Materials Today: Proceedings 47 (2021) 5135–5141

Contents lists available at ScienceDirect

Materials Today: Proceedings

journal homepage: www.elsevier.com/locate/matpr

Road traffic accidents: An overview of data sources, analysis techniques and contributing factors

Arun Chand *, S. Jayesh, A.B. Bhasi

Department of Mechanical Engineering, School of Engineering, Cochin University of Science and Technology, Kerala, India

ARTICLE INFO

Article history:
Received 17 April 2021
Received in revised form 15 May 2021
Accepted 18 May 2021
Available online 2 June 2021

Keywords:
Road Safety
Road Traffic Accidents
Statistical Data Analysis
Driver distraction

ABSTRACT

Road traffic accidents are one among the world's leading causes of injuries and fatalities and hence represent an important field of research towards the use of traffic accident analysis and prediction techniques and the determination of the most key factors contributing to road traffic accidents. This paper aims to provide an overview of road accident data sources, data analysis techniques, various algorithms used to build road accident forecasts, and also their suitability to the types of data being examined with the ease of interpretation. The paper also summarizes the operational problems of road traffic, identifies the risk factors, the efficacy of road safety measures when they contribute to the statistical analysis of the severity of motor vehicle accidents and offers an assessment of future methodological approaches. In this review, different gaps in the road traffic accident area were found and further fields of research have been mentioned.

Selection and peer-review under responsibility of the International Conference on Sustainable materials, Manufacturing and Renewable Technologies 2021.

1. Introduction

As a result of Road Traffic Accidents around the world, the lives of about 1.35 million people are cut short annually. About 20 to 50 million people experience nonfatal injuries due to these accidents and many are disabled permanently. Road traffic accidents are causing significant economic impacts on victims, thereby to the whole nation, by costing around 3% loss of the gross domestic product [1]. Thus, traffic accident has emerged as a topic of discussion and analyzing traffic accident data, becoming a major concern for researchers in search of coherent methods for road accidents forecasting. The main aim of accident data analysis is to identify the factors affecting road traffic accident occurrences, thus mitigate the main issues in the area of road safety. The effectiveness of accident prevention methods depends mostly on the genuineness of the gathered and estimated data and the suitability of the analysis methods [2].

The selection of the right data analysis method will help in uncovering the reasons for accident occurrence at a particular zone or location of study and to predict with reasonable precision, the probability of accident occurrence per day or the relative safety of various categories of road users in that region [4]. Thus the quality of research depends upon the selection of the appropriate statistical methods .

The statistical methods which are used to conduct the research study will give valid and reliable results if the researcher knows the basic concepts of it. Various statistical methods will have their unique assumptions when analysing the accident data in different studies. So to decide the appropriate method, their assumptions should be taken into account. Also, the selection of wrong statistical methods may lead to invalid conclusions of road accident analysis. This indicates for getting the proper insight into the traffic problems, interpretations of accident data analysis results are crucial. Thus sufficient understanding of statistical tools/methods is also very significant in creating quality research in road accident data analysis [5]. Sometimes it may be difficult for the researchers to find the appropriate statistical methods to be implemented while considering the given set of accident data as well as the objectives of the study.

This paper summarizes different road accident data sources from where the researchers could get sufficient data, various methods for the analysis of these data and identifies the uniqueness in each of them, when dealing with the road accident data, to answer the questions related to different analyses problems. In the final section of the paper, recommendations for the management of

* Corresponding author.
E-mail address: arunchand@cusat.ac.in (A. Chand).

https://doi.org/10.1016/j.matpr.2021.05.415
2214-7853/
Selection and peer-review under responsibility of the International Conference on Sustainable materials, Manufacturing and Renewable Technologies 2021.

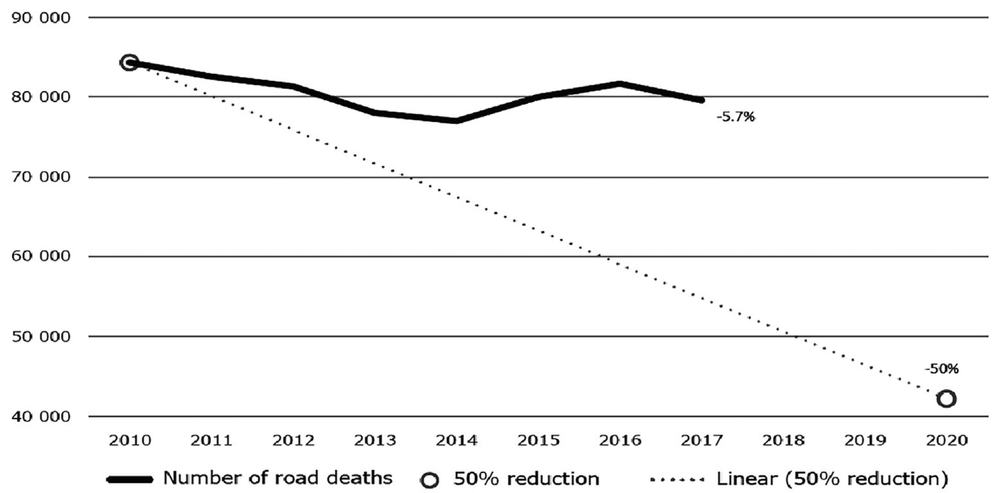

Fig. 1. Road traffic deaths over the past 10 years: a worldwide trend. (Source: Road Safety Annual Report 2019 [3].)

factors responsible for road accidents and future research are provided.

