Review and analysis of roadway crash prediction studies on urban roads under heterogeneous traffic conditions

Bodanapu Sony¹, Ch.Hanumantha Rao²

¹Research scholar, Department Civil Engineering, KL deemed to be University, Vijayawada, India.
²Professor and Head, Department Civil Engineering, KL deemed to be University, Vijayawada, India.

ABSTRACT: In recent decades, pre-predicting the roadway accidents is essential for real time traffic incident management that effectively minimizes the environmental pollution, traffic congestion and secondary incidents. Currently, the traffic data are available in thousands of public and private datasets and also generates terabytes of data each year. Though, it is infeasible to manage the huge datasets by utilizing traditional software and hardware. It is therefore essential that an automated system to predict road accidents is developed. The present review paper investigates the researches done on road accident prediction, particularly for urban roads under heterogeneous traffic conditions. It also explores the problems faced in existing works by researchers. This review paper helps researchers achieve a better solution for the current problems faced by heterogeneous traffic conditions when it comes to urban road accident prediction. The findings demonstrate that the operating speed and the disparities between the speed restrictions and the operating speed are the key factors influencing the accident frequency rate.

Keywords: Data Mining, Heterogeneous Traffic Condition, Machine Learning Algorithms, and Road Accident Prediction.

1. INTRODUCTION

In a developing country, traffic safety is a major problem. Studies have been widely reported on road safety since the 1960s, and Solomon (1964) has been reporting road crashes with speed, automobility and drivers on rural highways. Various studies found factors contributing to road crashes and reduced crash rates. Many of the reported studies deal with traffic collision geometry [1]. Road crashes are a complex occurrence and depend on a variety of facets, including traffic and composition, road design, speed differences, weather, motivation for travel, driver conditions, and so on. Road crashes (Aljanahi et al., 1999).

The transport conditions in India are highly heterogeneous and the transportation of vehicles includes a variety of static and dynamic vehicles [2-6]. In conditions of high volumes, vehicles of this heterogeneous traffic therefore move along Indian highways without adequate side and length of clearance, sharing the existing rooms. The slow motion also increases the complexity of the heterogeneous analysis/modelling of transport. Heterogeneous traffic should be tested for the effects of speed reduction on same-class road safety vehicles and different class vehicles and their speed differences along curves [7-12].
In prior research works, most of the developed algorithms mainly depend on relative distance and relative velocity using vehicle kinematics and dynamics (Kumar, and Toshniwal 2017). These metrics are easily obtained and understand in decision making development, and it is difficult to describe the road-way conditions, vehicle state, environmental characteristics, and driver behaviour (Hashimoto, et al, 2016).

1.1 Objective of the research

The investigators are pursuing several major road accident prediction objectives.

- Develop a new method of pre-processing to improve raw data quality. In addition, fill the absent values with suitable values by investigating the past and previous results.

- An effective optimization approach is developed to address “curse of dimensionality” problem. The developed optimization approach extracts the useful information from the collected data.

- Develop an unsupervised classifier for effectively predicting the road accident severity.

2. Literature review

Several research reports have reported about 0.5 million road accidents occurred in India in 2010, of which approximately 1 million such road accidents occurred, which indicates an estimated 56 accident numbers per hour. Fig. 2.1 shows that in recent years road accidents have substantially increased and that most deaths have resulted.

![Figure 2.1. Trends in population growth percentage changes, fatal and non-fatal accidents between 2004 and 2014. The Open Government Data Platform India is the source of road traffic accidents in India, 2013.](https://www.flickr.com)

A significant proportion of the car groups with a varying size, form and occupation consists of a heterogeneous traffic (Figure 2.2a) (two-wheelers, cars, three-wheelers, buses, trucks etc.). In contrast to heterogeneity, homogeneous traffic mainly includes motor vehicles (or any type of vehicle) with strict follow-up to lane discipline (Figure 2.2b).

