IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

REVIEW OF ROAD ACCIDENT ANALYSIS USING MACHINE LEARNING

Maneesha Kumari[1], Prof. Vivek Rai[2]

[1]M.Tech, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India

[2]Professors, Dept. of CSE, B N College of Engineering & Technology, (AKTU), Lucknow, India

Abstract**—** The auto industry has made efforts to create safer vehicles, but traffic accidents are inevitable, as this review paper explains. If we develop accurate prediction models that are able to automatically classify the type of injury and severity of various traffic accidents, we might be able to identify patterns in dangerous crashes. These social and street mishap examples can be valuable to foster traffic security control approaches. We believe that measures should be based on scientific and objective surveys of the causes of accidents and the severity of injuries in order to achieve the greatest possible accident reduction effects with limited budgetary resources. The results of four machine learning paradigms used to model the severity of injuries sustained in traffic accidents are summarized in this paper. We considered brain networks prepared utilizing cross breed learning draws near, support vector machines, choice trees and a simultaneous half and half model including choice trees and brain organizations. The results of the experiment show that the hybrid decision tree-neural network approach performed better than the other individual machine learning paradigms.

Keywords**—** safer vehicles, traffic accidents, accurate prediction, identify patterns, machine learning.

I. Introduction

Traffic accidents' costs, both in terms of deaths and injuries, have a significant impact on society. Researchers have paid more attention in recent years to identifying factors that significantly influence the severity of driver injuries resulting from traffic accidents [29][30]. Researchers have investigated this issue using a variety of methods. The log-linear model, fuzzy ART maps, the neural network, and the formulation of nesting logic are just a few examples. Model traffic accident data records can be used to learn about the drivers' behavior, road conditions, and weather that were linked to various injury severity by using data mining techniques.

Better traffic safety control policies can be formulated with the assistance of this information. Roh and co. By contrasting a model that was originally developed by Peltzman based on out-of-sample forecasts with a model specified using directed graphs, [22] demonstrated how statistical methods based on directed graphs, constructed over data for the recent period, may be useful in modeling traffic fatalities. [23] In terms of root mean squared forecast error, the directed graphs model performed better than the Peltzman model. Ossenbruggen and others [With the intention of using these models to carry out a risk assessment of a specific region, 24] utilized a logistic regression model to identify statistically significant factors that predict the probabilities of crashes and injury crashes.

A site's land use activity, roadside design, traffic control device use, and traffic exposure all played roles in these models. Their research demonstrated that village sites are safer than residential or commercial ones. Abdalla and others 25] investigated the relationship between casualty frequency and accident distance from residence zones. As was to be expected, casualty rates were higher closer to residence zones, possibly due to increased exposure. Residents of relatively deprived areas had significantly higher casualty rates than residents of relatively affluent areas, according to the study. Miaou and others 26] examined four regression models' statistical properties: two Poisson regression models and two conventional linear regression models in terms of their capacity to model highway geometric design relationships and vehicle accidents.

Street and truck mishap information from the Expressway Security Data Framework (HSIS) have been utilized to outline the utilization and the impediments of these models. It was demonstrated that conventional linear regression models do not possess the distributional property necessary to adequately describe road-based vehicle accident events that are typically sporadic, random, discrete, and nonnegative. In contrast, most of the desirable statistical properties for building relationships are found in the Poisson regression models.
In this review paper section I contains the introduction, section II contains the literature review details, section III contains the details about algorithms, section IV contains the software and language details, and section V provide conclusion of this review paper.

II. LITERATURE REVIEW

Yang and co. used a neural network approach to identify safer driving patterns that are less likely to result in injuries or deaths in car crashes [17]. In order to reduce the dimensions of the data, they carried out the Cramer's V Coefficient test [18] in order to locate significant variables that result in injury. After that, they used a frequency-based data transformation method to convert categorical codes into numerical values. Using a Backpropagation (BP) neural network, they made use of the University of Alabama-developed Critical Analysis Reporting Environment (CARE) system. They obtained a set of controllable cause variables that are likely causing the injury during a crash by utilizing the interstate alcohol-related data from 1997 Alabama and further investigating the weights on the trained network. There were two classes of the target variable in their study: injury and non-injury, where fatalities were included in the injury class. They discovered that they could potentially reduce fatalities and injuries by up to 40% by controlling a single variable, such as the driving speed or the lighting.

