Smart Cities, Smarter Roads: A Review of Leveraging Cutting-Edge Technologies for Intelligent Event Detection from Social Media

Ebtesam Ahmad Alomari[1], Rashid Mehmood[2]
Faculty of Computer Science and Information Technology, Albaha University, Albaha, Saudi Arabia[1]
Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, Saudi Arabia[2]

***Abstract*—The rapidly evolving landscape of smart cities and intelligent transportation systems makes the timely detection of traffic events a critical element for optimizing urban mobility. Furthermore, social media emerges as a valuable source of real-time information, with users acting as active sensors who spontaneously share observations and experiences related to traffic incidents. This review paper offers a comprehensive understanding of the state-of-the-art in traffic event detection from social media. The paper explores leveraging cutting-edge technologies including machine learning, and deep learning with big data technologies and high-performance computing. The discussion unfolds with an in-depth examination of the recent approaches for event detection followed by an exploration of the techniques of spatio-temporal information extraction and sentiment analysis, which are both considered fundamental aspects in enhancing the contextual understanding of traffic events. Further, the review explores the pivotal role of big data technologies in addressing scalability challenges inherent in the vast expanse of social data. The examination encompasses how big data frameworks facilitate efficient storage, processing, and analysis of large-scale social media datasets, thereby empowering machine learning and deep learning models for robust and real-time traffic event detection. Subsequently, the challenges and future directions have been highlighted. Addressing these challenges and leveraging advanced technologies, facilitates the proactive detection and management of these events, paving the way for smart mobility systems.**

***Keywords*—Mobility; smart cities; event detection; social media; big data analytics**

I. Introduction

The importance of detecting traffic events (incidents) cannot be overstated in the context of smart mobility and cities. Efficient traffic event detection lies in its impact on traffic flow optimization, safety enhancement, urban environment functionality and sustainability as well as the overall quality of life in smart cities. Therefore, rapid identification of incidents, such as accidents, road closures, or adverse weather conditions, helps authorities to make timely decisions, minimizing the negative impact of these events and contributing to the seamless operation of transportation systems.

Moreover, social media platforms become vast and decentralized sources of information, with their active users, who function as sensors spontaneously sharing observations, experiences, and updates related to topics in different domains including traffic. Subsequently, Twitter has widely been used to enable smart mobility systems, such as for traffic congestion estimation [1], passenger flow prediction in public metro transit systems [2], understanding taxi traffic dynamics [3], and detecting traffic anomalies caused by traffic accidents, disasters, etc. [4]. Besides, other social media have been used as well, for instance, public geotagged Instagram posts are used to detect an abnormal increase or decrease of the citizen's number in a specific area at a specific time by applying a density-based clustering algorithm [5]. Furthermore, several works have been focused on detecting events using social data, which raises the need for review papers that discuss the significant previous contributions and highlight future directions.

This review paper explores the detection of traffic events through analysis of social media data leveraging artificial intelligence with a specific focus on machine learning, and deep learning techniques in conjunction with big data technologies. The paper begins with examining the approaches of events detection and then investigates the techniques of Spatio-temporal information extraction and sentiment analysis. Furthermore, the review explores using big data technologies. The importance of utilizing big data technology in analyzing social data cannot be overstated, due to the volume, velocity, and variety of data generated on social media platforms, which require scalable and efficient processing frameworks. By leveraging advanced analytics and machine learning on large datasets, big data technology empowers the development of sophisticated models for traffic event detection, ensuring timely and accurate responses to dynamic urban challenges. To the best of our knowledge, only one review paper [6] provides a systematic review of traffic event detection techniques from social data but it does not includes the recent approaches such as using big data technology.

The main contribution of our work can be summarized as follows:

1) Cover the recent approaches including using deep learning and big data technologies as well as provide a taxonomy of the approaches.

2) Discuss the techniques of spatio-temporal information extraction and sentiment analysis, which are both considered

as fundamental aspects in enhancing the contextual understanding of traffic events.

3) Highlight the challenges and future directions to facilitate the proactive detection and management of the events and pave the way for smart mobility systems.

The rest of the paper is organized as follows. Section II reviews the works on traffic-related event detection from social data. Section III discusses using big data technologies in mobility and reviews the works related to traffic event detection using big data platforms and technologies. Section IV discusses the challenges and future directions. Finally, we draw our conclusions in Section V.

II. Social Media in Traffic Event Detection

With the exponential growth of social media platforms, users actively share real-time information, offering a valuable resource for monitoring and responding to traffic-related incidents. In this section, we review the existing work on road traffic analysis and event detection from social media. Fig. 1 depicts the number of reviewed papers in the period between 2012 and 2023. The total number of papers related to traffic event detection from social media included in this review is 61, only 15 of them are using big data technologies and platforms. Moreover, Fig. 2 shows the used social media and the number of reviewed papers that used big data technologies (light blue color) and that do not use them (blue color). It can be seen that most of the works have used Twitter while only one paper has used Instagram.

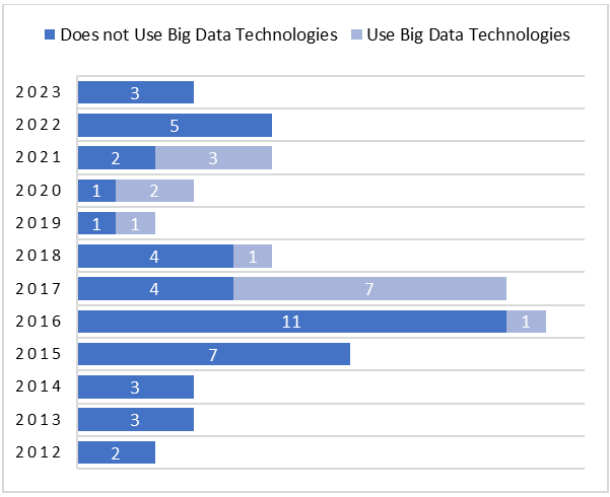

Fig. 1. Number of publications per year.

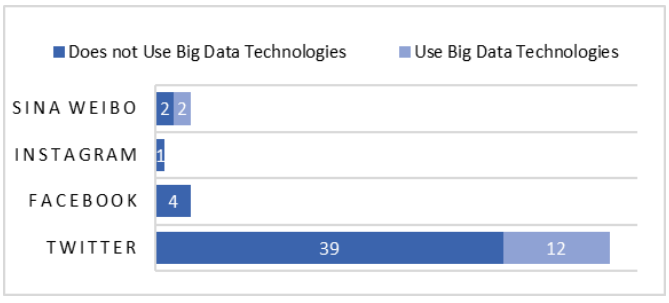

Fig. 2. Number of publications and social media platforms.

Fig. 3 shows the analysis dataset languages that have been used in the reviewed works, which include English, Arabic, Italian, Thai, Indonesian, Japanese, Malay, Korean, French, Spanish and Chinese. The blue color represents the number of papers that do not use big data technologies while the red represents the number of papers that use big data technologies. As depicted in the figure most of the works are focused on the English language.

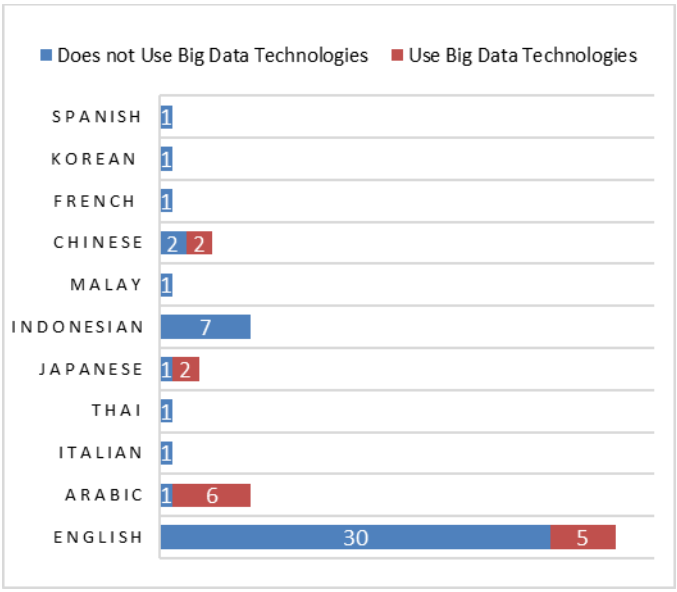

Fig. 3. Number of publications and dataset language.

