International Journal of Scientific Research in Science, Engineering and Technology
Print ISSN: 2395-1990 | Online ISSN : 2394-4099 (www.ijsrset.com)
doi : https://doi.org/10.32628/IJSRSET

A Survey on Road Accident Prediction Techniques Based on Various Methodologies

Atul Pandey, Virendra Pratap Yadav
Sheat College of Engineering, Varanasi, Uttar Pradesh, India

Article Info
Volume 9, Issue 4

Page Number : 350-357

Publication Issue :
July-August-2022

Article History
Accepted : 05 August 2022
Published: 16 August 2022

ABSTRACT

Since traffic accidents are a leading source of injury and death globally, there has been a lot of focus on developing more accurate methods of analysis and prediction in order to pinpoint the causes of these tragedies. Predicting traffic accidents is an effort to meet the problem of creating a safer transportation environment in order to save lives. The purpose of this study is to survey the current landscape of research into the use of convolutional neural networks, long short-term memory networks, and other deep learning architectures for the prediction of traffic accidents. In addition, the most popular data sources for predicting traffic accidents are compiled here and analyzed. Additionally, a categorization is recommended based on factors including its source and features, such as open data, measuring methods, onboard equipment, and social media data. In this section, we list and evaluate the many algorithms used to forecast traffic accidents, taking into account the data types for which each is most suitable, the accuracy of the findings, and the clarity with which they can be interpreted and studied. In order to further analyze the findings, the authors found that the best results were achieved by combining two or more analytic approaches. Many authors agree that using geospatial data, information from traffic volume, traffic statistics, video, sound, text, and sentiment from social media may improve the precision and accuracy of the analysis and predictions; this is one of the next challenges in road traffic forecasting.

Keywords - Traffic accident prediction, Road accident forecasting, Data analysis, Traffic engineering, Machine learning

I. INTRODUCTION

A. Overview

Many countries have paid little attention to reducing the severity of road traffic accidents (RTAs), despite the fact that they kill thousands of people and destroy millions of dollars' worth of property every day. The fact remains, however, that it is a leading global killer and destroyer of property. The severity of the impact on human life and property may be mitigated by finding and fixing the root causes of road traffic accidents. Road Adverse events of severe magnitude are not random occurrences, but rather follow predictable patterns that allow for their mitigation. Ergo, mishaps are "observable, measurable, and preventable occurrences" [20]. "Fatalities are not fated; accidents are not random; disease is not arbitrary; it is caused" [33] is how the workers' health association defines accidents. Every day, people were injured or killed in traffic accidents in Addis Ababa, Ethiopia's capital. In a matter of seconds, human lives may be lost

and property can be destroyed. It's a major killer in this nation, and it's one of the scariest things about it.

The severity of road traffic accidents has been a focus of study over the last two decades. Road accident severity categorization based models have been the subject of several intriguing research methodologies. The authors analyzed data in a conventional statistical manner to construct their models. These methods are useful for learning about and figuring out what contributes to road accidents. Due to the availability of large datasets, machine learning has surpassed traditional statistical methods for model prediction in recent years [41]. The causes of severe road traffic accidents have been the subject of several academic works [7, 9, 36, 37, 40, 43, 45] from a variety of nations. Studies aiming to forecast the severity of traffic accidents are still in their early stages of development. A mixed machine learning strategy was used to enhance classification accuracy in the prior work. This gap in the market is something we want to solve by developing a hybrid machine learning technique for road accident categorization, which will increase the efficiency and precision of our forecasts. The prior research focuses primarily on the efficiency of a Machine Learning-based categorization strategy. In contrast, there is a lack of research comparing deep learning algorithms with state-of-the-art Hybrid Machine Learning methods. Accurate predictions may often be improved by using the most appropriate method. Therefore, the most important causes of road accidents may be isolated with the aid of the best paradigm. Additionally, prior identification and concern for target-specific contributing elements was lacking. In order to forecast the severity of traffic accidents, the researchers utilized a combination of clustering and classification techniques. In this paper, we present a novel hybrid approach based on K-means clustering and random forest for predicting the severity of traffic accidents at a given location. The effectiveness of the created model was evaluated by comparing the suggested method to that of separate classifiers. Measures of accuracy, precision, specificity, and recall are employed to contrast the novel method with traditional approaches. There are a few stages to the new method: There are five main steps: (I) cleaning the data by removing unwanted noise and filling in missing data using the mean for numerical variables and the mode for the categorical variable, (II) dividing the data into a training and test dataset, (III) developing a novel feature through clustering, (IV) training classifiers, and (V) finally assessing the effectiveness of each classifier. The suggested method was further evaluated in comparison to state-of-the-art classification methods by using a deep neural network. The results of the study demonstrated that the suggested method outperformed the competition in terms of classification accuracy and overall performance.