2. Road accident data sources

There are several road accident data sources and an essential one is the data from the Government agencies. Here the data indexes are produced, gathered and made accessible to the public by government authorities. The police department, traffic department and traffic policy-making agencies include the Government agencies. The data collection is done through First Information Reports of Police dept., General Insurance company databases and Hospital records etc. The main equipments used for data collection are vehicle speed detection cameras, traffic police speed radars and signal breaker cameras etc. These data are utilized for making public strategy in every nation concerning road security plans and traffic infrastructure.

Another source is the database that can be openly accessible; most of them are generated by funding from the general public, which is accessible to the public without any restriction. Openly accessible data is accessible to the public is utilizing exclusive websites that empower search and uncover interfaces, including web administrations. This permits programmed utilization of the information and incorporate it to other sites or applications. The United States government open a data catalogue named 'catalog.data.gov/dataset' for traffic accidents [6] that includes data from all the Countries. United Kingdom open data list– 'data.gov.uk' and Australia open data catalogue-'data.gov.au' [7], that incorporates in-progress travel, vehicle accidents and public transport timetables.

Different electronic types of equipment are used to collect accident data from the road which includes technologies that are part of the road infrastructure, such as speed sensing cameras, vehicle radars, loop detectors fixed on the road itself and other video surveillance had been used in many studies by taking their availability and accessibility into consideration [8]. For collecting vehicle data from road intersections, Adaptive Signal Control datasets and Bluetooth detectors could also be used [9]. The video surveillance technologies are reliable and through which we can gather additional data, such as make and model of vehicle and vehicle utilization purpose. Video surveillance and Radars got some advantages like reliability, less cost, accuracy, and ease of use in collecting the data [10].

The Onboard equipment (OBE) installed on vehicles also can be used to collect the data related to vehicle conditions and driver distractions. These OBEs includes cameras set up to record road and vehicle conditions, global positioning systems, surveillance cameras for identifying, driver drowsiness or distractions [11]. They are also used for analysing the vehicle conditions, such as a change in speed, sudden braking, lane deviations etc, and direction and acceleration of the vehicle at the time of an accident [12].

Intelligent transportation systems (ITS) nowadays became the most significant data sources of big data related to traffic accident analysis and prediction. Huge data is collected through GPS information through On-Board electronic equipment and mobile phones [13]. ITS equipment, for example, microwave vehicle detection systems (MVDS) fitted on roadsides, constantly gather a large amount of data like Speed of the vehicle, volume of traffic, utilisation and type of vehicle from every area [14]. Vehicle License Plate Recognition (VLPR) system, is another type of ITS installation that can collect more than 8 million records each day [15]. Modelling of road accident risks and higher-level simulations can be anticipated using these big data statistics [16].

Social Media has emerged as the most recent source in road traffic accident data collection and related researches, and a large amount of data is utilized for the same purpose from Twitter, Google, Bing map and Inrix [17]. Because of the non-dependability of its source or distribution and difficulty in inferring, social media is considered untrustworthy because the users utilize local language to post their comments and the content may include spelling and linguistic errors [18]. In the case of road accidents, social media data could be biased since all the relevant information about accidents may not be reported by social media users [19].

For getting the superior perspective of the traffic accidents, many researchers had utilized at least two of the previously mentioned information sources, seeking to upgrade the precision of the outcomes gained from their analyses [20]. The combination of different data sources comprises the consolidation of heterogeneous data from road traffic accidents, traffic volume data, and climate condition, road facilities related to geometry and visibility and efficient use of land in connection with the region of analysis, which help in enhancing the accuracy of results of the traffic accidents analyzing using big data [21].

3. Accident data analysis techniques

To describe the information and variables of road traffic accidents, different analytic methods are utilized. Researchers analyses the data to find out hidden patterns and inferences thereby generate new rules. These processes help to create the following:

- To build and check classification rules associated with traffic accident data.
- To select preeminent factors causing the accidents thereby precise modelling of road accidents.
- To frame the driver's code and conduct on the road.
- To choose significant features to be utilized to train deep learning algorithms and artificial neural networks.
- To differentiate between safe and unsafe areas for driving.