Moreover, Fig. 4 describes the general workflow of event detection from social data. The main components are Data collection and filtering, Pre-processing, Event detection, Spatio-temporal information extraction, and Evaluation. Firstly, data are collected by using keywords, hashtags, accounts, geo-coordinates or a combination of these approaches. Then, the data are filtered to keep only traffic-related data. Data filtering can be done during data collection or it can be a separate step and various approaches can be used including machine learning algorithms. The next step is data pre-processing. There are common sub-components for pre-processing, which are tokenizer, normalizer, stop-words removal, and stemmer. Several tools and packages are available for pre-processing but mostly they are designed for a specific language. The next step is detecting the events and then extracting the time and location information. Subsequently, sentiment analysis is applied to understand the feelings and emotions regarding the detected events. This step is not mandatory. It depends on the interest of the researcher and the aim of the work. Finally, the tool is evaluated. The evaluation process depends on the method that has been used. For instance, if machine learning algorithms have been used to build classifiers for event detection, there are common evaluation metrics that can be used for evaluation such as accuracy, precision and recall. These are the main common steps that have been followed in literature to detect traffic events, more details about the components can be found in [7].

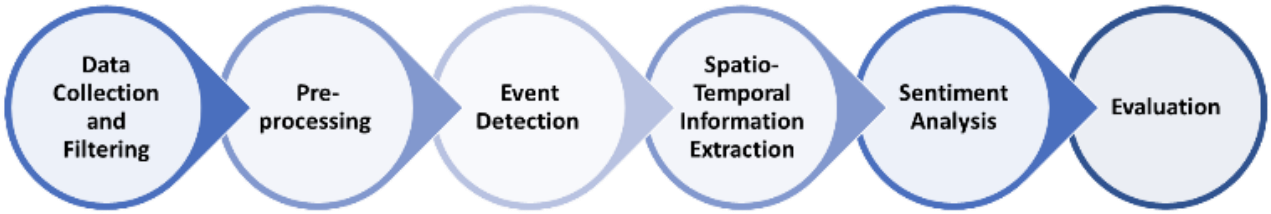

Fig. 4. General workflow of social data analysis for Traffic event detection.

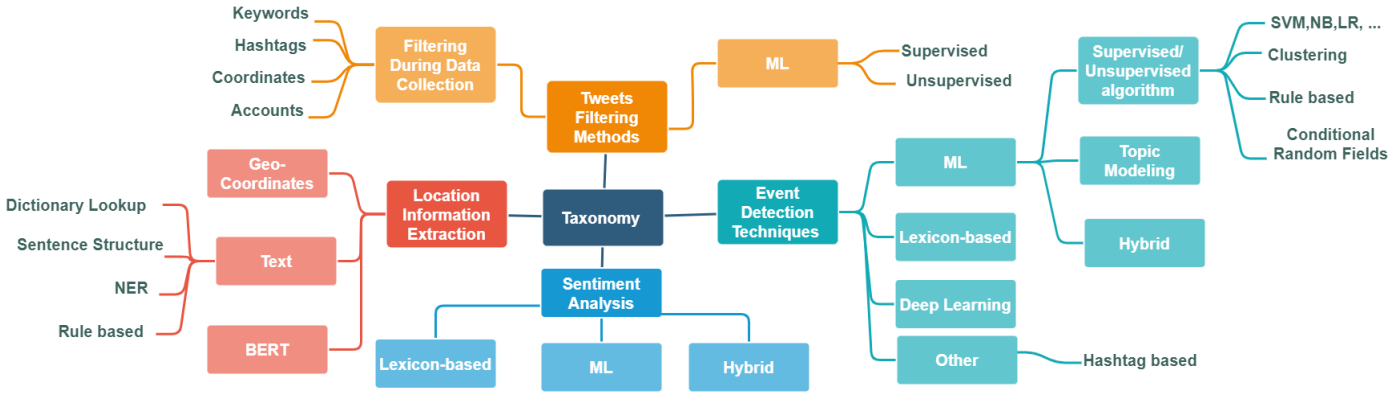

Fig. 5. Taxonomy of traffic events detection techniques from social data.

Fig. 5 shows the taxonomy of the approaches for each component in the general workflow except for the pre-processing component, which mainly depends on the language of the analyzed data and the evaluation component, which depends on the approach that has been used for the event detection. In the next subsection, we review the existing works and organize them based on the approaches that have been used. We started in section A with the event detection approaches because it is the main and the most important component. At the end of section A, we summarized the techniques for data filtering. In section B, we discuss the techniques of Spatio-temporal information extraction. Finally, section C covers the techniques of sentiment analysis

A. Event Detection Techniques

Several techniques have been used for traffic event detection including lexicon-based, machine learning (ML) and deep learning techniques. The next subsections provide a review of the existing works for each technique.

1) Lexicon-based: Istiq et al. [8] built an application to monitor road conditions. They collected road condition information from RSS fed on Facebook and stored them in the MySQL database. After Text Mining, the application categorized the information based on how much connectivity between words by categories. The information is categorized into six types including floods, traffic jams, congested roads, road damage, accidents, and landslides. Daly et al. [9] developed Dub-STAR system to extract the causes of traffic conditions from real-time tweets. They applied a simple dictionary approach to assigning tags to the input messages. The text is classified into classes such as delay, incident, roadwork, or concert. Moreover, Alomari and Mehmood [10] analyzed Arabic tweets related to traffic congestion in Jeddah city. They created custom dictionaries in SAP HANA for the common Arabic keywords about transportation and traffic congestion.

Alkouz and Alghbari [11] proposed a tool called SNSJam for traffic event detection using Arabic and English posts from Twitter and Instagram. They collected 50 million posts but after filtering they got around 2.3 million tweets and 4k Instagram posts. To identify events, they used a list of keywords.

2) Machine learning (ML)

a) Supervised and Unsupervised ML: Researchers in [12] and [13] used the SVM algorithm to classify tweets as relevant and irrelevant to traffic. But they don't detect event types. Dhavase and Bagade [14] provided a classification technique based on a machine-learning algorithm to detect crime-disaster events. Agarwal et al. [15] detected complaints tweets about road irregularities and bad road conditions by employing a rule-based classifier. They extracted the important information; such as the problem and the location, and then the tweets were categorized into useful, nearly-useful and irrelevant complaint reports.

Furthermore, Sakaki et al. [16] detected weather information and heavy-traffic information by classifying tweets into positive class means event-related and negative class means not related to events using the SVM algorithm. Klaithin and Haruechaiyasak [17] applied a machine learning classifier based on Naive Bayes Model. They classified the Thai tweets about traffic into three types: tweets about location, tweets about roads and tweets about traffic status. Kumar et al. [18] built a model to detect road hazards so they

classified the tweets into two classes namely, having negative or non-negative sentiment. They considered that all negative sentiment tweets have road hazard information. They applied three algorithms, which are Naïve Bayes (NB), K-nearest neighbors (KNN), and the Dynamic Language Model (DLM).

Moreover, D'Andrea et al. [19] applied text mining techniques and classified real-time Italian tweets into three classes, which are "traffic due to an external event", "traffic congestion or crash", and "non-traffic". They validated the detection of temporal information by their system using a generated dataset for traffic events from local newspapers and official news websites.