Sohn and co. used data fusion, ensemble, and clustering to boost the accuracy of individual classifiers for two categories of road traffic accident severity—physical injury and property damage—[15]. Neural network and decision tree classifiers were utilized as individual classifiers. After dividing the dataset into subsets with a clustering algorithm, they used each subset of data to train the classifiers. They discovered that when the variation in the observations is relatively large, as it is in the Korean data on road traffic accidents, clustering-based classification works better.

Mussone, others utilized neural networks to investigate a car accident that took place at an intersection in Milan, Italy [12]. BP learning-based feed-forward MLP was their choice. Eight variables—day or night, traffic flows circulating in the intersection, number of virtual conflict points, number of real conflict points, type of intersection, type of accident, condition of the road surface, and weather—had ten input nodes in the model. The ratio of the number of accidents at a given intersection to the number of accidents at the most dangerous intersection was used to calculate the output node, which was known as an accident index. According to the findings, the nighttime intersections with no traffic signals have the highest accident index for pedestrians being run over.

Dia and co. based a multi-layered MLP neural network freeway incident detection model on real-world data [5]. They thought about the presentation of the brain network model and the episode recognition model in procedure on Melbourne's expressways. The outcomes demonstrated that a neural network model could outperform the currently in use model in terms of incident detection speed and dependability. They also discovered that model performance in that section of the freeway could significantly suffer if speed data were not provided at a station.

Shankar and other used a nested logic formulation to estimate the likelihood of an accident's severity based on the likelihood of an accident happening [14]. They discovered that if at least one driver did not use a restraint system at the time of the accident, there is a greater chance of evident injury, disabling injury, or death than there is of no evident injury.

Kim et al. developed a log-linear model to explain how driver characteristics and actions contributed to more severe injuries. They discovered that driving under the influence of alcohol or drugs and not wearing a seat belt significantly raise the risk of more severe accidents and injuries [8].

Abdel-Aty and co used crash databases from the Fatality Analysis Reporting System (FARS) that covered the years 1975 to 2000 to look at how the rise in registrations for Light Truck Vehicles (LTV) affected fatal angle collision trends in the US [1]. They looked into the number of annual fatalities caused by angle collisions and the configuration of the collision (car-car, car-LTV, car-LTV, and LTV-car). The results of time series modeling indicated that fatalities as a result of angle collisions will rise over the next ten years, and that this rise will be influenced by the anticipated overall increase in the proportion of LTVs in traffic.

Bedard and co. utilized multivariate logistic regression to identify the independent contribution of driver, crash, and vehicle characteristics to the fatality risk of drivers [3]. They discovered that reducing speed, reducing the number and severity of driver-side impacts, and increasing seatbelt use may reduce fatalities. To ascertain the connection between accident notification times and fatalities, Evanco carried out a multivariate population-based statistical analysis [6]. The study found that the length of time it takes to notify drivers of an accident is a significant factor in the number of fatalities resulting from collisions on rural roads.

Ossiander and others utilized Poisson regression to examine the relationship between the speed limit increase and the fatal crash rate (fatal crashes per vehicle mile traveled) [13]. In Washington State, they discovered that an increase in the speed limit was linked to a higher rate of fatal crashes and an increase in fatalities on freeways.

Abdel-Aty and co used crash databases from the Fatality Analysis Reporting System (FARS) that covered the years 1975 to 2000 to look at how the rise in registrations for Light Truck Vehicles (LTV) affected fatal angle collision trends in the US [1]. They looked into the number of annual fatalities caused by angle collisions and the configuration of the collision (car-car, car-LTV, car-LTV, and LTV-car). The results of time series modeling indicated that fatalities as a result of angle collisions will rise over the next ten years, and that this rise will be influenced by the anticipated overall increase in the proportion of LTVs in traffic.