Dividing and attaching a set of entities into groups or clusters is called Cluster analysis. Thus the entities grouped in each cluster will have common features and they seem different from other entities grouped in other clusters. Normally, cluster analysis has been used to categorize data into structures, which are more clearly understood and manipulated [22]. According to the features on which clustering methods are applied, common features can be clarified as the level of interaction of entities. When compared to classification methods, clustering doesn't need any data that is previously assigned with any specific class for discriminating various groups in the data. The absence of these previous data reveals that the prime goal of clustering is to find out the fundamental structure in the accident data, instead of distinguishing future data into classes or categories. And thus, clustering brings off a more compact description of the available data [23]. The advantages of clustering algorithms are that they do not need prior data processing, work well with large data sets, and the results are graphically interpretable. Clustering Algorithms apply a distance function to analyse the resemblance in features and are sensitive to the possibility of detecting a local maximum instead of a global maximum on their optimization functions. These algorithms work with continuous elements and a measure of similarity for data with qualitative elements [24].

Association rule mining helps in fetching correlations among dissimilar features in a defined set of data. Rules that establish the correlation between separate groups of features in the data set can be derived using the association rule mining method. The rate of occurrence and consistency of a rule and support value are the factors that define the quality of the rules [25]. R. Bhavsar et al. applied a simplified linear modelling method for the analysis of traffic accidents on the National Highways of India. To improve road safety, they proposed a new perspective to comprise average spot speed and average daily traffic in the road accident mitigation model for rural highways. Since, the Statistical Models couldn't fully predict the features of all section, because of the heterogeneous character of traffic accidents, the association rule mining method was used by the researchers to identify accident spots and the findings from this model reflected a better estimate of road accidents [26].

In the Decision Tree method, the classification models are built in the form of trees and every leaf node symbolizes one of the key variables. The number of branches in each leaf node equals the number of probable values of the supposed key variable. Then, selected feature values are assigned to each node. Based on the value of the key variables starting from the root node to the leaf the decisions are made. The Decision Tree method does not need domain knowledge for its construction and thus it is widely used in pattern classification applications [27]. Data sets with partial information can also be processed using this tool. Decision Trees are a useful tool with high dimensional data and the mode of analysis is exploratory, not conclusive.

Traffic accident incidents reported from social media can be processed using Natural language processing (NLP) algorithms and deduce data such as geographic location, accident characteristics and other significant factors. To conduct categorization of social media information according to pre-set focus groups, as a binary classification, NLP algorithms are used, being the social media information related or not to road accidents [28]. When evaluated against other various classification models, including K-nearest neighbor (KNN), Naive Bayes classifier, PART classifier and C4.5 decision tree, the accuracy value was 0.9575, which is a point of reference in this field of knowledge [29].

4. Road accident forecasting through data analysis

Machine Learning (ML) methods had been utilized mostly for road traffic accident forecasting, depending upon their predictive capacities, the capability of machine learning methods to handle multi-dimensional data, flexible execution with coding. A predefined set of conditions or features are used to predict the circumstances of an accident by the researchers to carry out the analysis and represent traffic accident fatality models in the area of road accident forecasting [30].

Genetic Algorithms (GA) are search heuristics, used to solve optimization problems related to road traffic in the area of accident forecasting; both constrained and unconstrained, and are based on natural evolution [31]. At each step Genetic Algorithms repeatedly mutate a population of existing individual solutions, selecting randomly from the population to produce the next generation children. The finest individuals are constantly selected and managed on by crossover and mutation. Thus the population advances toward the best possible results over sequential generations and terminates if the population has converged [32]. GA is more robust compared to other conventional algorithms. Genetic Algorithms provides significant result in large and multi-modal space compared to other forecasting and searching techniques and it is capable of handling reasonably noise data [33].

Bayesian Networks (BN) are represented through the Directed Acyclic Graph in which each node correlates to an exclusive random variable and each edge correlates to a conditional dependency. The relationships that exist among the variables can be represented in the graph through the use of factors [34]. Road traffic and road infrastructure engineers can use the model for the classification of traffic accidents, hazardous locations of roads, factors affecting accidents severity and to conduct other parametric studies [35]. The Bayesian network analysis results help to trim down the number of accidents, accident severity and to develop safety regulations related to road traffic.

Support Vector Machines (SVM) can handle issues that involve both linear and nonlinear classification tasks, concerning separable or non-separable problems. The SVM model will be trained, tested, and compared using collected data. An SVM model maps the data points from the collected data to high-dimensional feature space, utilizing kernel function [36]. SVM models have considerably higher prediction accuracy as compared with the traditional approach and best for real-time traffic accident prediction.

Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are the two types of Deep Learning frameworks and a combination of the two are utilized to find out unknown relationships and configurations in high dimensional road accident data, text grouping, speech recognition, image processing and computer vision [37].

An artificial neural network (ANN) consists of a multi-layer feed-forward neural network with additional layers among the input and output layers. Through weight coefficients lines, every neuron in each layer is connected with each neuron in the adjoining layers. As a result, any change in the weight coefficients will change the outcome of the network. This will indicate the proposed values of weighted coefficients for getting the desired results. Thus neural network can be trained to perform a particular function by regulating the values of weights between elements [38]. To predict the accident causes, to forecast road accidents and to suggest traffic accident severity models are the common uses of Artificial Neural Networks in the accident forecasting area. ANN is exclusively proposed when the results from statistical analyses are difficult to interpret, better accuracy is needed for the forecasting results and recognising the method is unidentified for generating the data.