![Figure 2.2. Homogeneous and Heterogeneous traffic.](http://daily.blaskar.com)
While the difference between the same and heterogeneous traffic is very significant and evident, most heterogeneous traffic studies are based mainly on homogenous traffic methods. This generalization can lead to serious mistakes if its applicability is not tested and confirmed under heterogeneous conditions.

In addition, many assumptions of uniform transport are infringed under heterogeneous circumstances. Any transport study also involves testing assumptions.

Free flow speed is mainly determined by geometric design and traffic features of rural highways and multi-lane highways, while it is also influenced by the urban street environment in urban contexts aside from its road geometry.

**Antoine Hebert, 2019,** Suggested proposal Road accidents are a major issue for our modern societies, which every year in the world cause millions of deaths or injury. In 2018 alone, 359 deaths and 33 thousand injuries were caused in road accidents in Quebec. In this article we show how open datasets of a city such as Montreal, Canada can be used to produce high-resolution model accident prediction with large data analytics.

**Sobhan Moosavi, (2019),** A major problem for public safety are traffic accidents, with many research into the analysis and forecast of rare occurrences. Nevertheless, the majority of studies use small data sets, rely on extensive data, which other researchers are not easily accessible and which are not applicable in real time. We have introduced a new framework for traffic accident prevention to address those challenges in real time, based on a simple yet scanty amount of information..

**Shahriar Afandizadeh, (2020),** Examined the safety risk index reliability for the clusters, the results of the proposed models were analyzed for sensitivity between their model results. In addition, a high-risk cluster is the maximum risk index for the proposed study. In the sixth cluster of Leur and Sayed, however, the risk index is lower than that of the study proposed. The difference depends on the difference between speed limits and operating speed in the current study in the development of the rural risk index..

**Hui Zhang, (2020),** The study produced model crash risk analysis for a rear-end crash and side-effect collision using loop-detector traffic and historical crash data. Within 5 minutes prior to the colliding, the data from Wuhan's urban expressway was aggregated on the similarity of traffic conditions for the use of traffic investigations leading to accidents.

**Lin Hu, (2020),** Traffic accidents are often associated with a driver’s conduct that is mainly determined by his characteristics. On the basis of the China In-Depth Accident Study (CIDAS), age, driving experience and driving style were statistically examined to examine the association between traffic accident risk and driver. The Grey cluster assessment classified drivers into the four risk ranks of accident accidents: low, medium to low, medium to high and high.

**Angus Eugene Retallack, (2019),** In addition to many other benefits, the need to understand the effects of road congestions on traffic accidents is highlighted by the importance of mitigating congestion in reducing delays and the resulting lost economic productivity. There is still a strong debate in many studies on this relationship. Subtle differences in characteristics between locations of research and the large number of covariates but also limitations in the design and focus of past research can produce this lack of consensus.

**Wencheng Wang, (2018),** This study examined the influence of risk factors on the frequency of urban traffic accidents and examines whether traffic accidents are spatial or temporal/heterogeneity. The field of study will be divided into 100 TAZs, each of which will remove the frequency and attributes of urban traffic accidents. The T-FEEM and the T-FEEM are compared respectively with the linear regression model, the spatial lag model (SLM) and spatial error model. Data from 10 months of traffic accidents in the city of Guiyang, China, illustrated the methodologies proposed.
MohadFedder Musa, (2020), In developing effective strategies for reducing these fatal accidents, different risk factors, including road conditions, must be identified. In our public works and police databases, we identified these serious problems, consisting of 1067 instances of varying gravities, that occurred during the period 2008-2015 on Malaysian federal highways. Those records have been used to develop an orderly logistic accident severity regression model and nine variables have been analyzed.

3. Methods of crashes prediction analysis

In present scenario, numerous systems are developed by the researchers to predict the road accidents in urban areas [21]. Usually, the developed systems include four steps such as data acquisition, data denoising, optimization and classification. Detailed explanation about the overview of road accident prediction in urban areas is given below.