Bedard and co. utilized multivariate logistic regression to identify the independent contribution of driver, crash, and vehicle characteristics to the fatality risk of drivers [3]. They discovered that reducing speed, reducing the number and severity of driver-side impacts, and increasing seatbelt use may reduce fatalities. To ascertain the connection between accident notification times and fatalities, Evanco carried out a multivariate population-based statistical analysis [6]. The study found that the length of time it takes to notify drivers of an accident is a significant factor in the number of fatalities resulting from collisions on rural roads. Ossiander and others utilized Poisson regression to examine the relationship between the speed limit increase and the fatal crash rate (fatal crashes per vehicle mile traveled) [13]. In Washington State, they discovered that an increase in the speed limit was linked to a higher rate of fatal crashes and an increase in fatalities on freeways.

III. ALGORITHM

- **k-means clustering algorithm**

The well-known clustering problem can be solved with one of the simplest unsupervised learning algorithms, k-means. The method follows a straightforward method for classifying a given data set using a predetermined number of clusters (assume k clusters). The primary objective is to identify k centers, one for each cluster. Because different locations result in different outcomes, these centers should be strategically placed. Therefore, placing them as far apart as possible is the best option. The next step involves associating each point in a given data set with the closest center. The first step is finished and an early group age is completed when no point is pending. We need to recalculate k new centroids as the barycenter of the clusters from the previous step at this point. A new binding needs to be made between the same data set points and the nearest new center once we have these k new centroids. There has been created a loop. Because of this circle we might see that the k communities change their area bit by bit until no more changes are finished or at the end of the day places move no more.

- **Logistic Regression**

Logistic regression is the regression analysis and dependent upon the variables is binary numbers i.e. (0s and 1s), All regression analysis, the logistic regression is a prediction analysis. Logistic regression is used to details about data and to graphically explain the relationship between dependent binary variable and more nominal, ordinal, interval independent variables. Sometimes logistic regressions are difficult to describe the statistics tools are easily conduct and analysis the datasets, then in others plain word are as it is display in the output.

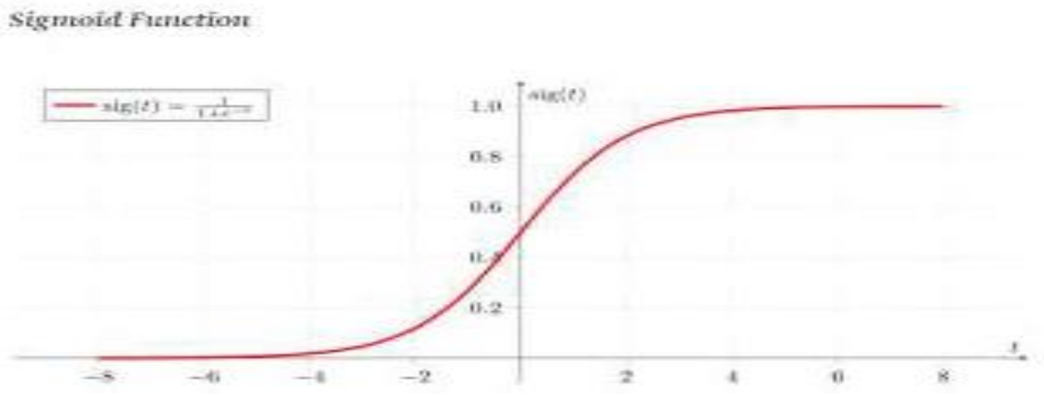

Figure 1: Signoid Function

IV. Software and Languages used

- **Jupyter**

Jupyter's mission is the creation of open-source software. It is utilized in dozens of programming languages for open standards and interactive computing services. It is a live document and code creation and sharing open-source web application. Which is a significant benefit of Jupyter? Cleaning and transforming data, numerical simulation, statistical modeling, machine learning, and many other applications are all possible with it. The algorithm was run with Jupyter.

