

Road Accidents Analysis – A Survey

¹Disha Jain, ²Aditya Magarde, ³Saad Yunus Sait

ABSTRACT--In India, road injuries are one of the top four causes of death and health loss among persons of age group 15-49 years. Road accidents also lead to economic loss to the country. The aim of this survey is to investigate the factors leading to road accidents and hence analyse the various algorithms to predict the severity of an accident.

Keywords--road accident, factors causing accident, data mining techniques, severity

I. INTRODUCTION

In the recent years, road accidents have become one of the biggest problems in the world as per WHO. Not only they cause economic damages, they have also become more likely to cause fatality. A recent article suggested that there are almost 400 deaths everyday due to accidents in India [1]. This has caused real tension causing bills from 2016 getting passed from the upper house in 2018. This goes to justify seriousness with which accidents should be taken. In India, road injuries are one of the top four causes of death and health loss among persons of age group 15-49 years. As per reports suggest, total number of road accidents reported in 2016 are 4,80,652 causing injuries to 4,94,624 persons and claiming 1,50,785 lives in the country. The number of persons killed in road crash has increased by 3.2 per cent over the previous year [2]. The report suggested various factors and statistics about the accidents that occurred in India. There are similar reports published by other countries too that consider road accidents to be one of the major reasons for loss of life and injury to individuals.

These statistics and the work going on, to study and analyse accidents to prevent damage emphasize the enormity of the need for research in this area. The occurrence of accidents and their severity has been a prime avenue for research in the recent years. Scientists have studied and analysed accidents and their severities in great detail. It is found that the occurrence of accidents depends on certain factors such as the prevailing weather conditions or the condition of the driver or the type of vehicle etc. Also, the studies have aimed at identifying the severity of accidents and using the documented circumstances and statistics to identify the severity of accidents. Lot of research work has been done on identifying the most important factors behind an accident and study how accidents become severe depending on these factors.

The remainder of the paper consists of two sections. Section 2 investigates the factors causing the accidents, followed by a table of various methodologies used to find these factors. Section 3 consists of the analysis of

¹ Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Kanchipuram Dist. – 603203, dishajain1211@gmail.com

² Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Kanchipuram Dist. – 603203, adityamagarde26@gmail.com

³ Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Kanchipuram Dist. – 603203, saad.y@ktr.srmuniv.ac.in

various algorithms to predict the severity of the accident followed by another table which summarizes various algorithms and their accuracies.

II. FACTORS CAUSING ACCIDENTS

Accidents are never merely coincidental and there are always factors and certain prevailing conditions that ultimately lead to an accident. There has been lots of research on finding and assessing the factors that were actually responsible for an accident. Uses of machine learning algorithms and data mining techniques have been deployed in the past to find out the important factors. A. Pakghokar et al. [3] in 2010 used CART and logistic regression and discovered that that people without a driving license or a conditional driving license were more prone to accidents. Further, it was 52% more if they had not fastened the seat belts. Also, they discovered that there was a significant correlation between age and commitment to fasten the seat belt. Tao Gang et al. [4] used the rough set theory (RST) to discover that road factors had the highest influence on accident mortality, whereas others like person, vehicle and environment did not affect much. It was discovered that five factors including ways of transportation, type of traffic signal, years of driving, type of road and visibility count for 58% of the total degree of importance and are most influential. Further in 2016, S. Vasavi [5] used machine learning techniques in order to extract and discover hidden patterns in traffic accidents. The research provides a general study and lists important factors that are majorly responsible for an accident. Entropy measures information gain, and Gain ratio and Gini index are used to choose relevant attributes useful for performing analysis. The author discovered 20 factors such as place of accident, lighting conditions, road condition, weather etc. which played a crucial role in accidents. Ultimately, k-medoids algorithm was used to cluster and further derive important findings from the data. The highlighting point of the research was that the biggest contributing factor of an accident was the combination of human and vehicle fault (83%). Further the reasons when evaluated, suggested that rash driving was the biggest cause of accidents. Research has always suggested that humans and the speed of the vehicle are majorly responsible for the mishaps on the roads. Sachin Kumar et al. [6] tried to identify the role of the location and characterize the role of the roads in any accident. They used data mining techniques and studied the accidents in areas around Dehradun. They categorized accidents locations as Low frequency accident location (LFAL), High frequency accident location (HFAL) and Medium frequency accident location (MFAL). Research signified that hilly roads and roads inside markets lie in HFAL. Also, authors pointed out that curves, bends near agricultural land are dangerous. LFAL regions are colonies and areas near hospitals. Another interesting discovery was that most accidents occur in winters (especially November) and after the sunset (after 6pm). Gagandeep Kaur et al. [7] studied the accidents on highways. The authors surveyed roads in Punjab wherein they discovered that there are higher number of accidents on the state highways as compared to the normal roads and further they suggest that straight roads are majorly responsible for accidents on the highway whereas cross-junctions are the major cause of accidents in case of ordinary roads. This helps us understand the accident-prone location inside the city and on highways. Table 1 summarises the factors causing accidents.

