---------|-----|-----------------|---------------|--------------------------|
| Gupta et al. [93] | Time-distrib-uted RNN | DETRAC dataset | 25 | 10 h recording of 376 videos, 99 frames selected from each video | Yes—link | Traffic Surveil-lance |
| Yang et al. [94] | Feature-fused SSD detector | Highway vehicle dataset, ImageNet VID dataset | – | 32,938 vehicle samples, 5354 videos | Yes—link | Traffic Surveil-lance |
| Ijjina et al. [83] | Mask R-CNN | YouTube acci-dent videos | 30 | 20 s video chunks, varied weather (harsh sunlight, day-light hours) | NA | Traffic Surveil-lance |
| You et al. [85] | Single-stream temporal action proposals (SST) | Causality in traffic accident (CTA) | – | 9.53 h video from 1935 videos, 18 semantic cause and 7 semantic effect labels | Yes—link | Dashcam |
| Srinivasan et al. [24] | Detection transformers and random forest classifier (DETR) | CADP | – | 1416 Accident footage | Yes—link | Traffic surveil-lance |
| Hui et al. [95] | Gaussian mixture model (GMM) | – | – | – | NA | Sensors |
| Min et al. [87] | Sparse topic model | QMUL junction dataset, AVSS dataset | 25 | 52 min traffic video, 4 s chunks | Yes—link | Traffic surveil-lance |
| Vatti et al. [96] | Accident detection and communication system | – | – | – | NA | Sensors |

similar result when compared to segmentation-based approaches and newer methods like Residual 2D+1D convolutional network-R(2+1)D and SlowFast network. Current state-of-the-art techniques still have limitations in accurately capturing traffic anomalies on the DoTA datasets, especially for occluded and distant objects. Further exploration of AR techniques such as the R(2+1)D, SlowFast network, and DETR algorithm has shown promising results in terms of accuracy, performance, and detection speed. Additionally, it would be beneficial to research methods that can better handle occluded objects, distant objects, and horizontal vehicle collisions. Improving the performance of these techniques and addressing their limitations will enhance the overall safety and efficiency of accident detection and autonomous transportation systems within smart cities.

### Key summary and recommendations RQ2

Yao et al. [86] proposed methodology is beneficial in unsupervised scenarios and can be deployed to detect traffic anomalies in real time especially with the lack of anno-tated datasets publicly available. Unsupervised video anomaly detection calculates traffic anomaly scores using the spatial–temporal area under the curve (STAUC), and

the future object localization (FOL) method in detecting anomalous events in videos. Another technique by Reddy et al. [26] combines GNN, RNN, and a transformer layer to model complex topological and temporal relationships in traffic video data. By capturing both spatial and temporal information, the spatiotemporal graph neural network outperforms RNN and LSTM-based methods in predicting traffic accidents. Similarly, Yu et al. [23] utilized a graph-based spatiotemporal model for predicting future traffic accidents by integrating spatial, temporal, and external features to improve the overall accuracy of accident prediction. The model's performance is around 5% better than spatial autoencoder (SAE). An other methodology for capturing external features combined graph convolutional network with a dynamic deep hybrid spatiotemporal neural network (DHSTNet) to capture spatial dependencies of dynamic traffic flow and uses LSTM cells to capture temporal correlations with external features [22]. Other taxonomies in accident detection and autonomous transportation includes deep Q-learning with YOLOv3. This hybrid approach combines the strengths of deep Q-learning and YOLOv3 for efficient object detection in traffic scenes [26]. Hierarchical recurrent neural network (HRNN) focused specifically on detecting accidents from CCTV surveillance. The method explores temporal and spatial features of video footage to identify accident occurrences effectively [92]. Yang et al. [94] proposed tracking-based object detection (TDO) and feature-fused SSD technique for improving detection results over state-of-the-art methods and established vehicle datasets for highway scene analysis. Based on our findings from RQ2, we recommend exploring the potential benefits of combining hybrid taxonomies and multiple algorithms to better capture the complex spatial and temporal relationships in traffic video data. For example, the spatiotemporal graph neural network by Reddy et al. [26] demonstrates the effectiveness of such an approach. However, it is essential to consider the potential trade-offs associated with these methods, such as increased computational costs and difficulties in real-time deployment. The implementation of complex models can pose challenges in terms of deployment on edge devices, such as cameras or other IoT devices due to their limited processing capability. Model compression or pruning techniques can optimize algorithms that require high processing power and memory capabilities.