3.1 Data collection and pre-processing

In road accident prediction, a few datasets are available online. In most of the cases, the researchers collected the data from their own on the basis of context information, vehicle data, roadway information and accident information.

- **Context information:** Lighting and weather conditions, while the accident occurred.
- **Vehicle data:** Number of vehicles involved in the accident.
- **Roadway information:** Directions, pavement surface condition, pavement type, grade, number of lanes, and horizontal alignment.
- **Accident information:** contributing conditions like accident pattern and accident type.

3.2 Optimizing the collected data

It is essential to calculate the feature space from the denoised data before performing classification. Though, the high dimension feature vectors make classification more difficult in order to find the correct patterns from the denoised data. The classification performance is enhanced by selecting the active feature vectors from the extracted features. Furthermore, it diminishes the computational requirements and storage space of the classifier.

3.2.1. Particle swarm optimization

It is a meta-heuristic evolutionary algorithm or population-based algorithm, based on the behavior of fish schools or bird flocks. Similar to other approaches, the PSO is also set with the random population of solutions. PSO approach handles a population of particles, where every particle signifies a potential solution to an optimization issue. Compared to other approaches, every potential solution in PSO is associated with the random velocity.

3.2.2. Whale optimization algorithm

It is a meta-heuristic algorithm that mimics the Hump-back whale’s behaviors. This optimization algorithm starts by generating a random population of whales. Then, the generated whales search for the optimum prey location by applying any one of these approaches; bubble-net or encircling. Finally, the whale position is updated by calculating a random search agent rather than the best search agent.

3.3 Classification of data

After optimizing the data, classification is performed by using an effective classifier. A few common classifiers used in road accident prediction are listed below.

- Convolutional Neural Network (CNN)
Deep Neural Network (DNN)
Adaptive Network based Fuzzy Inference System (ANFIS)
Random forest, etc.

3.3.1. Convolutional neural network

The CNN is a regularized form of multi-layer perceptron, which is also referred as fully connected network. Generally, CNN comprises of three layers such as input, multiple hidden, and output layers. Typically, the hidden layers of CNN contain a series layers, which convolve with a dot product or other multiplication. In CNN classification approach, the convolutional layer is a key layer that extracts the optimal information from the denoised data. In addition to this, the convolutional layer enhances the process of extracting the features from denoised data with low noise interference. In CNN classification approach, the mapping process is indicated in equation (1).

\[
x^i = f_c (\sum_{j \in \mathcal{M}} x^{i-1} \times k_j + \theta^i)
\]

\[
x^i = f_p (\beta^i \text{down}(x^{i-1}) + \theta^i)
\]

3.3.2. Deep neural network

In DNN classifier, the denoised data is transmitted from input to the output layer without any looping function. The possibilities of a lack of value are considered very low in classification as one of the great advantages of DNN [22]. In unsupervised pre-training phase, the DNN classifier executes only one layer. During prediction time, DNN classification approach assigns a score \(f(x)\).

In addition, is denoted as a function used for computation that is mathematically denoted in equation (3).

\[
Z_j = x \omega_{ij} Z_i = \sum_i Z_j + b_j X_j = g(Z_i)
\]

The DNN is a hierarchical feature learning classification method that helps to examine the unknown input feature coherences. In DNN classifier, greedy layer wise training is used to derive the low-level features from the high-level features. Hence, the key concern in DNN is to manage the complex functions that represents the high-level features. In addition, the DNN classifier used sparse auto encoders to support the successive input layers [23-25]. The general formula of auto encoder is mathematically shown in equation (4).

\[
x = h_{\omega \beta}(x) \approx x
\]