S. Y. Sohn et al. applied three data-mining techniques such as decision tree, neural network, and logistic regression, to decide a set of significant factors and established classification models for road accident severity. Then the three perspectives are compared concerning classification accuracy. It is found out that, accuracy does not differ significantly for each model and that the protective device is the key factor in the accident severity variation [39].

5. Advantage of statistical methods over others

S. Basu et al. statistically analysed strategies that were used to create models for crash expectation; and found that the greater part of them didn't think about the impact of heterogeneity of traffic blend. The authors constructed a statistical model which would forecast road accidents with reasonable precision [40].

M. Vilaca et al. executed a statistical examination to evaluate the severity and occurrence of road accidents associated with road users and this investigation established various relationships of accident situations, developments and improvements in road safety standards and new policies related to traffic safety [41].

M. Schlögl et al. compared a chain of statistical analysis techniques concerning the predictive performance to derive the influential factors correlated with road accident occurrence and found the importance of the ensemble of models. The findings of the research confirmed that a trade-off between sensitivity and accuracy is fundamental to variance classification problems and the analysis highlighted the advantages of using high-resolution data in the framework of road accident analysis [42].

By using a Zero Unitarization technique, I. Bak et al. conducted a statistical investigation of traffic safety, which empowered to rank voivodeship cities in Poland by the order of traffic crimes. The study examined the number of traffic offences and fatalities enlisted by the Police in road accidents, the detection pace of collected statistical data on traffic offences and applied multi-dimensional statistical analysis to examine the safety data concerning the culprits of traffic crimes [43].

To identify the correlation between road fatalities and shuttle distance travelled by various transport modes, R. Goel et al. has done a framework of an ecological regression model, at the state level in India. This framework helped in interfacing numerous modes on Indian roads and their approximate results related to road accidents. The final results revealed that 2 Wheelers, Heavy vehicles and passenger car are related to higher risk, while pedestrians, cycling and Intermediate Public Transport are correlated with a lower danger of road fatalities [44].

From different research literature related to the analysis of road accidents, it is found that the statistical methods will perform well when,

- The analyst wants to validate the statistical properties of the basic or hidden procedure that created the problem.
- There is a statistical method that resolves the specific accident associated issue better than neural networks.
- Better clarification of analysis results and fatalities are significant.
- Concerning the functional correlation between the risk factors in the issue, the data analyst possesses the expertise or strong theoretical knowledge.

Latest analytical frameworks, such as finite-mixture models, multi-state switching models and random-parameter simulations have shown tremendous potential to enhance the accuracy of statistical analysis with existing and new data sources [45].

6. Management of contributing factors to road accidents

Road accident data analysis is mostly driven by regression-based conventional statistical approaches, advanced heterogeneous statistical methods and data-driven methods such as neural networks, machine learning, artificial intelligent frameworks etc. These approaches have often been selected using accident data from reported accidents, although this may generate a challenge in uncovering the reasons behind the accident causalities due to the quality of the data. Thus the researchers sometimes have to sacrifice the predictive capability of the subsequent analysis and their ability to detect the causative nature of the accident contributing factors [46]. Various researches [47–51], conducted in the area of traffic accident analysis, detailed different elements that influence road accidents and identified the five key factors; are (i) human factors - debilitation, fatigue, distractions; (ii) vehicle design and physical condition; (iii) traffic conditions - traffic speed, thickness and variety in speed between vehicles; (iv) geometric characteristics of the road - the type of road, number of lanes, curvature, nearby ramps/intersections, etc.; and (v) weather conditions—rain, visibility, ice/sleet/snow, etc. The main factors are discussed below in detail.

6.1. Human factors - fatigue

From the heavy vehicle category, three explicit elements were identified, those have a direct effect on the driver fatigue and are; (i) irregular driver shifts; (ii) short or long delay time; and (iii) begin the workweek drained or not. Crum et al. have done a regression analysis with these elements for forecasting road accidents. The principal element, irregular driver shifts, was estimated by counting the number of six-hour time shifts the drivers regularly drove. The authors found that beginning the work week tired was the most significant indicator of the three. Long delay times were closely connected with narrow accident escapes and self-insight of fatigue. Paradoxically, the number of six-hour time shifts drivers driven per day was adversely connected with narrow escapes [52].

Since the complexity of managing a controlled experiment by enforcing treatments, most research designs are observational studies. They compared the consequences of factors that are observed, not enforced. Many researchers [53–56] used a group design or a case-control study. In the group design, several drivers are identified and calculated across time. In the case-control study, several accident cases were identified and were studied with controls. The focal point was then positioned on the variations between the accident cases and controls. Both group studies and case-control studies can help assess road safety.