Kurniawan et al. [20] built a machine learning classifier to classify real-time Indonesian tweets into traffic-related and not-related. They used NB, SVM and decision tree (DT) algorithms. They counted the occurrences of all the words in a tweet as features. Then, in the feature selection process, they used only 40 selected words. Nguyen et al. [21] developed a system called TrafficWatch to monitor traffic conditions in Australia using real-time and historical tweets. During the feature selection process, they used serval features including a bag of words, Pattern recognition, Lemma, part-of-speech and a bag of tags. They introduced the NER annotation schema for detecting traffic related entities. Additionally, they trained the model based on Conditional Random Fields (CRFs). The traffic-related entities are implemented as key features to train the ML classification model. Further, they implemented an online clustering algorithm to incrementally cluster the streaming tweets.

Yazici et al. [22] used TF-IDF and NB to detect events from personal and organizational Twitter accounts. They noticed that the tweets from personal accounts do not disseminate traffic incident information in a very structured manner in terms of grammar and spelling compared with the organizational account. On the other side, personal tweets are better in reporting events that have just happened. Besides, Semwal *et al.* [23] and Tejaswin *et al.* [24] predicted the traffic incidents from social media using a random forest classifier. Anantharam et al. [25] developed a framework for extracting city-related events. They created a training set for building a conditional random field (CRF) automatically, by using dictionary-based spotting of event and location terms, to reduce manual tagging effort.

Moreover, Langa and Moeti [26] filtered real-time tweets about congestion in South Africa into traffic-congested-route and non-traffic-congested tweets using the Naïve Bayes algorithm. Subsequently, they developed a mobile application to send notifications to users based on classified tweets about traffic. Suat-Rojas et al. [27] built a classifier to detect accidents from tweets in the Spanish language. They classified tweets into accident and non-accident. For validation, they compared the detected accidents with official data from the Bogota Mobility Secretariat. Subsequently, they found that using doc2vec for feature extraction and SVM for classification helped in achieving good accuracy results.

3) Topic Modeling: One of the popular topic modeling algorithms is Latent Dirichlet Allocation (LDA). It was introduced by Blei et al. [28]. Besides, it is a statistical classification model based on the word-topic frequency distribution. Wang et al. [29] focused on exploring the relationship between traffic conditions in daily matter and urban human events in Toronto, Canada. They developed a tweet-LDA engine to classify tweets into two classes namely, "traffic-relevant" and "traffic-irrelevant". To ensure that no relevant tweets are missed, they filter out traffic-irrelevant tweets using a set of keywords. Moreover, Huang et al. [30] used DBSCAN to build spatiotemporal clusters from the geotagged tweets. Then, the LDA was implemented to the tweets under each cluster to extract the topics. Ali et al. [31] applied OLDA-based traffic event labeling. They used a well-known Web Ontology Language (OWL) tool called Prot´eg´e to develop ontology before using LDA. They generated a list of the most frequent words related to traffic events and then used them to develop ontology-based semantic knowledge.

a) Hybrid (Topic Modeling and Supervised/ Unsupervised ML): Gu et al. [32] collected historical and real-time tweets about traffic in Pittsburgh and Philadelphia Metropolitan. They used Supervised Latent Dirichlet Allocation (sLDA). In addition, the Tweets are classified by a Semi-Naive-Bayes (SNB) classifier. Lau [33] used Latent Dirichlet Allocation (LDA) topic modeling to filter traffic messages in the Chinese language. Additionally, they built SVM, KNN and NB classifiers for traffic events detection.

4) Deep Learning: Chen et al. [34] extracted Chinese traffic information from the Sina Weibo platform. They employed deep neural networks to learn the abstract features and classify them into traffic relevant and irrelevant. Ji et al.[35] measured the similarity of events between texts using meta-path in heterogeneous information networks. Then, to get the best possible meta-path weights, they employed a graph neural network for semi-supervised learning. After that, they designed a clustering algorithm to identify the traffic event categories.

Moreover, Ambastha and Desarkar [36] used different ML algorithms including SVM, Naive Bayes and Random Forest as well as deep learning algorithms including LSTM, CNN, and Universal Language Model FIne-Tuning (ULMFIT). The developed classification models were used to classify the tweet into two categories, namely, "Traffic incident related" or "Non-Traffic incident related". The Transfer Learning approach using ULMFIT was employed to enhance the performance.

Kim et al. [37] classified tweets in the Korean language using NB, RF, SVC, linear SVC, BiLSTM, and TextCNN into six classes: construction, weather, accident, traffic jam, crowded event and others. Rifqi et al. [38] used CNN+Word2Vec, CNN+FastText, and SVM to classify Indonesian tweets into 2 classes, which are tweets about traffic jams and tweets about smooth traffic. Swapnika and Vasumathi [39] applied the DNN and Harris Hawk optimization (HHO) algorithm to detect the following events from Twitter: education, transportation, environment, geospatial and water events. They claimed that they addressed

the challenge of the scalability challenge, but they did not provide a detailed explanation.

Furthermore, Hodorog et al. [42] combined AWD-LSTM and ULMFiT to detect traffic-related events in Cardiff City in the UK. They focused on several events including congestion, accidents, social gatherings, thefts, bus queues, floods and electricity charges. Additionally, they studied the relationship between the events as well as assigned a citizen satisfaction value to each of them. Mehri et al. [43] used BERT to detect subway-related events from tweets in English and French Language. They manually labeled 10381 records in English and 11008 in French to build the training dataset. The finding indicated that BERT in zero-shot surpasses the performance of the baseline models.

TABLE I. EVENT DETECTION TECHNIQUES

Ref.	Detected Event types	Event Detection Technique
Lexicon-based		
[8]	Floods, traffic jams, road damage, accidents, landslides	Based on the connectivity between words
[9]	Traffic congestion causes	Dictionary-based
[10]	Accidents, weather, road works, social events.	Dictionary-based
[11]	Accidents	Dictionary-based
ML		
[14]	Crime and disaster	NB
[15]	Useful, nearly-useful, irrelevant complaint reports	Rule-based classifier
[16]	Related to roads, related to weather	SVM
[17]	Accident, announcement, question, request, sentiment.	NB
[18]	Hazard and non-hazard	NB, KNN, and DLM
[19]	Congestion, crash, non-traffic event, external event.	SVM, NB, C4.5 Decision tree, PART, KNN,
[20]	Traffic and non-traffic	NB, SVM, DT
[21]	Roadwork, queue, accident, activities, breakdown, police.	Conditional Random Fields (CRFs) labeling method.
[22]	Traffic-relevant, traffic-irrelevant	NB, dictionary of frequently occurring words
[23]	Heavy-vehicle, traffic-jam, park-footpath autometer, wrong-side, breakdown, jump-signal, U-turn, no-parking.	Random forest classifier
[24]	Weather	Random forest classifier
[40]	Roadwork traffic jam, freight traffic, road closure, weather, accident	Dictionary, clustering algorithm
[41]	Earthquakes, forest fires, floods, and droughts	Checking a set of predefined weighed keywords, KNN algorithm
[25]	City-related events	CRF model, dictionary-based spotting
[26]	Traffic-congested and non-traffic-congested	NB
[27]	Accident and non-accident	SVM, NB, RF and Neural Networks
[29]	Traffic-relevant, traffic-irrelevant	LDA
[30]	Leisure, sports, music, movies, art, and other	LDA
[31]	Traffic, non-traffic	OLDA
[32]	Roadwork accidents, weather, special events, obstacle vehicles.	sLDA, SNB
[33]	Accidents, traffic jams, weather	LDA, SVM, NB, K-Nearest
Deep Learning		
[34]	Traffic relevant, Traffic-irrelevant	CNN
[35]	Traffic Control, Weather Anomaly, Accident, Congestion, Road Construction, Vehicle Anchorage, Official Occupation, and Normal Traffic.	BS-GCN
[36]	Traffic relevant, Traffic-irrelevant	ULMFiT model
[37]	Construction, weather, accidents, traffic jams, crowded events and others.	NB, RF, SVC, linear SVC, BiLSTM, and TextCNN
[38]	Traffic jams, smooth traffic	CNN+Word2Vec, CNN+FastText, and SVM
[39]	Education, transportation, environment, geospatial and water event	DNN and HHO
[42]	Congestion, accidents, social gatherings, thefts, bus queues, floods, and electricity charges.	AWD-LSTM and ULMFiT
[43]	Incident and non-incident	BERT

5) Other Methods: He et al. [44] proposed MetroScope system which analyzed real-time tweets to detect events related to the Washington D.C. Metro system. They developed a phrase-level algorithm that groups events with similar key phrases into a story. To prioritize urgent events, they performed sentiment analysis and then implemented a function to automatically send emails to authorities regarding emergency events.