**Key summary and recommendations RQ3**

Our analysis highlights the importance of various datasets for accident detection and autonomous vehicles. These datasets are collected from different sources such as dashcams, traffic surveillance cameras, drones, and cameras installed on buildings. Examples of popular datasets include HighD, InD, NGSIM HW101, NGSIM I-80, PKU, Prevention dataset, and Apolloscape dataset. Features in these datasets vary but include environmental trajectories, vehicle trajectory data, and footage captured under diverse ambient conditions. To improve the generalizability of action recognition models for accident detection, it is recommended that future research should consider utilizing diverse datasets that encompass various traffic conditions, weather conditions, and geographical locations as this has demonstrated improved model performance [23, 92] and also ensure that developed models can perform well in real-world scenarios. Furthermore, there is a need for benchmark datasets that can help assess the quality of traffic accident detection across different action classes, facilitating a fair comparison of the performance of

various models and techniques. While simulated game video data has been shown to yield comparable results to real-life traffic videos [92], it is essential to prioritize real-life data to ensure effectiveness of the developed models in real-world situations. Limited annotated real-life accident datasets pose a challenge for researchers, we believe that investing in annotating and sharing these datasets will encourage more researchers to develop sophisticated methodologies and algorithms specific to accident detection and help improve the performance of action recognition models. This will ultimately contribute to advancement of the field but also support the goal of interconnected smart city automation, enhance traffic safety and efficiency in urban environments.

## Limitation

Our research focused on research papers relevant to action recognition, accident detection, and autonomous transportation. Our systematic literature review has some limitations due to the inclusion and exclusion criteria we applied during the search process. The time constraint of including only articles published within the last ten years (between 2012 and 2022) might lead to the exclusion of relevant research that was published before 2012. This could potentially limit our understanding of the evolution of action recognition techniques and their application in accident detection. Based on our criteria to exclude articles written in other languages other than English, we may have excluded research findings and advancements in action recognition and accident detection from non-English-speaking research communities. This language barrier may limit the comprehensiveness of our review and result in potential biases in our findings. Furthermore, the requirement for studies to include validation of their proposed techniques and contain analysis of spatial/temporal information may have led to the exclusion of some potentially relevant studies that focused on theoretical developments, proposed novel techniques without immediate validation, or used alternative methods for action recognition. These limitations could affect the overall comprehensiveness and generalizability of our systematic literature review.

## Conclusion

This systematic literature review aims to determine state-of-the-art action recognition for accident detection and autonomous transportation in smart cities. We used the PRISMA guideline for selecting seminary articles related to our topic domain, and this guideline was based on the inclusion and exclusion criteria discussed in "Literature search" section. We selected 21 papers from an initial list of 2030 publications, and we categorized and analyzed relevant papers based on the three pillars of our research question. This paper discussed the leading techniques and applications of action recognition in autonomous transportation. The study also explored the main taxonomies and algorithms used in AR for autonomous transportation. Finally, we presented an overview of datasets used in AR for autonomous transportation, features of the datasets, and download links to the datasets are embedded in Table 5 accessibility column.