- Accuracy: represent the rate of instances correctly classified over the total number of instances. The ideal value for accuracy is 1.00 (100% classification accuracy).
- Precision or confidence: defined as the proportion of predicted positive cases that are correctly predicted or labeled as real positives.
- Recall: calculated as the proportion of real positive cases that are accurately predicted positive.
- F-measure: weighted average of the precision and recall.
- Mean absolute error: also known as average prediction error, is the average of the difference between predicted and actual value in all the test cases. A low mean absolute error (MAE) indicates good predictive accuracy.
- Mean squared error: determined as the average of the squared differences between each computed value and its corresponding correct value.
- Root mean squared error (RMSE): RMSE is calculated as the square root of the MSE and is used as a measure of differences between valued predicted and the real values. A lower value of RMSE is an indicative of a higher prediction precision.

Traditional statistical model-based strategies were employed to forecast accident fatalities and severity in the field of road safety. Several standard statistical studies have been employed, including the mixed logit modeling technique [23, 26], the ordered Probit model [54], and the logit model [11]. The standard statistical approach was shown by some research to be more effective than other studies in distinguishing between independent and dependent accident variables [31].

However, traditional statistical methods cannot handle multidimensional data sets [16]. Many recent research have used the ML technique as a means of overcoming the shortcomings of conventional statistical models on account of its superiority in terms of prediction, efficiency, and depth of information. In the last ten years, ML has been used in a variety of fields, including building and construction [48], workplace accidents [41], farming [22], education [53] and sentiment analysis [50] and finance and insurance [46].

To construct an accident severity model, K-means, Support Vector Machines, K-Nearest Neighbors (KNN), Decision Tree (DT), Artificial Neural Network (ANN), Convolution Neural Network (CNN), and Logistic Regression (LR) are among the most effective clustering and classification techniques. Based on data gathered in California between 2004 and 2010, Kwon et al. [28] used Nave Bayes (NB) and Decision Tree (DT). Using binary regression, the authors found that although both the Nave Bayes and the Decision Tree models performed well, the former was more attuned to the presence of risk variables.

Using support vector machines and multilayer perceptrons, Sharma et al. [44] evaluated data on traffic accidents. The authors also relied heavily on only two independent variables (alcohol and speed) in their analysis. In the end, the SVM with the RBF kernel outperformed the MLP (64%) with a higher level of accuracy (94%). According to the results, driving too fast while under the influence of alcohol is the leading cause of collisions.

In order to assess classifiers and determine the most important causes of motorcycle accidents, Wahab and Jiang [51] used MLP, PART, and SimpleCART on the crash events in Ghana dataset. The authors employed Weka tools for data comparison and analysis, and they used InfoGainAttributeEval to identify the most significant factor in motorbike accidents in Ghana. In this regard, the simpleCART model outperformed competing classification methods.

II. DATA SOURCES

Road accident analysis and prediction data sources including government data, open data, measurement technologies, vehicle onboard equipment and social media.

Government data	Data sets that are generated, collected, preserved, stored and made available to the public by government entities or those that are delegated to exercise functions of control, execution or reporting of information concerning road accidents
Open data	Open data catalogs are maintained by government agencies and are available to all public without restriction. The data must comply all legislation regarding privacy and confidentiality
Onboard equipment	Onboard equipment refers to all devices installed on a vehicle that can store or transmit data concerning the vehicle variables and driver conditions
Measurement technologies	Measurement technologies include all kind of equipment that is part of the road infrastructure, such as radar, cameras, or equipment embedded on the road itself, i.e., loop detectors
Social media	Social media can be considered the newest developed data source in traffic and road accident related studies, and currently the most used data source comes from Waze, Inrix, Google Maps and Twitter streams

III. MEASUREMENT TECHNOLOGIES

Instruments like radar, cameras, and even hardware permanently installed in the road (like loop detectors) are all considered measurement technologies.