Shekhar et al. [45] collected data about traffic conditions from Facebook and Twitter. They constructed a decision tree to display traffic-sensitive optimized routes. After that, they categorized particular streets within a city into three categories, which are Moderate Congestion, Severe Congestion, and No Congestion based on the average user's sentiment on hourly time slots. Further, they detected the possible causes of traffic congestion in a particular area and enabled users to search for the cause of congestion at a particular time. Ni et al. [2] developed a hashtag-based event detection algorithm. To detect events, they examined tweets within a specific area and probe (the subway station and two stadiums) instead of detecting the exact topic of the events.

Table I shows the type of detected events by each of the reviewed papers as well as the techniques that have been used for event detection. Table II illustrates the filtering approaches that have been used to filter out irrelevant posts before detecting the events.

TABLE II. FILTERING TECHNIQUES

Tweets Filtering Techniques		References
During Data Collection	By Keywords	[8], [16], [18], [19], [41], [22], [30],
	By accounts	[9], [10], [14], [15], [17]
	By Geo-coordinates	[2], [25], [42]
Using Machine Learning		[46], [32], [33], [40], [29], [20], [21], [34], [12], [13]

TABLE III. LOCATION INFORMATION EXTRACTION TECHNIQUES

Technique		References
Geo-Coordinates		[1], [29], [16], [47], [18], [48], [33], [49]
Text Analysis	Dictionary lookup	[50], [17], [1], [9], [12], [16]
	Sentence structure	[13], [29], [12], [45], [51], [24]
	Rule-based	[17], [13], [14]
	NER	[52], [53], [15], [40], [33], [27] [46], [54], [27]
Deep Learning	BERT	[37]

B. Spatio-temporal Information Extraction

Chaniotakis et al. [50] analyzed tweets about the flood and the evacuation in Oroville, California USA. They used the WordNet dictionary to create a corpus to detect discussions concerning the evacuation. Muhammad and Khodra [53] used the conditional random field (CRF) model for event information extracted from Indonesian tweets. They filtered the event-related tweets by combining the rule-based method with the bag of words model. Kumar *et al.* [18] extracted coordinates from geo-tagged tweets. Then, the geographic coordinates are mapped to a specific road or road segment**.** Shekhar *et al.* [45] extracted location names from the text. They assumed that the location is almost preceded by a preposition. So, they created a list of all the prepositions (e.g. in, at, on, etc.). The extracted location name is sent to Google Maps API to return the geographical coordinates. Wang et al. [29] extracted the coordinates from geo-tagged tweets and mapped them to extract location in terms of road, street, and landmark. On the other side, the location information from non-geotagged tweets is extracted either from users' profiles or from tweet content by using semantic analysis to identify the key joint words of "between...and", "from...to" and "exit to...,".

Furthermore, Wang *et al.* [1] extracted the streets, landmarks, and direction information from the text by using a gazetteer. Additionally, for traffic estimation and prediction, they extracted two types of road features, which are physical features (such as the road segment length, the number of lanes and the number of intersections) and point of interest (POI) (such as schools, hospitals, shopping mall, etc.). Sakaki et al. [16] extracted driving information from Japanese tweets and transformed geographically related terms into geographical coordinates. They created a dictionary for the place names in Japan. In addition, they collected pairs of verbs and prepositions, which are dependent on the names of places. Then, they used such pairs to extract the names of places.

Moreover, Daly et al. [9] perform a dictionary lookup to extract location names from SMS messages or tweets. Alifi and Supangkat [12] classified the tweets by using SVM to distinguish between the data that are related and not related to the traffic condition. They suggested methods for extracting location information which is using location vocabulary, using the symbol "-" or based on the structure and words in a sentence. In addition, they obtained useful information from real-time streams involving congestion causes, traffic conditions, as well as weather conditions. They suggested using three approaches: (i) the existence of location words (such as from, to, toward), (ii) the use of the symbol "-" that usually used to link two location points at once, (iii) the location vocabulary (such as street name). Klaithin and Haruechaiyasak [17] extract words or phrases related to traffic information from Thai tweets using lexicon-based and rule-based methods. They extracted traffic information such as road names, locations, and traffic accidents. Hanifah et al. [13] applied a rule-based approach and obtained information regarding the time and date, location and image. Additionally, they employed SVM model for traffic congestion detection from tweets posted in Bandung, Indonesia. The developed model filters the tweets into relevant to traffic congestion and irrelevant.

Dhavase and Bagade [14] proposed an approach to extract location from tweets. They used three different parsers: (i) named location parser (e.g., use gazetteer matching) to check for locations in tweets, (ii) NER parser (like Stanford NER). (iii) street building parser using rule-based pattern matching.

Hasby and Khodra [52] developed an information extraction module to extract time, location, condition, direction, and causes from Indonesian tweets about traffic jam.

The module consists of five elements, which are tokenizer, normalizer, Named Entity Recognition (NER), template element task, information filling and relation extraction. Similarly, Muhammad and Khodra [53] extracted event name, location, time and additional event information using an extractor module that was built up by Tokenizer, NER and POS Taggercomponent.

Moreover, Agarwal et al. [15] applied a combination of NER (Indico Text Analysis API) and GeoCoding (OpenStreetMap API) APIs to obtain the geographical entities from tweets. Gutierrez et al. [40] extracted locations from tweet messages using four NER engines, which are: Alchemy, Stanford NER, OpenCalais and NERD. Further, they used three geolocation external applications, which are: GeoNames, Google Geocoding, and Nominatim. Raymond [33] and Musaev et al. [54] applied the open-source Stanford NER tool to extract the place name. Salas et al. [46] used NER to link a concept to a unique location through a knowledge base such as Wikipedia.

Subsequently, Xu *et al.* [47] proposed a model based on crowdsourcing for describing urban emergency incidents such as storms, fires or traffic jams. They proposed a 5W model for illustrating the data, which provides five basic elements 1) When: temporal information (e.g. the starting/ending time of the event), 2) Where: spatial information (places), 3) What: the semantic information for the event, 4) Who: personal information (e.g., participatory or witness) and 5) Why: the reason information. They extracted location information from the check-in information from Weibo.

Besides, Berlingerio *et al.* [48] developed a system named SaferCity based on a new spatiotemporal clustering algorithm for incident detection from Twitter. Singh [55] extracted location from tweet text. To address the issue of the lack of location information, they suggested using the historical locations of the user to predict the probable location by applying Markov chain model. Yang et al. [51] extracted information form tweets related to the traffic conditions in Malaysia. They suggested extracting the location and direction information from the text using prepositions and words like "from…to", "along", "heading" and etc. Tejaswin et al. [24] extracted location entities using a regular expression parser. After that, entity disambiguation is applied to verify if it is a location and ensure that the address belongs to the correct city. Kim et al. [37] built an algorithm for region extraction to extract keywords from Korean tweets with the help of entity name recognition API based on BERT.