In the quest for a smart city, automating city traffic by capturing spatial and temporal information from DNN is a significant step in smart city automation. Bao et al. [88] developed a model to handle the challenges of relational feature learning and uncertainty anticipation from traffic video to predict accident occurrence within 3.53 s with

an average precision of 72.22% using graph convolution network (GCN) and Bayesian neural networks (BNNs). Several factors are involved in traffic accident detection, including driver behavior, weather conditions, traffic flow, and road structure. Yu et al. [23] examined spatial–temporal relationships on heterogeneous data to develop a road-level accident prediction system. Besides sequential patterns in the temporal dimension, traffic flow is strongly affected by other road networks in the spatial dimension. Studies have been conducted on traffic flow prediction; however, many of them cannot account for spatial and temporal dependencies [80]. Reddy et al. [26] aimed to extract roadway characteristics that are relevant to the trajectory of an autonomous vehicle from real-world road conditions using deep Q-learning. Analyzing and forecasting dynamic traffic patterns within smart cities is necessary for planning and managing transportation. Forecasting traffic flow is more difficult because of the volatility of vehicle flow in the temporal dimension and the uncertainty related to accident occurrence and traffic movements. Ali et al. [22] proposed a hybrid model composed of GCN and DHSTNet, which can forecast short-term traffic patterns in urban areas for improved traffic management. Similarly, Alkandari et al. [89] developed a methodology for determining how long a vehicle stays in traffic based on traffic flow and congestion.

Automation of accident detection using AI systems based on traffic cameras will be a step towards the security of more lives. It will also support the transformation of traffic cameras to support smart city automation in providing first responders and law enforcement agencies with information about road accidents. Based on the foregoing we recommend;

- Experimental research study to combine action recognition techniques for objects and human action classification since they both have been developed using similar model architectures.
- Future reviews in this area should consider addressing the limitations of this study by including a broader range of publication years, languages, and publication types to ensure a more comprehensive understanding of action recognition techniques and their application in accident detection.
- Future research should focus on scaling up accident detection systems that can be integrated into smart city automation for alerting first responders about traffic accidents.

Finally, adopting automated accident detection system will support first responders in providing a quick response to victims thereby reducing human error and response time.

**Availability of data and materials**
The datasets used and spreadsheets analyzed during the current study are available from the corresponding author upon reasonable request.