3.3.3. Adaptive network based fuzzy inference system

It is a multi-layered neural network that inter-connects all the weights equally. It is also named as neuro fuzzy approach that includes the advantage of fuzzy logic and neural network. Initially, the learning procedure is performed on the extracted feature vectors. Then, the ANFIS classifiers basic rule is mathematically denoted in equation (5)

\[
Rules_i = a p(x_1^i) + b p(x_2^i) + c p(x_3^i) + f_i
\]

In layer three, every node is named as circle node that evaluates the strength of firing, which is stated in equation (6). Successively, in layer four, each node is named as square node that is mathematically denoted in equation (7).

\[
o_{3i} = w_i = \frac{w_i}{(w_{t1} + w_{t2})}, i = 1, 2
\]
\[ a_{ij} = w^i t_{i| Rules} j_i = 1.2 \] (7)

4. Results analysis from previous review

Due to the extreme 0.05 level of significance, the STATA 15.0 software was used to implement this method. In both models, all variables of traffic together with their multiplying combinations of interactions were taken as explicatory variables. All weather and traffic variables plus average speed and speed limit interaction were included in the best-equipped combination of collapses. The prediction performance was evaluated and the values of these measures were shown in Table 1.

**Table 1:** Estimation results of negative binomial model.

| Variables          | Rear-end collisions | Side-impact collisions |
|--------------------|---------------------|------------------------|
|                    | Mean                | P value                | Mean                | P value                |
| Intercept          | 11.4970             | 0.0000                 | 13.1129             | 0.0015                 |
| Average speed      | -0.1561             | 0.0080                 | -0.1236             | 0.0195                 |
| Volume             | 0.0027              | 0.0121                 | 0.0044              | 0.0431                 |
| Rain               | -1.4399             | 0.0000                 | -0.9787             | 0.0006                 |
| Speed limit        | -0.1917             | 0.0000                 | -0.2565             | 0.0007                 |
| Average speed*speed limit | 0.0028 | 0.0067 | 0.0026 | 0.0134 |
| AIC                | 490.61              |                        | 392.37              |                        |
| BIC                | 510.58              |                        | 412.66              |                        |
| MAD                | 1.42                |                        | 1.23                |                        |
| MSE                | 2.63                |                        | 3.45                |                        |
| \( R^2 \)          | 0.75                |                        | 0.63                |                        |

In the positive binomial model for a rear collision and a side-impact crash, the result shows a good prediction performance. The main variables of both models are average speed, traffic volume, meteorological limit and the interaction of average speed and limit. Table 1 shows that a positive sign (negative) is likely to cause more (less) car crashes for the variable component number. Certain variables affect both types of crashes similarly. For example, in the traffic volume parameter, the crash proportions in both models are positive and indicate a rise in the traffic volume that will lead to rear-end crashes and secondary crashes.

4.1 Analysis of Free-Flow Speeds of Different Vehicle Classes and Subclasses

One of the goals of this study is to investigate the differences between free flow and subclass. As shown in Table 2, the studied vehicles were divided into six classes and 14 subclasses. The estimated free flow rate was grouped into groups depending on the vehicle's class and subclass. The distribution of their mean free flow speeds of different vehicle classes shows in ascending order Figure 4.4.

**Figure 4.1:** The impact on the crash frequency of the interaction from average speed and velocity limit. (a) For the collision at the back end. (b) for a collision of side effects.
Analysis of Free-Flow Speeds on Arterial and Sub arterial Roads

The present study included two classes of roads (arterials and subarterials). Arterials, usually on continual routes, are these roads for traffic. Subarterials are mainly roads for continuous traffic, but offer a rather lower level of traffic than blood roads (IRC: 86, 1983). Freeflow data analyzes showed that the average overall freeflow rate on arterial routes was 54.6 km/h, compare to 42.7 km/h on subterranean roads. The class free-flow distributions on arterials and subarterials of vehicles in Chennai are shown in figure 4.5.
4.3 Analysis of Free-Flow Speeds in Urban and Suburban Areas

Vehicle classes have been analyzed for speed distribution on urban and suburban roads. Twenty of the urban areas were chosen and the other 4 were chosen in the 24 suburban areas. The distribution of free-flow speeds in city and suburban roads can be seen in Figure 4.6. The average free flow rate on suburban roads was 58.2 km/h, while urban roads were 50.9 km/h. The statistically significantly t-tests and F-tests have been used to monitor different media and variances on urban and suburban roads (Table 3).