Many investigations have made advantage of readily accessible technologies like the loop detector, video surveillance, microwave, and laser radar. Road junction data have also been collected using Bluetooth detectors and adaptive signal control databases. Since the variables in a loop detector record are limited to

vehicle type, vehicle speed, record time, and loop-specific information like localization and status, this data is not high-dimensional. When compared to other types of road infrastructure like cameras and radars, loop detector arrays are very cheap and may be installed along a major highway or expressway. However, loop detectors are not very trustworthy because to their susceptibility to failure in extreme temperatures, vibrations, and pavement fluctuations.

IV. ONBOARD EQUIPMENT

Any device put in a vehicle for the purpose of storing or transmitting information on vehicle variables and driving circumstances is considered onboard equipment. GPS units, cameras aimed at documenting road conditions or the driver's state of consciousness (Zheng et al., 2014), accelerometers, vehicle condition recorders that log data like speed, abrupt braking, lane changes, and impact or collision direction and acceleration may all be present. Using the PreScan platform to simulate traffic accidents is a novel strategy described by Xiong et al. (2017), who used a sophisticated system known as the chain road traffic incident.

V. ROAD ACCIDENT ANALYTIC METHODS

By using analytic methods, researchers seek to characterize the information and variables of the road accident, in order to discover hidden patterns, profile behaviors, generate rules and inferences. These patterns are useful to profile drivers or drivers' behavior on the road, to delimitate unsafe areas for driving, to generate classification rules related to road accident data, to perform selection of variables to be fetched in real-time model of accidents and to select relevant variables to be used to train other methods, such as artificial neural networks and deep learning algorithms.

On the aspect of the algorithms and computational methods reported by the authors employed to analyze road accident data, as summarized in Table 2, the most used are: i) clustering algorithms (Cao et al., 2015; Kumar and Toshniwal, 2015a; Moriya et al., 2018); ii) decision trees and classifiers (Castro and Kim, 2016; Gutierrez-Osorio and Pedraza, 2019; Scott-Parker and Oviedo-Trespalacios, 2017; Taamneh et al., 2017); iii) association rules (Ait-Mlouk et al., 2017; Ait-Mlouk and Agouti, 2019; Kumar and Toshniwal, 2015b) and iv) natural language processing algorithms (D'Andrea et al., 2015; Gu et al., 2016; Salas et al., 2018).

Representative studies and methods on road accident data analysis.

Author	Research problem-computational method	Data source	Result
Cao et al. (2015)	Correlate abrupt braking events in real time-batch clustering, fuzzy C-means and real time clustering	Data from driving events for seven vehicles by the DAP platform from Ford Motor Company	Correlations that indicate potentially dangerous places for driving, according to the time of day
Kaplan and Prato (2013)	Determinate the variables that influence the severity of road accidents between cyclist and drivers (latent class clustering)	Data reported in Denmark (2007–2011), accidents involving cyclist and drivers	13 clusters showing specific patterns of urban and rural road accidents; obtaining a high classification accuracy, with all the clusters being correctly assigned for more than 80 percent of the observati