## Declarations

**Competing interests**
The authors declare that there are no conflicts of interest.

## References

1. Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian VN (2017) Grad-cam: improved visual explanations for deep convolutional networks. arXiv:1710.11063
2. Al-Faris M, Chiverton J, Ndzi D, Ahmed AI (2020) A review on computer vision-based methods for human action recognition. J Imaging 6(6):46
3. Bo W, Fuqi M, Rong J, Peng L, Xuzhu D (2021) Skeleton-based violation action recognition method for safety supervision in the operation field of distribution network based on graph convolutional network. CSEE J Power Energy Syst
4. Muhammad K, Ullah A, Imran AS, Sajjad M, Kiran MS, Sannino G, de Albuquerque VHC et al (2021) Human action recognition using attention based lstm network with dilated cnn features. Futur Gener Comput Syst 125:820–830
5. Gabrielli M, Leo P, Renzi F, Bergamaschi S (2019) Action recognition to estimate activities of daily living (adl) of elderly people. In: 2019 IEEE 23rd international symposium on consumer technologies (ISCT), pp 261–264. IEEE
6. Ren H, Xu, G (2002) Human action recognition in smart classroom. In: Proceedings of Fifth IEEE international conference on automatic face gesture recognition, pp 417–422. IEEE
7. Gedamu K, Ji Y, Yang Y, Gao L, Shen HT (2021) Arbitrary-view human action recognition via novel-view action generation. Pattern Recogn 118:108043
8. Vrigkas M, Nikou C, Kakadiaris IA (2015) A review of human activity recognition methods. Front Robot AI. https://doi.org/10.3389/frobt.2015.00028
9. Lv F, Nevatia R (2007) Single view human action recognition using key pose matching and viterbi path searching. In: 2007 IEEE conference on computer vision and pattern recognition, pp 1–8. IEEE
10. Jhuang H, Gall J, Zuffi S, Schmid C, Black MJ (2013) Towards understanding action recognition. In: Proceedings of the IEEE international conference on computer vision, pp 3192–3199
11. Yao A, Gall J, Fanelli G, Van Gool L (2011) Does human action recognition benefit from pose estimation?. In: Proceedings of the 22nd British machine vision conference-BMVC 2011. BMV press
12. Xiaohan Nie B, Xiong C, Zhu S-C (2015) Joint action recognition and pose estimation from video. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1293–1301
13. Cheron G, Laptev I, Schmid C (2015) P-cnn: Pose-based cnn features for action recognition. In: Proceedings of the IEEE international conference on computer vision, pp 3218–3226
14. Raja K, Laptev I, Perez P, Oisel L (2011) Joint pose estimation and action recognition in image graphs. In: 2011 18th IEEE international conference on image processing, pp 25–28. IEEE
15. Iqbal A, Richard A, Gall, J (2019) Enhancing temporal action localization with transfer learning from action recognition. In: 2019 IEEE/CVF international conference on computer vision workshop (ICCVW), pp 1533–1540. IEEE
16. Sevilla-Lara L, Liao Y, Guney F, Jampani V, Geiger A, Black MJ (2018) On the integration of optical flow and action recognition. German conference on pattern recognition. Springer, Berlin, pp 281–297
17. Kuehne H, Arslan A, Serre T (2014) The language of actions: Recovering the syntax and semantics of goal-directed human activities. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 780–787
18. Lea C, Reiter A, Vidal R, Hager GD (2016) Segmental spatiotemporal cnns for fine-grained action segmentation. European conference on computer vision. Springer, Berlin, pp 36–52
19. Joshua SC, Garber NJ (1990) Estimating truck accident rate and involvements using linear and poisson regression models. Transp Plan Technol 15(1):41–58
20. Arvin R, Kamrani M, Khattak AJ (2019) How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data. Accid Anal Prev 127:118–133
21. Wang J, Chen Q, Gong H (2020) STMAG: A spatial-temporal mixed attention graph-based convolution model for multi-data flow safety prediction. Inf Sci 525:16–36. https://doi.org/10.1016/j.ins.2020.03.040
22. Ali A, Zhu Y, Zakarya M (2022) Exploiting dynamic spatio-temporal graph convolutional neural networks for city-wide traffic flows prediction. Neural Netw 145:233–247. https://doi.org/10.1016/j.neunet.2021.10.021
23. Yu L, Du B, Hu X, Sun L, Han L, Lv W (2021) Deep spatio-temporal graph convolutional network for traffic accident prediction. Neurocomputing 423:135–147. https://doi.org/10.1016/j.neucom.2020.09.043
24. Srinivasan A, Srikanth A, Indrajit H, Narasimhan V (2020) A novel approach for road accident detection using detr algorithm. In: 2020 international conference on intelligent data science technologies and applications (IDSTA), pp 75–80. IEEE
25. Huang T, Wang S, Sharma A (2020) Highway crash detection and risk estimation using deep learning. Accid Anal Prevent. https://doi.org/10.1016/j.aap.2019.105392

26. Reddy DR, Chella C, Teja KBR, Baby HR, Kodali P (2021) Autonomous vehicle based on deep q-learning and yolov3 with data augmentation. In: 2021 International conference on communication, control and information sciences (ICCISc), vol 1, pp 1–7. IEEE

27. Sharghi A, Haugerud H, Oh D, Mohareri O (2020) Automatic operating room surgical activity recognition for robot-assisted surgery. International conference on medical image computing and computer-assisted intervention. Springer, Berlin, pp 385–395

28. Ahmidi N, Tao L, Sefati S, Gao Y, Lea C, Haro BB, Zappella L, Khudanpur S, Vidal R, Hager GD (2017) A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery. IEEE Trans Biomed Eng 64(9):2025–2041

29. Zhou E, Zhang H (2020) Human action recognition toward massive-scale sport sceneries based on deep multi-model feature fusion. Signal Process Image Commun 84:115802

30. Davar NF, de Campos T, Windridge D, Kittler J, Christmas W (2011) Domain adaptation in the context of sport video action recognition. In: Domain adaptation workshop, in conjunction with NIPS