Table III summarizes the techniques for extracting location information. We grouped the approaches for extracting the location information into three main groups, which are: (i) from the coordinates attribute in the geotagged posts (ii) by extracting the location name from the text, and iii) using deep learning models such as BERT. The first approach is not always applicable since not all posts are geotagged because some users turn off location services on their smartphones to protect their privacy. For the second approach, the text will be analyzed using natural language processing (NLP) methods to extract the place name. The common methods that are applied to extract a placename from the text are as follows: i) Dictionary lookup: requires checking the text to discover place names listed in a gazetteer or glossary. ii) Sentence structure: use a list of prepositions (e.g., in, at, on, etc.) iii) Named Entity Recognition (NER): identify and categorize entities from text. iv) Rule-based pattern matching: implement the extraction based on certain written actions.

C. Sentiment Analysis

1) Lexicon-Based: Shekhar et al. in [45] and [41] categorized the users' emotions during a disaster by feeding English text from social media to sentiment analysis method. The users' emotions are sub-categorized as positive, negative, unhappy, depressed and angry. Additionally, they created a dictionary of weighted sentiment ratings for words and used SentiStrength online. SentiStrength [56] is a popular stand-alone online sentiment analysis tool. It uses a dictionary of sentiment words for assigning scores to negative and positive phrases in the text. Salas et al. [46] applied sentiment analysis to classify the tweets into positive, negative, or neutral class. They used TensiStrength for stress and relaxation strength detection.

2) Machine learning: A different sentiment classification method was applied by Kumar et al. [18] to categorize tweets into four sentiment classes: false negative, true negative, false positive and true positive. A true positive indicates that the tweet is accurately categorized as non-hazard whereas a true negative indicates that a tweet is accurately classified as a hazard. The false positive category refers to tweets that include some positive sentiment terms e.g. "awesome" and "enjoy", however, the actual sentiment is negative. On the other side, false negative refers to tweets that are incorrectly categorized as a hazard. They employed three ML algorithms, which are KNN, Naïve Bayes, and DLM.

Ohbe et al. [57] classify Japanese tweets about the local event into three categories: positive, negative, and other. They used a multinomial logistic regression analysis for the classifier. Berlingerio *et al.* [48] used Sentiment140 API for sentiment analysis. Sentiment140 [58] used a trained classifier built on large tweets with emoticons for distant supervised learning. Furthermore, Musaev et al. [54] developed a model to categorize tweets into happy or sad. They applied the Continuous Bag-of-Words and Skip-gram model. To do automatic labeling, they searched for tweets that contain ":-)" and ":-(" emoticons. After that, they utilized the Word2Vec repository to convert the tweets in the training set to their vector representations.

3) Hybrid: Adetiloye and Awasthi [59] applied sentiment analysis for traffic tweets using the lexicon of opinion words (LOWs) and the improved Naïve Bayes algorithms in [60]. Table IV summarizes the techniques for sentiment analysis that have been used in literature for event detection.

D. Big Data Tools and Platforms

The term big data refers to the extremely large amount of data that grows exponentially with time, which makes it difficult to conduct an efficient analysis using conventional IT and hardware solutions within a reasonable amount of time

[61]. Subsequently, the four main characteristics of big data are i) Volume: refers to the size of data that might be measured by Zettabytes (ZB), or Yottabytes (YB) ii)Velocity: which refers to the processing speed and considered as crucial characteristic for the performance. iii) Variety: refers to the diversity of the data, which can be structured, unstructured, or semi-structured. iv) Value: refers to valuable and reliable data.

TABLE IV. SENTIMENT ANALYSIS APPROACHES FOR EVENT DETECTION

Ref.	Lexicon	ML	Hybrid	Categories
[45], [41]	✓			Positive, Negative, Unhappy, Depressed, Angry
[18]		✓		True Negative, False Negative, True Positive, False Positive.
[46]	✓			Positive, Negative, Natural
[57]		✓		Positive, Negative, and other
[48]		✓		Positive, Negative, Natural
[59]			✓	Positive, Negative, Natural
[54]		✓		Sad, Happy

III. BIG DATA TECHNOLOGIES IN MOBILITY

Furthermore, the traditional data storage, processing, and analysis applications are insufficient to address the challenges that come from the massive continuously generated transportation and traffic-related data from various sources such as sensors, digital cameras, and social media. This raises the need for big data platforms and technologies, based on distributed data management and parallel processing. Big data storage solutions, such as NoSQL databases are ideal solutions for the storage issue since they have more flexible and adaptable data models and schemas compared to relational databases. Subsequently, big data platforms have integrated libraries including machine learning, deep learning, or data mining algorithms, which facilitate smart analysis [62].

The next subsection explains the existing approaches for traffic event detection using big data.

A. Traffic Events Detections Using Big Data Technologies

Nguyen and Jung [63] detected events by applying density-based spatial clustering. Additionally, to evaluate the proposed method, they used datasets (about 'FA Cup'' and ''Super Tuesday) employed in a previous study. They evaluated the performance using Hadoop. Khazaei *et al.* [64] proposed a big data analytics platform, named Sipresk that was built over Apache Spark to detect traffic events from different sources including social media, cameras, mobile devices, etc. Lau [33] suggested using topic model-based for text filtering then they built a classifier to identify traffic events such as traffic jams, road accidents, weather, etc. They fed the message corpus to the proposed Latent Dirichlet Allocation (LDA) model for topic learning. Subsequently, they applied the probabilistic language model to estimate the generation probabilities of the labeled message based on a mined unlabeled topic. Further, they implemented ML classifier using Spark Machine Learning (MLib) library.

Salas et al. [46] fetched real-time tweets through Kafka and Flume and stored them in HBase storage. They employed Spark machine learning library to build SVM classifier and filter the tweets into traffic or non-traffic-related tweets. The tweets are processed using Natural Language Processing (NLP) methods before they are passed to the trained classifier.

Suma *et al.* [49] used Apache Spark for spatio-temporal events detection in London City. For the data pool, they utilized the power of the Fujitsu Exabyte File System (FEFS). Further, they installed both FEFS and spark technologies on top of the HPC cluster. Pandhare and Medha [65] analyzed tweets related to traffic and accidents to detect road traffic events using Spark. They used a regular expression filter to separate unnecessary information. Then, they applied some text mining steps including tokenization, creating term frequency vectors by using HashingTF and TF-IDF to reflect the importance of a token in a document. After that, they classified the tweets by employing Logistic regression and SVM algorithms. Kousiouris *et al.* [66] identified large crowd concentration events that might affect the user journey. They used Spark, Apache AVRO and Cloud-based solutions (OpenStack Swift).

Moreover, Alomari and Mehmood [10] developed a lexicon-based approach to filter Arabic tweets about traffic congestion in Jeddah city using SAP HANA. Then, they extended their word and performed sentiment analysis [67]. Subsequently, they developed multiple big data pipelines and architectures for social text event detection using cutting-edge technologies consisting of machine learning algorithms and high-performance computing as well as Apache Spark. Furthermore, they proposed supervised [68], [7] and unsupervised [69] machine learning methods to enable smarter transportation by detecting events using social data in the Arabic language. Subsequently, they detected several events including congestion, roadwork, fire, social events, weather, government measures, and public concern. Additionally, to improve the performance of detecting events from Saudi dialectical Arabic text, they proposed a pre-processing pipeline that includes a tokenizer, irrelevant characters removal, normalizer, stop words removal, as well as an Arabic light stemmer. Also, they built a tool for spatio-temporal clustering and visualization. Furthermore, they proposed methods for validating the detected events through internal sources (Twitter data) and external sources (e.g. official newspaper websites). Moreover, to address the challenges that come from manual labeling of large datasets, they proposed an automatic labeling method [70] using predefined dictionaries for detecting events.