Table 1: Summary of the research which finds the factors causing accidents.

S. No.	Paper	Methodology	Data Source	Summary
--------	-------	-------------	-------------	---------

1.	Alireza Pakgohar et al. [3]	Use of CART (Classification and Regression Trees) and logistic regression.	Road Accidents in Iran	Identified that people without license were more involved into accidents and not wearing a seat belt further promoted it.
2.	Tao Gang et al [4]	Rough Set Theory	2008 to year 2013 accident data from China	Highest level of influence on accident mortality comes from road factors while levels of influence from person, vehicle and environment on accident mortality don't vary too much.
3.	S. Vasavi [5]	K-medoids to form clusters Used a priori algorithm.	2013 data of major national highways that pass through Krishna district.	Biggest contributing factor - combination of human and vehicle fault (83%). Rash driving was the biggest cause of accidents.
4.	Sachin Kumar et al [6]	K-mean to derive different clusters. Next, performed association rule mining.	Emergency Management Research Institute (GVK-EMRI), Dehradun	Hilly roads, and roads inside markets more prone to accidents. Winter nights more accident prone.
5.	Mr. G. Kaur et al [7]	Data Mining using R.	(2012-2015) of State Highways and Ordinary District Roads from Punjab	Straight roads are major cause of accidents on the highway whereas cross-junctions are the major cause of accidents in case of ordinary roads.

III. SEVERITY OF THE ACCIDENTS

Loss of life is an irreparable loss and road accidents have been one of the major contributors to this. Road accidents can cause life time health problems in some cases depending on the type of accident. Road accidents can be of various severities: accident with no injury, accident with slight injury, accident with serious injury and accident with fatality. A lot of research has been going on lately in order to predict the possible accidents and what could be the severity of the accidents. While some researchers have focused their research on the severity prediction on the basis of single factor while others have taken into account multiple factors which can lead to the accident, hence detecting the severity on the basis of the accident. Others predict only the death toll due to

road accident. QIN Liyan et al. [8] used neural networks and genetic algorithm to predict the death toll due to the accidents. From the results, it can be seen that Genetic Algorithm – Back Propagation (GA-BP) prediction model has higher accuracy than the BP neural network. Maximum relative error of GA-BP algorithm is 7.49% while the BP algorithm is 13.37% making GA-BP a better choice for prediction model. Yu Rujun et al. [9] used RBF neural networks to predict the number of deaths and economy loss. They used lab sampled data as the training data and were able to predict the death toll with maximum relative error of 5.98% and maximum accuracy of 99.74%. Galal A. Ali et al. [10] used Artificial Neural Networks (ANN) in 2012 to predict the casualties during 2010-2014 using the Sudan accident dataset from 1991-2009 to train the model. The results show that the ANN model for prediction of casualty has max difference of 1.84% between the actual observed accidents and ANN forecast accidents with an accuracy of 98.16% in predicting casualties. C. Sugetha et al. [11] used Association Rule Mining and Clustering to study and analyse the influence of various factors that lead to road accident and performance of various algorithms to predict severity of the accident. They used Naïve Bayes, ZeroR, IBK (Instance-Bases learning with parameter k) implementing K-Nearest Neighbour and Random Forest algorithm to train the model using training data and predicted the severity of the accident using test data. The complete data set consisted of 2549 instances of accident with 16 attributes. The results shows that IBK algorithm has highest accuracy with 100% on discrete dataset and 99.63% on mixed dataset. ZeroR algorithm has the least accuracy while Naïve Bayes is slightly better than ZeroR. Random Forest works well on both discrete and mixed dataset with accuracy around 98%. Tadesse Kebede Bahiru et al. [12] made a comparative study of various data mining techniques to predict severity. They used J48, ID3, CART, and Naïve Bayes classifiers to predict the accident severity. UK traffic accident data of 3050 instances was taken for training the classifier. J48 algorithm had the highest accuracy of 96.30% followed by the CART algorithm with 96.22%. Naïve Bayes and ID3 had slightly less accuracy of 94.62% and 94.09% respectively. Table 2 summarises the research on severity/death prediction

Table 2: Summary of research on death/severity prediction.