31. Al Zamil MG, Samarah S, Rawashdeh M, Karime A, Hossain MS (2019) Multimedia-oriented action recognition in smart city-based iot using multilayer perceptron. Multimed Tools Appl 78(21):30315–30329

32. Dhulekar P, Gandhe S, Chitte H, Pardeshi K (2017) Human action recognition: an overview. In: Proceedings of the international conference on data engineering and communication technology, pp 481–488. Springer

33. Kamthe U, Patil C (2018) Suspicious activity recognition in video surveillance system. In: 2018 fourth international conference on computing communication control and automation (ICCUBEA), pp 1–6. IEEE

34. Adewopo V, Elsayed N, Anderson K (2022) Baby physical safety monitoring in smart home using action recognition system. arXiv preprint arXiv:2210.12527

35. Yan J, Yan S, Zhao L, Wang Z, Liang Y (2019) Research on human-machine task collaboration based on action recognition. In: 2019 IEEE international conference on smart manufacturing, industrial and logistics engineering (SMILE), pp. 117–121. IEEE

36. Khan FS, Van De Weijer J, Anwer RM, Felsberg M, Gatta C (2014) Semantic pyramids for gender and action recognition. IEEE Trans Image Process 23(8):3633–3645

37. Badawy M, Ramadan N, Hefny HA (2023) Healthcare predictive analytics using machine learning and deep learning techniques: a survey. J Electr Syst Inf Technol 10(1):40

38. Kruger V, Kragic D, Ude A, Geib C (2007) The meaning of action: A review on action recognition and mapping. Adv Robot 21(13):1473–1501

39. Rodrıguez-Moreno I, Martınez-Otzeta JM, Goienetxea I, Rodriguez-Rodriguez I, Sierra B (2020) Shedding light on people action recognition in social robotics by means of common spatial patterns. Sensors 20(8):2436

40. Fortun D, Bouthemy P, Kervrann C (2015) Optical flow modeling and computation: a survey. Comput Vis Image Underst 134:1–21. https://doi.org/10.1016/j.cviu.2015.02.008

41. Montori F, Bedogni L, Bononi L (2018) A collaborative internet of things architecture for smart cities and environmental monitoring. IEEE Int Things J 5(2):592–605. https://doi.org/10.1109/JIOT.2017.2720855

42. Elsayed N, Zaghloul ZS, Azumah SW, Li C (2021) Intrusion detection system in smart home network using bidirectional lstm and convolutional neural networks hybrid model. In: 2021 IEEE international midwest symposium on circuits and systems (MWSCAS), pp 55–58. IEEE

43. Azumah SW, Elsayed N, Adewopo V, Zaghloul ZS, Li C (2021) A deep lstm based approach for intrusion detection IoT devices network in smart home. In: 2021 IEEE 7th world forum on internet of things (WF-IoT), pp 836–841. IEEE

44. Romero D, Salamea C (2019) Convolutional models for the detection of firearms in surveillance videos. Appl Sci 9(15):2965

45. Jamil H, Ali KM, Kim D-H (2023) Federated recognition mechanism based on enhanced temporal-spatial learning using mobile edge sensors for firefighters. Fire Ecol 19(1):44

46. Patil VK, Pawar VR, Randive S, Bankar RR, Yende D, Patil AK (2023) From face detection to emotion recognition on the framework of raspberry pi and galvanic skin response sensor for visual and physiological biosignals. J Electr Syst Inf Technol 10(1):1–27

47. Stewart T (2022) Overview of motor vehicle crashes in 2020. Technical report

48. Celaya-Padilla JM, Galvan-Tejada CE, Lozano-Aguilar JSA, Zanella-Calzada LA, Luna-Garcia H, Galvan-Tejada JI, Gamboa-Rosales NK, Velez Rodriguez A, Gamboa-Rosales H (2019) "Texting and driving" detection using deep convolutional neural networks. Appl Sci 9(15):2962