Chen et al. [71] proposed a semi-supervised deep-learning model for detecting traffic events. They built a multi-model feature learning architecture to transform data from sensor time series and Twitter posts into a unified multi-modal feature representation. They built two encoders: the first one to extract features from sensor data using the Recurrent Neural Network (RNN) while the second encoder is designed for social data.

IV. DISCUSSION, CHALLENGES AND FUTURE DIRECTIONS

We illustrated in Section II the general workflow for event detection. In this section, we discuss the challenges for each step in the workflow.

Firstly, for data collection, the existing API such as Twitter API[1] limits the number of collected data for free. Additionally, even though they have paid APIs with fewer restrictions on the number of fetched tweets, it is too expensive. Secondly, for data pre-processing, although there are several tools and packages, most of them are built for the English language. Thus, there is a need for more efficient tools for other languages. Subsequently, there is a need to improve pre-processing and NLP methods to work on the dialectical short text. Thirdly, posts on social media are usually short and thus may not include all the important information about the events. Therefore, finding the exact location or time of occurrence might be difficult. The information might either not exist because users disable the location service for privacy reasons, or it exists but does not reflect the time or the place where the event occurred since people can post about events in other cities and countries. Additionally, in some cases, the post carries more than one location or time information. For instance, the information attached to the posts such as the geo-coordinates and the information users wrote in the text. Thus, the existing approaches for Spatio-temporal information extraction need improvement to consider the different scenarios and extract or predict the right information. Subsequently, recent approaches need to be used including deep learning and Large Language Model (LLM) such as the work in [37].

Moreover, several approaches have been used for traffic event detection including the lexicon-based approach, ML algorithms and deep learning. We divided the works that used ML into four categories. The first category includes works that used the common supervised or unsupervised algorithms such as SVM, Naïve based, clustering, etc. The second category includes the works that used topic modeling algorithms such as LDA. The third category contains the works that applied deep learning. The last category includes the works that used other methods such as hashtags-based techniques. Each approach has it is own challenges. One of the major challenges of using supervised classification is labeling the data for training the model. Manual labeling takes significant time and effort, especially for big data.

Furthermore, supervised classification requires defining the classes and then building and training the models, which means that we need to specify the types of events that we want to detect in advance. Thus, supervised classification is not appropriate if we do not want to limit the detected event types. The other challenge in supervised classification is having an imbalanced dataset where the number of posts in the training dataset for each class label is not balanced. On the other side, topic modeling has its challenges. One of the challenges is understanding the topic and finding the category that belongs to it. Additionally, we need to test different parameters to find the best number of topics and iterations, which is a difficult process and takes a long running time, especially for big data (for more details see [69]). Furthermore, the works in literature that used deep learning for traffic event detection are very limited. Additionally, more work is required to study the feelings and emotions regarding the detected events, which could help the authorities and decision-makers understand the situation and get more involved in addressing the difficulties.

Finally, detecting events from big social data is difficult due to its daunting characteristics -- volume, variety, velocity and veracity. The state-of-the-art on using big data technology for traffic event detection from social media is limited. Therefore, many more works are needed to improve the breadth and depth of the studies in this area regarding the size and diversity of the data, as well as the applicability, accuracy, performance, and scalability of the analysis and detection methods.

V. CONCLUSION

This paper comprehensively reviews recent advancements in the fusion of AI with a specific focus on machine learning, and deep learning techniques in conjunction with big data technologies for traffic event detection from social media. Furthermore, we presented the general workflow, which includes the following steps: Data collection and filtering, Pre-processing, Event detection, Spatio-temporal information extraction, and Evaluation. Before detecting events, data are filtered to keep only traffic-related posts. We found that this process is done either during the data collection phase or after through machine learning algorithms. After that, we divided the techniques for event detection into categories and then discussed the works based on the applied technique. The first technique for event detection is lexicon-based. The second technique is using machine learning including, supervised and unsupervised algorithms, topic modeling, or hybrid. The third technique is using deep learning. The last category includes other techniques such as hashtags-based techniques.

Moreover, we grouped the techniques for extracting location information into two main groups, which are using geo-coordinates attributes in the geotagged posts and extracting location names by text analysis. The first approach is not always applicable since not all posts are geotagged. For the second approach, several methods are applied to extract location from the text including using NER, Dictionary lookup, Rule-based pattern matching, deep learning models and sentence structure by using a list of prepositions.

Furthermore, we classified the existing approaches for sentiment analysis for traffic-related events into three categories, which are lexicon-based, using machine learning, and hybrid approaches. Additionally, we reviewed the works that used big data technology for traffic event detection from social media.

However, the state of the art in this area that uses deep learning, LLM or big data technologies is limited, and thus many more works are needed to improve the breadth and depth of the studies since using them helps to improve the efficiency, scalability, performance, flexibility as well as support multilingual. Subsequently, big data technologies and platforms are very important in this domain due to the characteristics -- volume, variety, velocity, and veracity of the social data. In conclusion, this review consolidates the current state of research, offering a valuable resource for researchers, practitioners, and policymakers seeking to leverage cutting-

[1] https://developer.twitter.com/en/docs/twitter-api

edge technologies for enhancing urban mobility and smart cities.

REFERENCES

[1] S. Wang, L. He, L. Stenneth, P. S. Yu, and Z. Li, "Citywide traffic congestion estimation with social media," in Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15, 2015, pp. 1–10, doi: 10.1145/2820783.2820829.

[2] M. Ni, Q. He, and J. Gao, "Forecasting the Subway Passenger Flow under Event Occurrences with Social Media," IEEE Trans. Intell. Transp. Syst., vol. 18, no. 6, pp. 1623–1632, 2017, doi: 10.1109/TITS.2016.2611644.

[3] F. Wu, H. Wang, and Z. Li, "Interpreting traffic dynamics using ubiquitous urban data," Proc. 24th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. - GIS '16, pp. 1–4, 2016, doi: 10.1145/2996913.2996962.

[4] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," Acm Sigspatial, pp. 334–343, 2013, doi: 10.1145/2525314.2525343.

[5] D. R. Domínguez, R. P. Díaz Redondo, A. F. Vilas, and M. Ben Khalifa, "Sensing the city with Instagram: Clustering geolocated data for outlier detection," Expert Syst. Appl., vol. 78, pp. 319–333, 2017, doi: 10.1016/j.eswa.2017.02.018.

[6] S. Xu, S. Li, and R. Wen, "Sensing and detecting traffic events using geosocial media data: A review," Comput. Environ. Urban Syst., no. June, 2018, doi: 10.1016/j.compenvurbsys.2018.06.006.

[7] E. Alomari, I. Katib, and R. Mehmood, "Iktishaf: A Big Data Road-Traffic Event Detection Tool Using Twitter and Spark Machine Learning," Mob. Networks Appl., pp. 1–16, 2020.

[8] A. F. Istiq Septiana, Setiowati, Yuliana, Arna Fariza, "Road Condition Monitoring Application Based on Social Media With Text Mining System," Int. Electron. Symp. Road, pp. 148–153, 2016.

[9] E. M. Daly, F. Lecue, and V. Bicer, "Westland row why so slow?Fusing Social Media and Linked Data Sources for Understanding Real-Time Traffic Conditions," Proc. 2013 Int. Conf. Intell. user interfaces - IUI '13, no. March, p. 203, 2013, doi: 10.1145/2449396.2449423.

[10] E. Alomari and R. Mehmood, "Analysis of Tweets in Arabic Language for Detection of Road Traffic Conditions," in in Proceedings of the First EAI Conference on Smart Societies, Infrastructure, Technologies and Applications (SCITA 2017), Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering (LNICST), pp. 98–110.

[11] B. Alkouz and Z. Al Aghbari, "SNSJam: Road traffic analysis and prediction by fusing data from multiple social networks," Inf. Process. Manag., vol. 57, no. 1, p. 102139, 2020, doi: 10.1016/j.ipm.2019.102139.