• **Python**

Python is a high-level, interpreted programming language with a wide range of applications. Python, developed by Guido van Rossum and first released in 1991, places an emphasis on code readability through the use of a lot of whitespace in its design. It is currently the programming language that is used the most. It offers structures that make it possible to program clearly at both small and large scales. The system's logistic regression is carried out using jupyter, and the algorithm is written in python.

• **HTML, CSS, and JSCRIPT** are the web development languages that are utilized the most frequently. These programming languages were used to create the prediction system's user interface. The website serves as an interface, passing the various constraints entered by users to the program for it to work with.

V. CONCLUSION

Approaches to predicting the severity of drivers' injuries in head-on front impact point collisions are discussed in this review paper. There are many things that can lead to a road accident. After reading all of the research papers, it can be concluded that the types of vehicles, driver age, vehicle age, weather, road structure, and other factors have a significant impact on road accident cases. As a result, we have developed an application that uses the aforementioned factors to accurately predict road accidents.

REFERENCES

[1] Abdel-Aty, M., and Abdelwahab, H., Analysis and Prediction of Traffic Fatalities Resulting From Angle Collisions Including the Effect of Vehicles' Configuration and Compatibility. *Accident Analysis and Prevention*, 2003.
[2] Abdelwahab, H. T. and Abdel-Aty, M. A., Development of Artificial Neural Network Models to Predict Driver Injury Severity in Traffic Accidents at Signalized Intersections. *Transportation Research Record 1746*, Paper No. 01-2234.
[3] Bedard, M., Guyatt, G. H., Stones, M. J., & Hireds, J. P., The Independent Contribution of Driver, Crash, and Vehicle Characteristics to Driver Fatalities. *Accident analysis and Prevention*, Vol. 34, pp. 717-727, 2002.
[4] Buzeman, D. G., Viano, D. C., & Lovsund, P., Car Occupant Safety in Frontal Crashes: A Parameter Study of Vehicle Mass, Impact Speed, and Inherent Vehicle Protection. *Accident Analysis and Prevention*, Vol. 30, No. 6, pp. 713-722, 1998.
[5] Dia, H., & Rose, G., Development and Evaluation of Neural Network Freeway Incident Detection Models Using Field Data. *Transportation Research C*, Vol. 5, No. 5, 1997, pp. 313-331.
[6] Evanco, W. M., The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities. Accident Analysis and Prevention, Vol. 31, 1999, pp. 455-462.
[7] Hand, D., Mannila, H., & Smyth, P., Principles of Data Mining. The MIT Press, 2001.
[8] Kim, K., Nitz, L., Richardson, J., & Li, L., Personal and Behavioral Predictors of Automobile Crash and Injury Severity. *Accident Analysis and Prevention,*
Vol. 27, No. 4, 1995, pp. 469-481.
[9] Kweon, Y. J., & Kockelman, D. M., Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Severity Models. Accident Analysis and Prevention, Vol. 35, 2003, pp. 441-450.
[10] Martin, P. G., Crandall, J. R., & Pilkey, W. D., Injury Trends of Passenger Car Drivers In the USA. Accident Analysis and Prevention, Vol. 32, 2000, pp. 541-557.
[11] Mayhew, D. R., Ferguson, S. A., Desmond, K. J., & Simpson, G. M., Trends In Fatal Crashes Involving Female Drivers, 1975-1998. Accident Analysis and Prevention, Vol. 35, 2003, pp. 407-415.
[12] Mussone, L., Ferrari, A., & Oneta, M., An analysis of urban collisions using an artificial intelligence model. *Accident Analysis and Prevention*, Vol. 31,
1999, pp. 705-718.
[13] Ossiander, E. M., & Cummings, P., Freeway speed limits and Traffic Fatalities in Washington State. Accident Analysis and Prevention, Vol. 34, 2002,
pp. 13-18.
[14] Shankar, V., Mannering, F., & Barfield, W., Statistical Analysis of Accident Severity on Rural Freeways. *Accident Analysis and Prevention*, Vol.
28, No. 3, 1996, pp.391-401.
[15] Sohn, S. Y., & Lee, S. H., Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science*, Vol. 4, issue1, February 2003, pp. 1-14.
[16] Tavris, D. R., Kuhn, E. M, & Layde, P. M., Age and Gender Patterns In Motor Vehicle Crash injuries: Improtance of Type of Crash and Occupant Role. Accident Analysis and Prevention, Vol. 33, 2001, pp. 167-172.
[17] Yang, W.T., Chen, H. C., & Brown, D. B., Detecting Safer Driving Patterns By A Neural Network Approach. *ANNIE '99 for the Proceedings of Smart Engineering System Design Neural Network, Evolutionary Programming, Complex Systems and Data Mining*, Vol. 9, pp 839-844, Nov. 1999.
[18] Zembowicz, R. and Zytkow, J. M., 1996. From Contingency Tables to Various Forms of Knowledge in Database. Advances in knowledge Discovery and Data Mining, editors, Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R. AAAI Press/The MIT Press, pp.329-349.
[19] Abraham, A., Meta-Learning Evolutionary Artificial Neural Networks, Neurocomputing Journal, Elsevier Science, Netherlands, Vol. 56c, pp. 1-38, 2004.
[20] Moller, A.F., A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural
Networks, Volume (6), pp. 525-533, 1993.
[21] National Center for Statistics and Analysis http://www-nrd.nhtsa.dot.gov/departments/nrd-30/ncsa/NASS.html
[22] Roh J.W., Bessler D.A. and Gilbert R.F., Traffic fatalities, Peltzman's model, and directed graphs, Accident Analysis & Prevention, Volume 31, Issues 1-2, pp. 55-61, 1998.