49. Avazov K, Mukhiddinov M, Makhmudov F, Cho YI (2021) Fire detection method in smart city environments using a deep-learning-based approach. Electronics 11(1):73

50. Cai Y, Wang H, Chen X, Jiang H (2015) Trajectory-based anomalous behaviour detection for intelligent traffic surveillance. IET Intel Transp Syst 9(8):810–816. https://doi.org/10.1049/IET-ITS.2014.0238

51. Morris BT, Trivedi MM (2011) Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach. IEEE Trans Pattern Anal Mach Intell 33(11):2287–2301

52. Huang X, He P, Rangarajan A, Ranka S (2019) Intelligent intersection: two-stream convolutional networks for real-time near accident detection in traffic video. ACM Trans Spat Algorithms Syst 6(2):23. https://doi.org/10.1145/3373647.1901.01138

53. Saunier N, Sayed T (2007) Automated analysis of road safety with video data. Transp Res Rec 2019:57–64. https://doi.org/10.3141/2019-08

54. Robles-Serrano S, Sanchez-Torres G, Branch-Bedoya J (2021) Automatic detection of traffic accidents from video using deep learning techniques. Computers. https://doi.org/10.3390/COMPUTERS10110148

55. Lim W, Jang D, Lee T (2016) Speech emotion recognition using convolutional and recurrent neural networks. In: 2016 Asia-pacific signal and information processing association annual summit and conference (APSIPA), pp 1–4. IEEE

56. Elsayed N, Maida AS, Bayoumi M (2019) Reduced-gate convolutional lstm architecture for next-frame video prediction using predictive coding. In: 2019 international joint conference on neural networks (IJCNN), pp 1–9. IEEE