[12] M. R. Alifi and S. H. Supangkat, "Information Extraction for Traffic Congestion in Social Network," in International Conference on ICT For Smart Society, 2016, no. July, pp. 20–21.

[13] R. Hanifah, S. H. Supangkat, and A. Purwarianti, "Twitter information extraction for smart city," Proc. - 2014 Int. Conf. ICT Smart Soc. "Smart Syst. Platf. Dev. City Soc. GoeSmart 2014", ICISS 2014, pp. 295–299, 2014, doi: 10.1109/ICTSS.2014.7013190.

[14] N. Dhavase and A. M. Bagade, "Location identification for crime & disaster events by geoparsing Twitter," 2014 Int. Conf. Converg. Technol. I2CT 2014, pp. 2–4, 2014, doi: 10.1109/I2CT.2014.7092336.

[15] S. Agarwal, N. Mittal, and A. Sureka, "Potholes and Bad Road Conditions- Mining Twitter to Extract Information on Killer Roads," ACM India Jt. Int. Conf. Data Sci. Manag. Data CoDS-COMAD 2018, 2018.

[16] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa, "Real-time event extraction for driving information from social sensors," in Proceedings - 2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems, CYBER 2012, 2012, pp. 221–226, doi: 10.1109/CYBER.2012.6392557.

[17] S. Klaithin and C. Haruechaiyasak, "Traffic Information Extraction and Classification from Thai Twitter," Comput. Sci. Softw. Eng. (JCSSE), 2016 13th Int. Jt. Conf., pp. 1–6, 2016, doi: 10.1109/JCSSE.2016.7748851.

[18] A. Kumar, M. Jiang, and Y. Fang, "Where not to go?: detecting road hazards using twitter," in Proceedings of the 37th international ACM …, 2014, vol. 2609550, pp. 1223–1226, doi: 10.1145/2600428.2609550.

[19] E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni, "Real-Time Detection of Traffic from Twitter Stream Analysis," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 4, pp. 2269–2283, 2015, doi: 10.1109/TITS.2015.2404431.

[20] D. A. Kurniawan, S. Wibirama, and N. A. Setiawan, "Real-time Traffic Classification with Twitter Data Mining," in In 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), 2016, pp. 1–5, doi: 10.1109/ICITEED.2016.7863251.

[21] and F. C. Hoang Nguyen, Wei Liu, Paul Rivera, "TrafficWatch: Real-Time Traffic Incident Detection and Monitoring Using Social Media," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9651, pp. 540–551, 2016, doi: 10.1007/978-3-319-31753-3.

[22] M. A. Yazici, S. Mudigonda, and C. Kamga, "Incident Detection through Twitter Organization vs. Personal Accounts," no. 212, pp. 1–17, 2017.

[23] D. Semwal, S. Patil, S. Galhotra, A. Arora, and N. Unny, "STAR: Real-time Spatio-Temporal Analysis and Prediction of Traffic Insights using Social Media," in In Proceedings of the 2nd IKDD Conference on Data Sciences, 2015, p. 7, doi: 10.1145/2778865.2778872.

[24] P. Tejaswin, R. Kumar, and S. Gupta, "Tweeting Traffic: Analyzing Twitter for generating real-time city traffic insights and predictions," in Proceedings of the 2nd IKDD Conference on Data Sciences - CODS-IKDD '15, 2015, pp. 1–4, doi: 10.1145/2778865.2778874.

[25] P. Anantharam, P. Barnaghi, K. Thirunarayan, and A. Sheth, "Extracting City Traffic Events from Social Streams," ACM Trans. Intell. Syst. Technol., vol. 6, no. 4, pp. 1–27, 2015, doi: 10.1145/2717317.

[26] M. R. Langa and M. N. Moeti, "A Real-Time Notification System for Traffic Congestion on South African National Routes," pp. 79–91, 2022.

[27] N. Suat-rojas, C. Gutierrez-osorio, and C. Pedraza, "Extraction and Analysis of Social Networks Data to Detect Traffic Accidents," 2022.

[28] D. M. Blei, B. B. Edu, A. Y. Ng, A. S. Edu, M. I. Jordan, and J. B. Edu, "Latent Dirichlet Allocation," J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003, doi: 10.1162/jmlr.2003.3.4-5.993.

[29] D. Wang, A. Al-Rubaie, S. S. Clarke, and J. Davies, "Real-Time Traffic Event Detection From Social Media," ACM Trans. Internet Technol, vol. 18, no. 23, pp. 1–23, 2017, doi: 10.1145/3122982.

[30] W. Huang, S. Xu, Y. Yan, and A. Zipf, "An exploration of the interaction between urban human activities and daily traffic conditions: A case study of Toronto, Canada," Cities, no. July, 2018, doi: 10.1016/j.cities.2018.07.001.

[31] 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," Accid. Anal. Prev., vol. 151, no. December 2020, p. 105973, 2021, doi: 10.1016/j.aap.2021.105973.

[32] Y. Gu, Z. (Sean) Qian, and F. Chen, "From Twitter to detector: Real-time traffic incident detection using social media data," Transp. Res. Part C Emerg. Technol., vol. 67, pp. 321–342, 2016, doi: 10.1016/j.trc.2016.02.011.

[33] R. Y. K. Lau, "Toward a social sensor based framework for intelligent transportation," in 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2017, pp. 1–6, doi: 10.1109/WoWMoM.2017.7974354.

[34] Y. Chen, Y. Lv, X. Wang, and F.-Y. Wang, "A convolutional neural network for traffic information sensing from social media text," 2017 IEEE 20th Int. Conf. Intell. Transp. Syst., pp. 1–6, 2017, doi: 10.1109/ITSC.2017.8317650.

[35] Y. Ji, J. Wang, Y. Niu, and H. Ma, "Reliable Event Detection via Multiple Edge Computing on Streaming Traffic Social Data," IEEE Access, pp. 1–14, 2021, doi: 10.1109/ACCESS.2021.3060624.

[36] P. Ambastha and M. S. Desarkar, "Incident Detection from Social Media Targeting Indian Traffic Scenario Using Transfer Learning," 2020 IEEE 23rd Int. Conf. Intell. Transp. Syst. ITSC 2020, 2020, doi: 10.1109/ITSC45102.2020.9294295.

[37] Y. Kim et al., "Regional Traffic Event Detection Using Data Crowdsourcing," Appl. Sci., vol. 13, no. 16, 2023, doi: 10.3390/app13169422.

[38] R. R. Almassar and A. S. Girsang, "Detection of traffic congestion based on twitter using convolutional neural network model," IAES Int. J. Artif. Intell., vol. 11, no. 4, pp. 1448–1459, 2022, doi: 10.11591/ijai.v11.i4.pp1448-1459.

[39] K. Swapnika and D. Vasumathi, "a Hybrid Dnn-Hho Approach for Event Detection in Big Data," Indian J. Comput. Sci. Eng., vol. 13, no. 5, pp. 1401–1411, 2022, doi: 10.21817/indjcse/2022/v13i5/221305086.

[40] C. Gutierrez, P. Figuerias, P. Oliveira, R. Costa, and R. Jardim-Goncalves, "Twitter mining for traffic events detection," Proc. 2015 Sci. Inf. Conf. SAI 2015, pp. 371–378, 2015, doi: 10.1109/SAI.2015.7237170.

[41] H. Shekhar and S. Setty, "Disaster analysis through tweets," 2015 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2015, no. August, pp. 1719–1723, 2015, doi: 10.1109/ICACCI.2015.7275861.

[42] A. Hodorog, I. Petri, and Y. Rezgui, "Machine learning and Natural Language Processing of social media data for event detection in smart cities," Sustain. Cities Soc., vol. 85, no. June, p. 104026, 2022, doi: 10.1016/j.scs.2022.104026.