[23] Peltzman, S., The effects of automobile safety regulation. Journal of Political Economy 83, pp. 677–725, 1975.
[24] Ossenbruggen, P.J., Pendharkar, J. and Ivan, J., Roadway safety in rural and small urbanized areas. Accid. Anal. Prev. 33 4, pp. 485–498, 2001.
[25] Abdalla, I.M., Robert, R., Derek, B. and McGuicagan, D.R.D., An investigation into the relationships between area social characteristics and road accident casualties. Accid. Anal. Prev. 29 5, pp. 583–593, 1997.
[26] Miaou, S.P. and Harry, L., Modeling vehicle accidents and highway geometric design relationships. Accid. Anal. Prev. 25 6, pp. 689–709, 1993.
[27] SVMlight. http://www.cs.cornell.edu/People/tj/svm_light/. Access date: May, 2003.
[28] Vapnik, V. N., The Nature of Statistical Learning Theory. Springer, 1995.
[29] Chong M., Abraham A., Paprzycki M., Traffic Accident Data Mining Using Machine Learning Paradigms, Fourth International Conference on Intelligent Systems Design and Applications (ISDA'04), Hungary, ISBN 9637154302, pp. 415- 420, 2004.
[30] Chong M., Abraham A., Paprzycki M., Traffic Accident Analysis Using Decision Trees and Neural Networks, IADIS International Conference on Applied Computing, Portugal, IADIS Press, Nuno Guimarães and Pedro Isaías (Eds.), ISBN: 9729894736, Volume 2, pp. 39-42, 2004.
[31] Eui-Hong (Sam) Han, Shashi Shekhar, Vipin Kumar, M. Ganesh, Jaideep Srivastava, Search Framework for Mining Classification Decision Trees, 1996. umn.edu/dept/users/kumar/dmclass.ps
[32] N. Cristianini and J. Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.
[33] Abraham, Intelligent Systems: Architectures and Perspectives, Recent Advances in Intelligent Paradigms and Applications, Abraham A., Jain L. and Kacprzyk J. (Eds.), Studies in Fuzziness and Soft Computing, Springer Verlag Germany, Chapter 1, pp. 1-35, 2002.