S.No.	Paper	Data Source	Algorithm	Accuracy	Summary
1	QIN Liyan et al. [8]	1978-2004, China road accident data	GA-BP	92.51% (min)	Genetic Algorithm (GA-BP) is better and more accurate to predict death toll than BP algorithm.
			BP	86.63% (min)	
2	Yu rujun et al. [9]	1990 – 2006, Lab sampled data	RBF neural network	94.02% (min)	RBF neural network shows a variable accuracy of

					94.02% - 99.74%.
3	Galal A. Ali et al. [10]	1991-2009, Sudan road accident data	Artificial Neural Networks (ANN)	98.16% (min)	ANN prediction model shows that the maximum prediction accuracy is 99.85%
4	C. Sugetha et al. [11]	2016, Columbia road accident data	Naïve Bayes	88.54%	IBK algorithm has the highest accuracy while ZeroR has the least.
			Random Forest	98%	
			IBK	99.63%	
			ZeroR	87.81%	
5	Tadesse Kebede Bahiru et al. [12]	2016, Riyadh road accident data	ID3	96.30%	ID3 algorithm has the highest accuracy while J48 has the least.
			J48	94.09%	
			Naïve Bayes	94.62%	
			CART	96.22%	

IV. CONCLUSION

From the above survey, it can be concluded that:

1. Table 1 indicates strong connection between accidents and the factors such as type of roads and speed of driving or driver authorization. Research on factors of accident indicates that strong rules must be imposed to monitor speeds and they should be implemented even more strictly at certain hotspots such as highways or crossings inside marketplaces.
2. On comparing the various algorithms from Table 2 to predict the severity or death toll due to a road accident, we can see that the IBK[11] algorithm has the highest accuracy among all algorithms while BP-neural network[8] algorithm has least accuracy of 86.63%.

REFERENCES

1. Sanjay, PR, "400 deaths a day are forcing India to take car safety seriously" The Economic Times 10th January 2018. Published: www.economictimes.indiatimes.co/news/politics-and-nation/. Web.
2. Government of India, Ministry of Road Transport and Highways, Transport Research Wing. (2017). Road Accidents in India – 2017. Retrieved from www.indiaenvironmentportal.org.in

3. Alireza Pakgohar, Reza Sigari Tabrizi, Mohadeseh Khalili, Alireza Esmaeili, The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach, *Procedia Computer Science*, Volume 3, 2011, Pages 764-769, ISSN 1877-0509, <https://doi.org/10.1016/j.procs.2010.12.126>.
4. T. Gang, S. Huan-Sheng, Y. Yong-Gang and M. Jafari, "Cause Analysis of Traffic Accidents Based on Degrees of Attribute Importance of Rough Set," 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, 2015, pp. 1665-1669. doi:10.1109/UIC-ATC-ScalCom-CBDCCom-IoP.2015.303
5. Vasavi S., Mishra D., Azar A., Joshi A. Extracting Hidden Patterns Within Road Accident Data Using Machine Learning Techniques. *Information and Communication Technology. Advances in Intelligent Systems and Computing*, vol 625. Springer, Singapore
6. Kumar, S. & Toshniwal, D. J. *Mod. Transport.* (2016) 24: 62. <https://doi.org/10.1007/s40534-016-0095-5>
7. G. Kaur and E. H. Kaur, "Prediction of the cause of accident and accident prone location on roads using data mining techniques," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-7. doi: 10.1109/ICCCNT.2017.8204001
8. Q. Liyan and S. Chunfu, "Macro Prediction Model of Road Traffic Accident Based on Neural Network and Genetic Algorithm," *2009 Second International Conference on Intelligent Computation Technology and Automation*, Changsha, Hunan, 2009, pp. 354-357. doi: 10.1109/ICICTA.2009.93
9. R. Yu and X. Liu, "Study on Traffic Accidents Prediction Model Based on RBF Neural Network," *2010 2nd International Conference on Information Engineering and Computer Science*, Wuhan, 2010, pp. 1-4. doi: 10.1109/ICIECS.2010.5678126
10. Galal A. Ali, Awadalla Tayfour, "Characteristics and Prediction of Traffic Accident Casualties In Sudan Using Statistical Modeling and Artificial Neural Networks", *International Journal of Transportation Science and Technology*, Volume 1, Issue 4, 2012, Pages 305-317, ISSN 2046-0430, <https://doi.org/10.1260/2046-0430.1.4.305>.
11. C. Sugetha, L. Karunya, E. Prabhavathi and P. K. Sujatha, "Performance Evaluation Of Classifiers For Analysis Of Road Accidents," *2017 Ninth International Conference on Advanced Computing (ICoAC)*, Chennai, 2017, pp. 365-368. doi: 10.1109/ICoAC.2017.8441188
12. T.K.Bahiru, D. Kumar Singh and E.A. Tessfaw, "Comparative Study on Data Mining Classification Algorithms for Predicting Road Traffic Accident Severity," *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Coimbatore, 2018, pp. 1655-1660. doi: 10.1109/ICICCT.2018.8473265