4.4 Analysis of lane choice of vehicles under free-flow conditions

Under free flow conditions, lane change is declared to be minimum. The position of the lane has been significantly influenced by free-flow rates of vehicle classes. The four- and six-lane divided-road analysis of proportions of vehicle classes showed specific lane choice patterns.

![Figure 4.5: Box plot: Arterial and subarterial free-flow speed distribution](image)

A very large number of smaller and slower vehicles, such as double-wheelers and three-wheeled cars, have preferred a kerb lane on four roads (Figure 4.7), while large vehicles (busses, trucks, cars and LCVs) have preferred medium-range lanes. Speed is the priority for aggressive and bigger vehicles. The majority of these are therefore selected as median paths that have negligible friction impedance.

A large number of vehicles preferred the mid lane in all vehicle classes. For kernels or medium-sized lines, a greater number of slower speeding vehicles are chosen for the kerb lane (two and three-wheelers), whereas the median lane is chosen by a greater percentage of fast-moving vehicles.

![Figure 4.6: Box plot: Free-flow speed distribution in urban and suburban areas](image)
Table 3: Statistical tests for comparing free-flow distributions in urban and suburban areas.

| Class | Area type pair tested | Mean FFS (km/h) | F–test | t–test |
|-------|-----------------------|-----------------|--------|--------|
|       | (a) - (b)             | (a)             | F–stat | p–value | t–stat | p–value |
| 2W    | Suburban - Urban      | 54.30           | 1.826  | 0.000   | 13.816 | 0.000   |
| 3W    | Suburban - Urban      | 41.00           | 1.242  | 0.076   | 1.707  | 0.088   |
| Car   | Suburban - Urban      | 72.58           | 1.747  | 0.000   | 16.502 | 0.000   |
| LCV   | Suburban - Urban      | 57.82           | 1.091  | 0.329   | 6.034  | 0.000   |
| Bus   | Suburban - Urban      | 53.05           | 0.845  | 0.197   | 6.934  | 0.000   |
| Truck | Suburban - Urban      | 46.03           | 1.117  | 0.284   | 0.264  | 0.792   |

Figure 4.7: Proportion of vehicles on four-lane, four-lane roads.

Figure 4.8: Proportion of vehicles choosing different lanes in six-lane divided roads.

5. CONCLUSION

- Road accidents are currently the main cause of injury and death. There is an increase in the incidence of accidents across India. In order to diminish the roadway accidents, careful analysis need to be done to find out the factors that mainly leads to accidents.
- This review paper gives an overview of road accident prediction for urban roads under heterogeneous traffic conditions and also analysis the existing research works by means of method, advantage, dataset, limitation and performance measures. From the investigation, this review paper states that still more work need to be done on road accident prediction for attaining better outcome.
- This review paper assists the researchers to understand the state-of-the-art in road accident prediction for urban roads under heterogeneous traffic conditions and also motivates more meaningful research works in the future.