6.2. Human factors - distracted driving

World Health Organization [57] identified that the use of cell phones while driving is the major issue of distraction in vehicle drivers. They noticed that use of mobile phones has increased to 11% in the previous 5 to 10 years. Their report states that mobile phones and other electronic gadget utilization increase the probability of a road accident by a factor of four. A. Chand et al. [58] noticed that, around one-fourth of the road accidents in heavy trucks and passenger vehicles were because of driver distractions from cell phones and other gadgets in vehicles. Olson et al. [59] examined distracted driving in 203 heavy-duty truck drivers. The information included 4452 significant incidents, like road crashes, narrow escapes and unexpected lane changes, 19,888 time-frames that involved unremarkable incidents. The authors found that 71% of road accidents and 46% of narrow escapes were associated with drivers who were involved in activities not connected with driving. The result showed that 60% of significant incidents happened when the driver was carrying out non-driving activities.

Klauer et al. [60] investigated the distraction in drivers, in which 42 amateur drivers (16 to 17 years old) who had recently obtained their driving licence and 109 professional drivers were examined. Accelerometers and cameras were utilized to recognize distracted movements while vehicle driving. They found that distracting activities like eating or mobile phone calling or messaging will assist in widening the danger from road accidents. Therefore, those activities should not be permitted while driving, to reduce accidents and maintain road safety.

6.3. Vehicle design and physical condition

The mechanical characteristics of the vehicle involved in road accidents contribute to this factor, such as poorly maintained vehicle; lack of modern safety features like lane assist, brake assist, electronic stability control and Anti-lock Braking System [61]. The driving environment design is also a prime factor. Driver fatigue, a major factor in vehicle accident risks, is due to the poorer driving environment design, especially in heavy-duty vehicles. The design should provide the least fatigue to the driver from long term driving for both commercial and passenger vehicles [62]. The Large Truck Crash Causation Study conducted by the U.S. Dept. of Transportation, Federal Highway Administration in 2001–2003 found that fatigue is a related cause in 13% of grievance and fatal crashes linking at the minimum one large truck [63].

6.4. Weather

The main variables determining whether conditions are visibility, temperature, wind speed, precipitation and moistness and represent the principal external factors that have an impact on the crash likelihood and seriousness [64]. These factors should be considered very important since their interactions are unpredictable and can significantly change accident probability.

6.5. Traffic conditions and road geometry

Wang et al. [65] investigated the significant elements that could give rise to high-risk traffic conditions. The authors analysed traffic, road geometry, weather, and driver behavioural attitudes, as the basic attributes of the area surrounding the accident. They did not consider the people in the accident. They utilized a case-control framework with a 10:1 proportion of non-accident to accidents, employed support vector machines (SVM) for element selection and Bayesian logistic regression for getting the results. They established the prominent factor as the home-based trip generation percentage, which integrates routine travellers and it had a greater correlation to high-risk road accidents.

7. Discussion and future challenges

The road accident data source selection for the analysis relies upon the traffic issue to be confronted. The techniques for investigation could be finalised only after selecting the factors and their impact on road accidents. The classification algorithms and decision trees are two broadly utilized analytic methods for traffic accident analysis due to their interpretability, even though they don't suggest results with elevated levels of correctness and precision when comparing with other analytic techniques. When Deep learning algorithms are utilized for accident forecasting in the domains of image processing and signal; it gives high potential results to analyse, scrutinise and predict road crashes Deep learning architectures possess the disadvantage of is their high computational necessities and the need for huge data sets for building over-fitting models.

Statistical methods got their advantages when using particular problems related to accident data. Using improved quality road accident data, such as naturalistic driving data and direct and multiple measurements for safety-related factors, should be detailed well before the analysis, which will give precise results.

Incorporating heterogeneous traffic data sources, which comprises geographic facts, statistics of traffic movements, traffic volume data, sound, video or text from social and other online media, can enhance the accuracy and precision of the results and accident forecasts and thereby improve the range of prediction models. Also, there must be a re-investigation on different causation factors of road accidents, using the latest technologies and advanced frameworks to collect more precise data, to re-assess the impact of them on transport safety.

Many of the behavioural and psychological aspects of human could not be efficiently captured in the existing road accident datasets, methods and frameworks, which is a significant drawback of the current scenario for accidents and forecast modelling. Researchers should aim to combine stress analysis and human emotions, derived from social media information, and find a way to resolve this drawback to increase the predictability of road accidents.

8. Conclusion

The design and implementation of analytical techniques are growing to a new era of unparalleled scenarios in the field of road accident analysis. Such change has been driven by a combination of recent developments in accident research models and the availability of promising new traffic data sources. With the introduction of more suitable and advanced data processing techniques and frameworks, accident research has gained significantly. The use of these techniques has allowed the researchers to derive substantial judgements from the analysis of traffic data. The accident analysts must pursue exploring basic methodological issues and continuously attempt to extend the technological boundaries, as research on the data analysis of road accident advances.

An overview of road accident data sources, techniques used to analyze the data, frameworks for predicting traffic accidents and the management of contributing factors were presented in this paper. Sophisticated, nonlinear techniques have both advantages and disadvantages and sometimes simpler models offer excellent results as compared to the sophisticated frameworks. The insight gained, from all modelling methods, is that the purposes of the data analysis are more important than the techniques used.