			ons, and reporting an entropy criterion of 0.86
Depaire et al. (2008)	Find patterns of severity of injuries resulting from road accidents in a heterogeneous data set (latent class clustering)	Accident data reported by the Belgian road police (1997–1999), 29 variables and 4028 accident records	7 clusters showing a high level of accidents for motorcyclist and cyclists under 19 years old
De Oña et al. (2013)	Identify the key factors that affect the severity of injuries caused by a rural road accident (latent class clustering, Bayesian networks)	3229 traffic accidents reported by the police in rural roads of Granada (Spain), occurred between 2005 and 2008	Results depends strongly on the initial data set analyzed and the techniques used
Zheng et al. (2014)	Determinate the variables that identify a driving style or driver with high risk of vehicular collision (cluster K-means)	31 vehicles with a GPS for 60 days, driving recorders and cameras to capture the road and driver's	3 clusters for road accident risk levels, a correlation between driving events and maximum
		facial expression	deceleration
Kumar and Toshniwal (2015a)	Determinate the variables that influence the event of road accidents (cluster K-means, association rules model)	11,574 traffic events on the roads of Dehradun (India) (2009–2014)	6-cluster model as input to a model of association rules. Severity of accident, type of road, lighting and surrounding area affect the aggregation of the clusters
Taamneh et al. (2017)	Determination of the most important variables for severity prediction of traffic accident (J48 decision tree, rule induction PART, Naive Bayes)	5973 traffic accident records occurred in Abu Dhabi between 2008 and 2018	Age, gender, nationality, year of the accident affect the severity of the accident
Beshah et al. (2011)	Understand the interaction between the different actors that intervene in a road accident (CART and random forest)	14,254 traffic accidents with 48 attributes (May 2005–September 2008) Ethiopia	CART and RF behave similarly but RF has lower failure rates to predict the probability of someone

			emerging unscathed from a road accident
Ahmed and Abdel-Aty (2012)	Prediction of traffic accidents in real time with information provided by an automatic vehicle identification (AVI) (random forest)	Real-time data on speed, average speed and traffic volume obtained from AVI along 125 km highway at Orlando (FL) 2008	The model is sensitive to the distance between each tag to AVI, over 4 km nor statistically or predictive significant values were obtained

VI. ROAD ACCIDENT FORECAST METHODS

Representative results on road accident forecast methods.

Road accident analytic method	Metric	Best result
Clustering algorithms	Bayesian information criterion (BIC), Akaike information criterion (AIC)	Moriya et al. (2018) with minimum values of AIC and BIC at 3 clusters
Classification algorithms and decision trees	Accuracy, precision, recall and F-measure, using receiver operating characteristic (ROC) curve	Tiwari et al. (2017), with an accuracy of 0.8235
Natural language processing	Accuracy, precision, recall and F-measure, using receiver operating characteristic (ROC) curve	D'Andrea et al. (2015) reported and accuracy value of 0.9575

Regarding road accident forecasting, as shown in Table 5, deep learning architectures, usually employed in the fields of signal and image processing, shows promising results to identify, analyze and forecast traffic accidents. The drawback of deep learning algorithms is their elevated computational requirements and the need of extensive data sets that can be subject to the possibility of produce over fitting models. The model proposed by (Ren et al., 2017) can be considered a baseline model for predicting traffic accident risk, since it incorporates big traffic accident data, as called by the authors, and proposed a novel deep learning architecture based on LSTM to predict the risk with accurate results. It can be remarked the novel approach proposed to model the data, using an encoding matrix that represents the spatial–temporal frequency of traffic accidents. Furthermore, the encoding matrix was developed using a heat map, which allowed visually highlighting the space-time zones with the highest road accident frequency values.

Representative metrics and results on road accident forecast.

Road accident forecasting method	**Metric**	**Best result**
Bayesian networks	Accuracy, precision, recall and F-measure, using receiver operating characteristic (ROC) curve	Castro and Kim (2016), accuracy is 0.8159, precision is 0.7239, recall value is 0.7239, F-measure is 0.723
Genetic algorithms and evolutionary computing	Accuracy, precision, recall and F-measure	Hasheminejad (2017), precision is 0.885, recall is 0.889, accuracy is 0.8820, F-measure is 0.8875

Support vector machines	Accuracy, precision, recall and F-measure	Xiong et al. (2017), accuracy is 0.8730
Artificial neural networks	Accuracy measure correlation, R-squared, MSE and RMSE	Alkheder et al. (2017), accuracy is 0.7460
Deep learning	Mean absolute error (MAE), mean relative error (MRE), mean squared error (MSE) and root mean squared error (RMSE)	Ren et al. (2017), MAE is 0.014, MSE is 0.001, RMSE is 0.0340

VII. CONCLUSIONS AND FUTURE WORK

The researches reviewed, were limited by the lack of incorporation of other relevant factors and variables, such as traffic flow, human mobility and special events that can affect traffic and accident risk, i.e., massive events. Furthermore, in order to provide an effective forecast and analysis, the models output was coarse-grained, using data that comprise in spatial variables, road segments or city grids, and in temporal terms, day, or hours, that cannot be disaggregated. The results lacking predictions and analysis that can provide road segment level results and temporal analysis that cannot be drill down to minutes.