57. Kattenborn T, Leitloff J, Schiefer F, Hinz S (2021) Review on convolutional neural networks (cnn) in vegetation remote sensing. ISPRS J Photogramm Remote Sens 173:24–49
58. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2016) Lstm: a search space odyssey. IEEE Trans Neural Netw Learn Syst 28(10):2222–2232
59. Elsayed N, Maida AS, Bayoumi M (2020) Reduced-gate convolutional long short-term memory using predictive coding for spatiotemporal prediction. Comput Intell 36(3):910–939
60. Liang C (2023) Intelligent monitoring methodology for large-scale logistics transport vehicles based on parallel internet of vehicles. EURASIP J Wirel Commun Netw 2023(1):75
61. Carreira J, Zisserman A (2017) Quo Vadis, action recognition? A new model and the kinetics dataset. In: Proceedings - 30th IEEE conference on computer vision and pattern recognition, CVPR 2017, vol 2017, pp 4724–4733. https://doi.org/10.1109/CVPR.2017.502.1705.07750. https://arxiv.org/abs/1705.07750
62. Jordao A, Nazare AC, Sena J, Robson Schwartz W (2018) Human activity recognition based on wearable sensor data: a standardization of the state-of-the-art
63. Sensor Data: A Standardization of the State-of-the-Art. Technical report. https://arxiv.org/abs/1806.05226v3 http://www.sense.dcc.ufmg.br/activity-recognition-based-wearable-
64. Stisen A, Blunck H, Bhattacharya S, Prentow TS, Kjærgaard MB, Dey A, Sonne T, Jensen MM (2015) Smart devices are different: Assessing and mitigatingmobile sensing heterogeneities for activity recognition. In: Proceedings of the 13th ACM conference on embedded networked sensor systems, pp 127–140
65. Sayed SA, Abdel-Hamid Y, Hefny HA (2023) Artificial intelligence-based traffic flow prediction: a comprehensive review. J Electr Syst Inf Technol 10(1):13
66. Singh D, Mohan CK (2018) Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. IEEE Trans Intell Transp Syst 20(3):879–887
67. Chan F-H, Chen Y-T, Xiang Y, Sun M (2016) Anticipating accidents in dashcam videos. Asian conference on computer vision. Springer, Berlin, pp 136–153
68. Shi X, Chen Z, Wang H, Yeung D-Y, Wong W-K, Woo W-C (2015) Convolutional lstm network: a machine learning approach for precipitation nowcasting. Adv Neural Inf Process Syst **28**
69. Elsayed N, Maida AS, Bayoumi M (2018) Empirical activation function effects on unsupervised convolutional LSTM learning. In: 2018 IEEE 30th international conference on tools with artificial intelligence (ICTAI), pp 336–343. IEEE
70. Rashidi TH, Abbasi A, Maghrebi M, Hasan S, Waller TS (2017) Exploring the capacity of social media data for modelling travel behaviour: opportunities and challenges. Transp Res Part C Emerg Technol 75:197–211
71. Adewopo V, Gonen B, Elsayed N, Ozer M, Elsayed ZS (2022) Deep learning algorithm for threat detection in hackers forum (deep web). arXiv preprint arXiv:2202.01448
72. Xu S, Li S, Wen R (2018) Sensing and detecting traffic events using geosocial media data: a review. Comput Environ Urban Syst 72:146–160
73. Xu S, Li S, Wen R, Huang W (2019) Traffic event detection using twitter data based on association rules. ISPRS Ann Photogramm Remote Sens Spat Inf Sci 4:543–547
74. Salas A, Georgakis P, Nwagboso C, Ammari A, Petalas I (2017) Traffic event detection framework using social media. In: 2017 IEEE international conference on smart grid and smart cities (ICSGSC), pp 303–307
75. Gu Y, Qian ZS, Chen F (2016) From twitter to detector: Real-time traffic incident detection using social media data. Transp Res Part C Emerg Technol 67:321–342
76. Harris JD, Quatman CE, Manring M, Siston RA, Flanigan DC (2014) How to write a systematic review. Am J Sports Med 42(11):2761–2768
77. Wright RW, Brand RA, Dunn W, Spindler KP (2007) How to write a systematic review. Clin Orthop Relat Res 1976–2007(455):23–29
78. Gough D, Thomas J, Oliver S (2017) An introduction to systematic reviews. Introd Syst Rev 1–352
79. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C, et al (2021) The PRISMA 2020 statement: An updated guideline for reporting systematic reviews [Internet], vol 372, The BMJ. BMJ Publishing Group
80. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Moher D (2021) Updating guidance for reporting systematic reviews: development of the prisma 2020 statement. J Clin Epidemiol 134:103–112
81. Smola AJ, Scholkopf B (2004) A tutorial on support vector regression. Stat Compu 14(3):199–222
82. Wang X, Ma Y, Wang Y, Jin W, Wang X, Tang J, Jia C, Yu J (2020) Traffic flow prediction via spatial temporal graph neural network. In: Proceedings of the web conference 2020, pp 1082–1092
83. Zhang J, Shi X, Xie J, Ma H, King I, Yeung D-Y (2018) Gaan: Gated attention networks for learning on large and spatiotemporal graphs. arXiv preprint arXiv:1803.07294
84. Zhao L, Song Y, Zhang C, Liu Y, Wang P, Lin T, Deng M, Li H (2019) T-gcn: a temporal graph convolutional network for traffic prediction. IEEE Trans Intell Transp Syst 21(9):3848–3858
85. Ijjina EP, Chand D, Gupta S, Goutham K (2019) Computer vision-based accident detection in traffic surveillance. In: 2019 10th international conference on computing, communication and networking technologies (ICCCNT), pp 1–6. IEEE
86. Fernandez-Llorca D, Biparva M, Izquierdo-Gonzalo R, Tsotsos JK (2020) Two-stream networks for lane-change prediction of surrounding vehicles. In: 2020 IEEE 23rd international conference on intelligent transportation systems (ITSC), pp 1–6. IEEE
87. You T, Han B (2020)Traffic accident benchmark for causality recognition. In: Computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, pp 540–556. Springer
88. Yao Y, Wang X, Xu M, Pu Z, Wang Y, Atkins E, Crandall D (2022) DoTA: unsupervised detection of traffic anomaly in driving videos. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2022.3150763
89. Xia L-M, Hu X-J, Wang J (2018) Anomaly detection in traffic surveillance with sparse topic model. J Central South Univ 25(9):2245–2257. https://doi.org/10.1007/s11771-018-3910-9