REFERENCES

[1] Yannis G, Dragomanovits A, Laiou A, Richter T, Ruhl S, La Torre F, Domenichini L, Graham D, Karathodorou N, Li H (2016). Use of accident prediction models in road safety management—an international inquiry. Transportation research procedia 14: 4257-4266.
[2] Galatito F, Catalano M, Shaikh N, McCormick E, Johnston R (2018). Advanced accident prediction models and impacts assessment. IET Intelligent Transport Systems 12(9): 1131-1141.
[3] Xiong X, Chen L, Liang J (2017). A new framework of vehicle collision prediction by combining SVM and HMM. IEEE Transactions on Intelligent Transportation Systems 19(3): 699-710.
[4] Abadi A, Rajabioun T, Ioannou PA (2014). Traffic flow prediction for road transportation networks with limited traffic data. IEEE transactions on intelligent transportation systems, 16(2): 653-662.
[5] La Torre F, Meocci M, Domenichini L, Branzi V, Paliotto A. (2019). Development of an accident prediction model for Italian freeways. Accident Analysis Prevention, 124: 1-11.
[6] Ravindran V, Viswanathan L, Rangaswamy S. (2016). A novel approach to automatic road-accident detection using machine vision techniques. International Journal of Advanced Computer Science and Applications, 7(11): 235-242.
[7] Kumar S, Toshniwal D. (2017). Severity analysis of powered two wheeler traffic accidents in Uttarakhand, India. European transport research review 9(2): 24.
[8] Hashimoto S, Yoshiki S, Saeki R, Mimura Y, Ando R, Nanba S. (2016). Development and application of traffic accident density estimation models using kernel density estimation. Journal of traffic and transportation engineering (English edition) 3(3): 262-270.
[9] Yan L, Zhang Y, He Y, Gao S, Zhu D, Ran B, Wu Q (2016). Hazardous traffic event detection using Markov Blanket and sequential minimal optimization (MB-SMO). Sensors 16(7): 1084.
[10] Alkheder S, Taamneh M, Taamneh S. (2017). “Severity prediction of traffic accident using an artificial neural network”, Journal of Forecasting 36(1): 100-108.
[11] Z. Christoforou, S. Cohen, and M. G. Karlaftis, “Identifying crash type propensity using real-time traffic data on freeways,” Journal of Safety Research, vol. 42, no. 1, pp. 43–50, 2011.
[12] T. Bhowmik, S. Yasmin, and N. Eluru, “Do we need multivariate modeling approaches to model crash frequency by crash types? A panel mixed approach to modeling crash frequency by crash types,” Analytic Methods in Accident Research, vol. 24, p. 100107, 2019.
[13] F. Chen, S. Chen, and X. Ma, “Crash frequency modeling using real-time environmental and traffic data and unbalanced panel data models,” International Journal of Environmental Research and Public Health, vol. 13, no. 6, p. 609, 2016.
[14] F. Chen, S. Chen, and X. Ma, “Analysis of hourly crash likelihood using unbalanced panel data mixed logit model and real-time driving environmental big data,” Journal of Safety Research, vol. 65, pp. 153–159, 2018.
[15] P. Chouhdary, M. Imprialou, N. R. Velaga, and A. Chouhdary, “Impacts of speed variations on freeway crashes by severity and vehicle type,” Accident Analysis & Prevention, vol. 121, pp. 213–222, 2018.
[16] Q. Shi, M. Abdel-aty, and R. Yu, “Multi-level Bayesian safety analysis with unprocessed automatic vehicle identification data for an urban expressway,” Accident Analysis & Prevention, vol. 88, pp. 68–76, 2016.
[17] R. Yu, X. Wang, K. Yang, and M. Abdel-aty, “Crash risk analysis for shanghai urban expressways: a Bayesian semiparametric modeling approach,” Accident Analysis & Prevention, vol. 95, pp. 495–502, 2016.
[18] R. Yu, M. Quddus, X. Wang, and K. Yang, “Impact of data aggregation approaches on the relationships between operating speed and traffic safety,” Accident Analysis & Prevention, vol. 120, pp. 304–310, 2018.
[19] L. Aarts and I. van Schagen, “Driving speed and the risk of road crashes: a review,” Accident Analysis & Prevention, vol. 38, no. 2, pp. 215–224, 2006.
[20] M. J. M. Imprialou, M. Quddus, and D. E. Pitfield, “Predicting the safety impact of a speed limit increase using conditionbased multivariate Poisson lognormal regression,”