Considering the analytic methods for road accident analysis, the classification algorithms and decision trees are widely employed by their interpretability, but, in the other hand, they do not offer results with such high levels of precision and accuracy compared to other methods. Because of this, it can be considered that the approach proposed by Tiwari et al. (2017), as shown in Table 4, is valuable, since their research obtain better results by using clustering algorithms to preprocess the data set, in this particular case, hierarchical clustering and K-modes clustering were evaluated. The results obtained improved the performance of the classifiers methods. Regarding the natural language processing of social media information related to traffic accidents, the work by D'Andrea et al. (2015) was innovative and a baseline model to other authors, since their model was compared against other algorithms, such as Naive Bayes classifier, C4.5 decision tree, K-nearest neighbor (KNN) and PART classifier, and their model was put to test in task of classification of real-time twitter streams, with successful results.

VIII. REFERENCES

[1]. R. Nayak et al., "Road Crash Proneness Prediction using Data Mining". Ailamaki, Anastasia & Amer-Yahia , Sihem (Eds.) Proceedings of the 14th International Conference on Extending Database Technology, Association for Computing Machinery (ACM), Upp-sala, Sweden, pp. 521-526, 2019.

[2]. J. Hipp, U. Güntzer, G. Nakhaeizadeh, "Algorithms for Association Rule Mining &Mdash; a General Survey and Comparison", SIGKDD Explor Newsl, pp. 58–64, 2020.

[3]. A.T. Kashani et al., "A Data Mining Approach to Identify Key Factors of Traffic Injury Severity", PROMETTraffic& Transportation, pp. 11-17, 2021.

[4]. P. J. Ossenbruggen, J. Pendharkar et. al., "Roadway safety in rural and small urbanized areas", Accidents Analysis & Prevention, 33(4), pp. 485-498, 2021.

[5]. R. Agrawal, T. Imieliski, A. Swami, "Mining Association Rules Between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, ACM, New York, NY, USA, pp. 207–216, 2018.

[6]. R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp 487-499, 2019.

[7]. L. Breiman, "Random Forests", Machine Learning, Vol. 45, pp. 5 - 32, 2021

[8]. Savolainen P, Mannering F, Lord D, Quddus M. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accid Anal Prev. 2021;43:1666–76.

[9]. Depaire B, Wets G, and Vanhoof K. Traffic accident segmentation utilizing latent class clustering, accident analysis, and prevention, vol. 40. Elsevier; 2018.

[10]. Jones B, Janssen L, Mannering F. Analysis of the frequency and duration of freeway accidents in Seattle, accident analysis and prevention, vol. 23. Elsevier; 2021.

[11]. Miaou SP, Lum H. Modeling vehicle accidents and highway geometric design relationships, accident analysis and prevention, vol. 25. Elsevier; 2020.

[12]. Abdel-Aty MA, Radwan AE. Modeling traffic accident occurrence and involvement. Accid Anal Prev Elsevier. 2021;32.

[13]. Han J, Kamber M. Data Mining: Concepts and Techniques. USA: Morgan Kaufmann Publishers; 2021.

[14]. Pardillo-Mayora JM, Domínguez-Lira CA, Jurado-Pina R. Empirical calibration of a roadside hazardousness index for Spanish two-lane rural roads. Accid Anal Prev. 2019;42:2018–23.

[15]. Goodman LA. Exploratory latent structure analysis using both identifiable and unidentifiable models. Biometrica. 2020.

Cite this article as :

Sonam Singh, Shailesh Singh , "A Survey on Road Accident Prediction Techniques Based on Various Methodologies", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 9 Issue 4, pp. 350-360, July-August 2022.
Journal URL : https://ijsrset.com/IJSRSET229460