90. Bao W, Yu Q, Kong Y (2020) Uncertainty-based traffic accident anticipation with spatio-temporal relational learning. In: Proceedings of the 28th ACM international conference on multimedia. MM '20, pp 2682–2690. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3394171.3413827

91. Alkandari A, Aljandal M (2015) Theory of dynamic fuzzy logic traffic light integrated system with accident detection and action, pp 62–68. https://doi.org/10.1109/ICCTIM.2015.7224594

92. Riaz W, Chenqiang G, Azeem A, Saifullah-Bux JA, Ullah A (2022) Traffic anomaly prediction system using predictive network. Remote Sens 14(3):447. https://doi.org/10.3390/rs14030447

93. Tang X, Huang X-L, Sun S-Y, Dong H, Zhang X, Gao Y, Liu N (2017) Intelligent recognition of traffic video based on mixture lda model. In: Lecture Notes of the institute for computer sciences, social-informatics and telecommunications engineering, vol 183, pp 356–363. Springer

94. Bortnikov M, Khan A, Khattak AM, Ahmad M (2019) Accident recognition via 3d cnns for automated traffic monitoring in smart cities. Advances in Intelligent Systems and Computing

95. Gupta G, Singh RK, Patel AS, Ojha M (2020) Accident detection using time-distributed model in videos

96. Yang B, Zhang S, Tian Y, Li B (2019) Front-vehicle detection in video images based on temporal and spatial characteristics, vol 19 https://doi.org/10.3390/s19071728.

97. Hui Z, Yao-hua X, Lu M, Jiansheng F (2014) Vision-based real-time traffic accident detection. In: Proceeding of the 11th world congress on intelligent control and automation, pp 1035–1038

98. Vatti NR, Vatti PL, Vatti RA, Garde CS (2018) Smart road accident detection and communication system. In: 2018 International conference on current trends towards converging technologies (ICCTCT), pp 1–4

99. Nowozin S (2014) Optimal decisions from probabilistic models: the intersection-over-union case. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 548–555

100. Krajewski R, Bock J, Kloeker L, Eckstein L (2018) The highd dataset: a drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems. In: 2018 21st international conference on intelligent transportation systems (ITSC), pp 2118–2125. IEEE

101. Zhan W, Sun L, Wang D, Shi H, Clausse A, Naumann M, Kummerle J, Konigshof H, Stiller C, de La Fortelle A et al (2019) Interaction dataset: an international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps. arXiv preprint arXiv:1910.03088

102. Colyar J, Halkias J (2007) Ngsim-us highway 101 dataset. Federal Highway Administration (FHWA), Tech. Rep. FHWA-HRT-07-030, DC, USA

103. Halkias J, Colyar J (2006)Ngsim interstate 80 freeway dataset. US Federal Highway Administration, FHWA-HRT-06-137, Washington, DC, USA

104. Zhao H, Wang C, Lin Y, Guillemard F, Geronimi S, Aioun F (2017) On-road vehicle trajectory collection and scene-based lane change analysis: Part i. IEEE Trans Intell Transp Syst 18(1):192–205. https://doi.org/10.1109/TITS.2016.2571726

105. Izquierdo R, Quintanar A, Parra I, Fernandez-Llorca D, Sotelo M (2019) The prevention dataset: a novel benchmark for prediction of vehicles intentions. In: 2019 IEEE intelligent transportation systems conference (ITSC), pp 3114–3121. IEEE

106. Wang P, Huang X, Cheng X, Zhou D, Geng Q, Yang R (2019) The apolloscape open dataset for autonomous driving and its application. IEEE Trans Pattern Anal Mach Intell 42(10):2702–2719

## Publisher's Note