CRediT authorship contribution statement

Arun Chand: Conceptualization, Methodology, Validation, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing. **S. Jayesh:** Writing - review & editing. **A.B. Bhasi:** Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] World Health Organization, Road traffic injuries, News-Room/Fact-Sheets/Detail/Road-Traffic-Injuries. (2020). https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (accessed August 14, 2020).
[2] S. Kumar, D. Toshniwal, A data mining framework to analyze road accident data, J. Big Data. 2 (2015), https://doi.org/10.1186/s40537-015-0035-y.
[3] Internationaltransportforum, Road Safety Annual Report 2019, 2019.
[4] P.T. Savolainen, F.L. Mannering, D. Lord, M.A. Quddus, The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives, Accid. Anal. Prev. 43 (2011) 1666–1676, https://doi.org/10.1016/j.aap.2011.03.025.
[5] A.K. Prabhaker Mishra, C.M. Pandey, U. Singh, M. Sabaretnam, Selection of appropriate statistical methods for data analysis, Ann. Card. Anaesth. 22 (2019) 297–301, https://doi.org/10.4103/aca.ACA_248_18.
[6] US Gov, Open Data in the United States, Data.Gov. (2014).
[7] K. Kapoor, V. Weerakkody, U. Sivarajah, Open data platforms and their usability: Proposing a framework for evaluating citizen intentions, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2015. 10.1007/978-3-319-25013-7_21.
[8] D. Topolšek, E.A. Herbaj, M. Sternad, The accuracy analysis of measurement tools for traffic accident investigation, J. Transp. Technol. 04 (2014) 84–92, https://doi.org/10.4236/jtts.2014.41008.
[9] J. Yuan, M. Abdel-Aty, P.E. Pegasus, L. Wang, J. Lee, R. Yu, X. Wang, Utilizing Bluetooth And Adaptive Signal Control Data For Urban Arterials Safety Analysis, 2017.
[10] V.C. Maha Vishnu, M. Rajalakshmi, R. Nedunchezhian, Intelligent traffic video surveillance and accident detection system with dynamic traffic signal control, Cluster Comput. 21 (2018) 135–147, https://doi.org/10.1007/s10586-017-0974-5.
[11] Automatic Vehicle Accident Detection and Reporting With Black Box, (n.d.). https://www.researchgate.net/publication/290061841_Automatic_Vehicle_Accident_Detection_and_Reporting_With_Black_Box (accessed November 26, 2020).
[12] K.A. Khaliq, O. Chughtai, A. Shahwani, A. Qayyum, J. Pannek, Road accidents detection, data collection and data analysis using V2X communication and edge/cloud computing, Electron 8 (2019), https://doi.org/10.3390/electronics8080896.
[13] C. Chen, Analysis and forecast of traffic accident big data, ITM Web Conf. 12 (2017) 04029, https://doi.org/10.1051/itmconf/20171204029.
[14] A. Mehdizadeh, M. Cai, Q. Hu, M.A.A. Yazdi, N. Mohabbati-Kalejahi, A. Vinel, S. E. Rigdon, K.C. Davis, F.M. Megahed, A review of data analytic applications in road traffic safety. Part 1: Descriptive and predictive modeling, Sensors (Switzerland) 20 (2020) 1–24, https://doi.org/10.3390/s20041107.
[15] H. Rajput, T. Som, S. Kar, An automated vehicle license plate recognition system, Computer (Long Beach. Calif). 48 (2015) 56–61, https://doi.org/10.1109/MC.2015.244.
[16] U. Lokala, S. Nowduri, P.K. Sharma, Road Accidents Bigdata Mining and Visualization using Support Vector Machines., (2017). https://corescholar.libraries.wright.edu/knoesis (accessed November 26, 2020).
[17] F. Dai, M. Sujon, Measuring Current Traffic Safety Culture via Social Media Mining, n.d.
[18] B. Zhen Qian, Real-time Incident Detection Using Social Media Data, 2016.
[19] Z. Zheng, C. Wang, P. Wang, Y. Xiong, F. Zhang, Y. Lv, Framework for fusing traffic information from social and physical transportation data, PLoS One. 13 (2018), https://doi.org/10.1371/journal.pone.0201531.
[20] C. Gutierrez-Osorio, C. Pedraza, Modern data sources and techniques for analysis and forecast of road accidents: a review, J. Traffic Transp. Eng. (English Ed.) 7 (2020) 432–446, https://doi.org/10.1016/j.jtte.2020.05.002.
[21] E. Abdullah, A. Emam, Traffic accidents analyzer using big data, in: Proc. - 2015 Int. Conf. Comput. Sci. Comput. Intell. CSCI 2015, Institute of Electrical and Electronics Engineers Inc., 2016: pp. 392–397. 10.1109/CSCI.2015.187.
[22] A. Sarkar, U.C. Sahoo, G. Sahoo, Accident prediction models for urban roads, Int. J. Veh. Saf. 6 (2012) 149–161, https://doi.org/10.1504/IJVS.2012.049020.
[23] C.P. Castañeda, J.G. Villegas, Analyzing the response to traffic accidents in Medellín, Colombia, with facility location models, IATSS Res. 41 (2017) 47–56, https://doi.org/10.1016/j.iatssr.2016.09.002.
[24] J.T. Wong, Y.S. Chung, Comparison of methodology approach to identify causal factors of accident severity, Transp. Res. Rec. (2008) 190–198, https://doi.org/10.3141/2083-22.
[25] J. Xi, Z. Gao, S. Niu, T. Ding, G. Ning, A hybrid algorithm of traffic accident data mining on cause analysis, Math. Probl. Eng. 2013 (2013), https://doi.org/10.1155/2013/302627.
[26] R. Bhavsar, A. Amin, L. Zala, Development of model for road crashes and identification of accident spots, Int. J. Intell. Transp. Syst. Res. (2020), https://doi.org/10.1007/s13177-020-00228-z.
[27] M.G. Karlaftis, E.I. Vlahogianni, Statistical methods versus neural networks in transportation research: differences, similarities and some insights, Transp. Res. Part C. 19 (2011) 387–399, https://doi.org/10.1016/j.trc.2010.10.004.