[43] B. Mehri, M. Trépanier, and Y. Goussard, "Multilingual Text Classification on Social Media Data for Incident Alert in Subway Multilingual Text Classification on Social Media Data for Incident Alert in Subway Transportation Network," no. January, 2023.

[44] J. He et al., "MetroScope: An Advanced System for Real-Time Detection and Analysis of Metro-Related Threats and Events via Twitter," SIGIR 2023 - Proc. 46th Int. ACM SIGIR Conf. Res. Dev. Inf. Retr., pp. 3130–3134, 2023, doi: 10.1145/3539618.3591807.

[45] H. Shekhar, S. Setty, and U. Mudenagudi, "Vehicular traffic analysis from social media data," 2016 Int. Conf. Adv. Comput. Commun. Informatics, pp. 1628–1634, 2016, doi: 10.1109/ICACCI.2016.7732281.

[46] I. Salas, A., Georgakis, P., Nwagboso, C., Ammari, A. and Petalas, "Traffic Event Detection Framework Using Social Media," in IEEE International Conference on Smart Grid and Smart Cities, 2017, no. July, pp. 303–307, doi: 10.1109/ICSGSC.2017.8038595.

[47] Z. Xu et al., "Crowdsourcing based Description of Urban Emergency Events using Social Media Big Data," IEEE Trans. Cloud Comput., pp. 1–1, 2016, doi: 10.1109/TCC.2016.2517638.

[48] M. Berlingerio, F. Calabrese, G. Di Lorenzo, X. Dong, Y. Gkoufas, and D. Mavroeidis, "SaferCity: A system for detecting and analyzing incidents from social media," Proc. - IEEE 13th Int. Conf. Data Min. Work. ICDMW 2013, pp. 1077–1080, 2013, doi: 10.1109/ICDMW.2013.39.

[49] S. Suma, R. Mehmood, N. Albugami, I. Katib, and A. Albeshri, "Enabling Next Generation Logistics and Planning for Smarter Societies," Procedia - Procedia Comput. Sci., pp. 1–6, 2017.

[50] E. Chaniotakis, C. Antoniou, and ..., "Enhancing resilience to disasters using social media," … Technol. …, pp. 699–703, 2017, doi: 10.1109/MTITS.2017.8005602.

[51] L. C. Yang, B. Selvaretnam, P. K. Hoong, I. K. T. Tan, E. K. Howg, and L. H. Kar, "Exploration of road traffic tweets for congestion monitoring," J. Telecommun. Electron. Comput. Eng., vol. 8, no. 2, pp. 141–145, 2016, [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84984848398&partnerID=40&md5=aa3409237b2ff2788facabd0f6edd723.

[52] M. Hasby and M. L. Khodra, "Optimal Path Finding based on Traffic Information Extraction from Twitter," Int. Conf. ICT Smart Soc., pp. 1–5, 2013, doi: 10.1109/ICTSS.2013.6588076.

[53] F. Muhammad and M. L. Khodra, "Event information extraction from Indonesian tweets using conditional random field," ICAICTA 2015 - 2015 Int. Conf. Adv. Informatics Concepts, Theory Appl., pp. 0–5, 2015, doi: 10.1109/ICAICTA.2015.7335383.

[54] A. M. B, Z. Jiang, S. Jones, and P. Sheinidashtegol, Detection of Damage and Failure Events of Road Infrastructure Using Social Media, vol. 10966. Springer International Publishing, 2018.

[55] J. P. Singh, Y. K. Dwivedi, N. P. Rana, A. Kumar, and K. K. Kapoor, "Event classification and location prediction from tweets during disasters," Ann. Oper. Res., pp. 1–21, 2017, doi: 10.1007/s10479-017-2522-3.

[56] M. Thelwall, K. Buckley, G. Paltoglou, and D. Cai, "Sentiment Strength Detection in Short Informal Text," Am. Soc. Informational Sci. Technol., vol. 61, no. 12, pp. 2544–2558, 2010, doi: 10.1002/asi.

[57] T. Ohbe, T. Ozono, and T. Shintani, "Developing a sentiment polarity visualization system for local event information analysis," Proc. - 2016 5th IIAI Int. Congr. Adv. Appl. Informatics, IIAI-AAI 2016, pp. 19–24, 2016, doi: 10.1109/IIAI-AAI.2016.118.

[58] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification using Distant Supervision," Processing, vol. 150, no. 12, pp. 1–6, 2009, doi: 10.1016/j.sedgeo.2006.07.004.

[59] T. Adetiloye and A. Awasthi, Traffic Condition Monitoring Using Social Media Analytics, vol. 44. Springer Singapore, 2018.

[60] H. Kang, S. J. Yoo, and D. Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews," Expert Syst. Appl., vol. 39, no. 5, pp. 6000–6010, Apr. 2012, doi: 10.1016/J.ESWA.2011.11.107.

[61] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," Mob. Networks Appl., vol. 19, no. 2, pp. 171–209, 2014, doi: 10.1007/s11036-013-0489-0.

[62] R. M. Ebtesam Alomari, Iyad Katib, "Big Data Technologies for Arabic Social Media Analysis to enable Smarter Transportation," King AbdualAziz University, 2021.

[63] D. T. Nguyen and J. E. Jung, "Real-time event detection for online behavioral analysis of big social data," Futur. Gener. Comput. Syst., vol. 66, pp. 137–145, 2017, doi: 10.1016/j.future.2016.04.012.

[64] H. Khazaei, R. Veleda, M. Litoiu, and A. Tizghadam, "Realtime big data analytics for event detection in highways," 2016 IEEE 3rd World Forum Internet Things, WF-IoT 2016, pp. 472–477, 2017, doi: 10.1109/WF-IoT.2016.7845461.

[65] K. R. Pandhare and M. A. Shah, "Real time road traffic event detection using Twitter and spark," 2017 Int. Conf. Inven. Commun. Comput. Technol., no. Icicct, pp. 445–449, 2017, doi: 10.1109/ICICCT.2017.7975237.

[66] G. Kousiouris et al., "An integrated information lifecycle management framework for exploiting social network data to identify dynamic large crowd concentration events in smart cities applications," Futur. Gener. Comput. Syst., vol. 78, pp. 516–530, 2018, doi: 10.1016/j.future.2017.07.026.

[67] E. Alomari, R. Mehmood, and I. Katib, "Sentiment Analysis of Arabic Tweets for Road Traffic Congestion and Event Detection," in In: Mehmood R., See S., Katib I., Chlamtac I. (eds) Smart Infrastructure and Applications: Foundations for Smarter Cities and Societies, Springer (https://www.springer.com/us/book/9783030137045), 2020, pp. 37–54.

[68] E. Alomari, R. Mehmood, and I. Katib, "Road Traffic Event Detection Using Twitter Data, Machine Learning, and Apache Spark," in 2019 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Aug. 2019, pp. 1888–1895, doi: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00332.

[69] E. Alomari, I. Katib, A. Albeshri, and R. Mehmood, "COVID-19: Detecting Government Pandemic Measures and Public Concerns from Twitter Arabic Data Using Distributed Machine Learning," Int. J. Environ. Res. Public Health, vol. 18, no. 1, p. 282, Jan. 2021, doi: 10.3390/ijerph18010282.

[70] E. Alomari, I. Katib, A. Albeshri, T. Yigitcanlar, and R. Mehmood, "Iktishaf+: A big data tool with automatic labeling for road traffic social sensing and event detection using distributed machine learning," Sensors, vol. 21, no. 9, pp. 1–33, 2021, doi: 10.3390/s21092993.

[71] Q. Chen, W. Wang, K. Huang, S. De, and F. Coenen, "Multi-modal generative adversarial networks for traffic event detection in smart cities," Expert Syst. Appl., vol. 177, no. March, p. 114939, 2021, doi: 10.1016/j.eswa.2021.114939.