[28] M.A.M. Al-bared, A. Marto, Applications of Deep Learning in Severity Prediction of Traffic Accidents, Springer, Singapore, 2019, 10.1007/978-981-10-8016-6.
[29] A. Irfan, R. Al Rasyid, S. Handayani, Data mining applied for accident prediction model in Indonesia toll road, AIP Conf. Proc. 1977, 2018, 10.1063/1.5043013.
[30] D. Lord, F. Mannering, The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives, Transp. Res. Part A. 44 (2010) 291–305, https://doi.org/10.1016/j.tra.2010.02.001.
[31] D.D. Clarke, R. Forsyth, R. Wright, Behavioural factors in accidents at road junctions: the use of a genetic algorithm to extract descriptive rules from police case files, Accid. Anal. Prev. 30 (1998) 223–234, https://doi.org/10.1016/S0001-4575(97)00080-8.
[32] K. Assi, S.M. Rahman, U. Mansoor, N. Ratrout, Predicting crash injury severity with machine learning algorithm synergized with clustering technique: a promising protocol, Int. J. Environ. Res. Public Health. 17 (2020) 1–17, https://doi.org/10.3390/ijerph17155497.
[33] Y. Li, D. Ma, M. Zhu, Z. Zeng, Y. Wang, Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network, Accid. Anal. Prev. 111 (2018) 354–363, https://doi.org/10.1016/j.aap.2017.11.028.
[34] Y. Castro, Y.J. Kim, Data mining on road safety: factor assessment on vehicle accidents using classification models, Int. J. Crashworthiness 21 (2016) 104–111, https://doi.org/10.1080/13588265.2015.1122278.
[35] M. Deublein, M. Schubert, B.T. Adey, J. Köhler, M.H. Faber, Prediction of road accidents: a Bayesian hierarchical approach, Accid. Anal. Prev. 51 (2013) 274–291, https://doi.org/10.1016/j.aap.2012.11.019.
[36] Z. Li, P. Liu, W. Wang, C. Xu, Using support vector machine models for crash injury severity analysis, Accid. Anal. Prev. 45 (2012) 478–486, https://doi.org/10.1016/j.aap.2011.08.016.
[37] A. Naseer, M.K. Nour, B.Y. Alkazemi, Towards Deep Learning based Traffic Accident Analysis, in: 2020 10th Annu. Comput. Commun. Work. Conf. CCWC 2020, Institute of Electrical and Electronics Engineers Inc., 2020: pp. 817–820. 10.1109/CCWC47524.2020.9031235.
[38] F. Rezaie Moghaddam, Sh. Afandizadeh, M. Ziyadi, Prediction of accident severity using artificial neural networks, Int. J. Civ. Eng. 9 (2010). www.iust.ac.ir (accessed January 15, 2021).
[39] P. Taylor, S.Y. Sohn, H. Shin, Pattern recognition for road traYc accident severity in Korea, 2010.
[40] S. Basu, P. Saha, Regression models of highway traffic crashes: a review of recent research and future research needs, Procedia Eng. 187 (2017) 59–66, https://doi.org/10.1016/j.proeng.2017.04.350.
[41] M. Vilaça, N. Silva, M.C. Coelho, Statistical analysis of the occurrence and severity of crashes involving vulnerable road users, Transp. Res. Procedia. 27 (2017) 1113–1120, https://doi.org/10.1016/j.trpro.2017.12.113.
[42] M. Schlögl, R. Stütz, G. Laaha, M. Melcher, A comparison of statistical learning methods for deriving determining factors of accident occurrence from an imbalanced high resolution dataset, Accid. Anal. Prev. 127 (2019) 134–149, https://doi.org/10.1016/j.aap.2019.02.008.
[43] I. Bąk, K. Cheba, B. Szczecińska, The statistical analysis of road traffic in cities of Poland, Transp. Res. Procedia. 39 (2019) 14–23, https://doi.org/10.1016/j.trpro.2019.06.003.
[44] R. Goel, Modelling of road traffic fatalities in India, Accid. Anal. Prev. 112 (2018) 105–115, https://doi.org/10.1016/j.aap.2017.12.019.
[45] F.L. Mannering, Analytic Methods in Accident Research : Methodological Frontier and Future Directions, (2013) 1–69.
[46] F. Mannering, C.R. Bhat, V. Shankar, M. Abdel-Aty, Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis, Anal. Methods Accid. Res. 25 (2020), https://doi.org/10.1016/j.amar.2020.100113 100113.
[47] J.J. Rolison, S. Regev, S. Moutari, A. Feeney, What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records, Accid. Anal. Prev. 115 (2018) 11–24, https://doi.org/10.1016/j.aap.2018.02.025.
[48] K.J. Parnell, N.A. Stanton, K. Plant, Where are we on driver distraction? Methods, approaches and recommendations, Theor. Issues Ergon. Sci. 19 (2018) 578–605, https://doi.org/10.1080/1463922X.2017.1414333.
[49] R.L. Hammond (Olson), S.A. Soccolich, R.J. Hanowski, The impact of driver distraction in tractor-trailers and motorcoach buses, Accid. Anal. Prev. 126 (2019) 10–16, https://doi.org/10.1016/j.aap.2018.03.015.
[50] European Road Safety Observatory, Driver Distraction 2015, 2015.
[51] F.L. Mannering, V. Shankar, C.R. Bhat, Unobserved heterogeneity and the statistical analysis of highway accident data, Anal. Methods Accid. Res. 11 (2016) 1–16, https://doi.org/10.1016/j.amar.2016.04.001.
[52] M.R. Crum, P.C. Morrow, The influence of carrier scheduling practices on truck driver fatigue, Transp. J. 42 (2002) 20–41 (accessed January 15, 2021) https://www.jstor.org/stable/20713513?seq=1.
[53] C. Chen, J. Zhang, Exploring background risk factors for fatigue crashes involving truck drivers on regional roadway networks: a case control study in Jiangxi and Shaanxi, China, Springerplus 5 (2016) 1–12, https://doi.org/10.1186/s40064-016-2261-y.
[54] Z. Gao, X. Wang, Y. Yang, C. Mu, Q. Cai, W. Dang, S. Zuo, EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation, IEEE

Trans. Neural Networks Learn. Syst. 30 (2019) 2755–2763, https://doi.org/10.1109/TNNLS.2018.2886414.
[55] M. Gastaldi, R. Rossi, G. Gecchele, Effects of driver task-related fatigue on driving performance, Procedia - Soc. Behav. Sci. 111 (2014) 955–964, https://doi.org/10.1016/j.sbspro.2014.01.130.
[56] W.J. Horrey, M.F. Lesch, A. Garabet, L. Simmons, R. Maikala, Distraction and task engagement: How interesting and boring information impact driving performance and subjective and physiological responses, Appl. Ergon. 58 (2017) 342–348, https://doi.org/10.1016/j.apergo.2016.07.011.
[57] World Health Organization; 2018. Licence: CC BY- NC-SA 3.0 IGO, Global Status Report On Road Safety 2018, Geneva, 2018. 10.29333/aje.2019.423a.
[58] A. Chand, A.B. Bhasi, Effect of driver distraction contributing factors on accident causations-a review, AIP Conf. Proc. 2134 (2019), https://doi.org/10.1063/1.5120229.
[59] R.L. Olson, R.J. Hanowski, J.S. Hickman, J. Bocanegra, Driver Distraction In Commercial Vehicle Operations, 2009. FMCSA-RRR-09-042.
[60] B.G. Simons-Morton, F. Guo, S.G. Klauer, J.P. Ehsani, A.K. Pradhan, Keep your eyes on the road: young driver crash risk increases according to duration of distraction, J. Adolesc. Heal. 54 (2014) S61–S67, https://doi.org/10.1016/j.jadohealth.2013.11.021.
[61] J.B. Cicchino, Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates, Accid. Anal. Prev. 99 (2017) 142–152, https://doi.org/10.1016/j.aap.2016.11.009.
[62] Y. Rong, Z. Akata, E. Kasneci, Driver intention anticipation based on in-cabin and driving scene monitoring, ArXiv (2020), https://doi.org/10.1109/itsc45102.2020.9294181.
[63] Federal Motor Carrier Safety Administration, The Large Truck Crash Causation Study - Analysis Brief, 2007. FMCSA-RRA-07-017.
[64] R. Bergel-Hayat, M. Debbarh, C. Antoniou, G. Yannis, Explaining the road accident risk: weather effects, Accid. Anal. Prev. 60 (2013) 456–465, https://doi.org/10.1016/j.aap.2013.03.006.
[65] L. Wang, C. Yu, Y. Zhang, L. Luo, G. Zhang, An analysis of the characteristics of road traffic injuries and a prediction of fatalities in China from 1996 to 2015, Traffic Inj. Prev. 19 (2018) 749–754. 10.1080/15389588.2018.1487061.