CAHIERS MAGELLANES-NS
Volume 06 Issue 2
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632574
http://magellanes.com/

MACHINE LEARNING IN TRAFFIC SAFETY: A REVIEW OF TECHNIQUES FOR INJURY SEVERITY PREDICTION

[1]Tanuj Nangia, [2]Umesh Sharma
[1]Ph.D. Scholar, Punjab Engineering College (Deemed to be University), Chandigarh, 160012, India, tanujnangia9@gmail.com
[2]Professor, Punjab Engineering College (Deemed to be University), Chandigarh, 160012, India, umesh1651@gmail.com

Abstract
Vehicle crashes are a significant cause of fatalities and serious injuries annually, presenting an ongoing public health problem on a global scale. In order to effectively address specific solutions, improve emergency response, and elevate road safety regulations, researchers must obtain an accurate evaluation of the intensity of collision injuries. This study explores differentmachine-learning techniques to evaluate their effectiveness in predicting the severity of injuriesin traffic crashes. The techniques include neural networks, support vector machines, decision trees, random forests, logistic regression, and gradient boosting. When it comes to road safety and the effects of crashes, having a comprehensive understanding of the degree of injuries sustained. This investigation of the literature analyzes the merits and disadvantages of variousmachine learning techniques to predict the severity of injuries sustained in a collision. Consequently, the generalized study showed that sophisticated models such as deep learning and ensemble analysis are more accurate in their predictions. Situations like this, on the other hand, are good for models such as logistic regression and decision trees because they are moreeasily interpretable and usable. As a result, difficulties related to feature selection, dealing withimbalanced data situations, and a lack of understanding of the offered models persist. The purpose is to raise awareness of how to use a range of multifaceted methods and combine knowledge about the domain with references to improve forecasts as well as develop more secure transportation systems.
Keywords: Machine Learning, Traffic Crashes, Road Safety, Deep Learning, Support Vector Machines, Neural Networks.

1.0. Preamble

According to information provided by the World Health Organization (WHO, 2023), every year, there are around 1.3 million fatalities and between 20 and 50 million injuries that do notresult in deaths that are caused by road traffic crashes. This results in disability for those affected individuals, leading to a three percent drop in economic output for most areas where this happens. This indicates that a significant proportion of VRUs, particularly bicycle riders,pedestrians, and motorcyclists, are involved in fatal crashes. They mostly occur amongst the most populous group of the population, those between the ages of five and twenty-nine years,and are mostly male as established by the WHO (WHO, 2023). Analyzing the factors denotingtraffic and population, it is possible to conclude that the number of traffic crashes worldwide and the people they affect has risen due to the increase in car usage and population growth. Itwas established that there are many factors affecting traffic collisions and these

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632574
http://magellanes.com/

factors dependon each other, which means that when explaining the impact of the crash characteristics, it is impossible to single out one of the characteristics and explain the overall picture of traffic collisions without referring to all the remaining characteristics.

Studies have demonstrated the effectiveness of utilizing KNN, RF, SVM, and DT algorithms in addressing the issue of crash severity prediction. According to a study conducted by (Santoset al., 2022), the four machine learning methods mentioned above are considered to be the most superior. In addition, developing and comparing the models for GB and XGBoost. The XGBoost algorithm is known for its high efficiency and flexibility (Chen & Guestrin, 2016). Since XGBoost is different from gradient tree boosting which is popularly known, this study also decided to use GB (Duncan et al., 1998). The severity of traffic collisions has been predicted in the past using various statistical modeling approaches. The ordered probit (OP) model (Kockelman & Kweon, 2002), ordered logit (OL) model (O'Donnell & Connor, 1996),

multinomial logit (ML) model (Khorashadi et al., 2005), and logistic regression LR model (AlGhamdi, 2002) are among the statistical models in consideration. Many machine learning algorithms have been created to simulate crash severity to overcome these constraints (Li et al., 2012) regression tree cart model (Karlaftis & Golias, 2002), and Artificial Neural Networks(ANN) (Karlaftis & Golias, 2002).

1.1. Utilization of Statistical Models for Predicting Crash Severity

The collision severity is determined by the use of statistical modeling, where the crash severityserves as the dependent variable. The surroundings, car, road, and driver each function as separate factors. Regression techniques, such as logit and probit, have been utilized to assess the magnitude of crash rates. (Kashani & Mohaymany, 2011) multinomial logit and probit were used in research to describe several severity levels of crashes, binary probit and logit models were employed in others. To take into consideration the ordinal character of variabilityand the link between variables in traffic crash data, several intricate models have been devised.The research of traffic crashes involves applying such models as probit models, ordered logit models, and nested logit models. This field also utilizes Bayesian hierarchical models as wellas their hybrids. The logistic regression model originated in the United States, and its usage isdirectly aimed at exploring and forecasting the interactions between the environment and the roads, as well as the severity of traffic crashes (Wang & Zhang, 2017). The severity of a crashis significantly influenced by the following factors: It examines aspects like the geographical location, the kind of road, the speed limit, road direction, lighting conditions, and the quality of pavement.

1.2. Implementation of Machine Learning Models in Predicting the Severity of Crashes Without relying on statistical assumptions, researchers have used a variety of machine-learningtechniques to forecast the severity of crashes in traffic. Machine learning algorithms can beemployed to forecast the complex correlation between the severity of crashes and associatedparameters. (Sameen & Pradhan, 2017) utilized recurrent neural networks (RNN) to predictthe extent of injuries resulting from vehicle crashes in Malaysia. The RNN approach achievedthe highest accuracy of 71.77%, while Bayesian logistic regression (BLS) attained an accuracyof 58.30% and multilayer perceptron neural networks (MLP) achieved an accuracy of 65.48%.(Abdel-Aty &

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632574
http://magellanes.com/

Abdelwahab, 2004) evaluated the effectiveness of both the OP and ANN modelsusing a similar methodology. The study employed fuzzy adaptive resonance theory (ART) and multilayer perceptron (MLP) to analyze the severity of crashes. According to the research,MLP fared better than ART, with a classification accuracy rate of 73.6% compared to ART's70.6%. The OP model's findings exhibited the lowest accuracy level, measuring 61.7%. Thecrash's severity was predicted using ANN, using six years of Abu Dhabi crash data. Theaccuracy of the Artificial Neural Network (ANN) surpassed that of the OP when their resultswere compared. (Alkheder et al., 2017) assessed the efficacy of pattern search (PS), geneticalgorithm (GA), combination genetic algorithm (CGA), and Artificial Neural Network (ANN)in forecasting the severity of traffic collisions (Kunt et al., 2011). The study's findings revealedthat the Artificial Neural Network (ANN) approach outperformed the other three techniques,with an R-value of 0.87. When comparing them, the Genetic Algorithm (GA) and PSprocedures got R-values of 0.79. The study conducted by (Li et al., 2018) investigated thevariables that influence the intensity of vehicular crashes in Washington, USA. This wasachieved by combining genetic algorithms (GA) with artificial neural networks (ANN).

To identify the crucial factors accurately, the contributors integrated the Artificial Neural Network (ANN) architecture with the Genetic Algorithm (GA) methodology. Driver behavior,vehicle action, age, driver restraint, and roadway condition had a substantial impact on the severity of the incident. A convolutional neural network (CNN) based on deep learning was used to forecast the magnitude of crashes in traffic. The efficacy of different statistical and machine learning methods, such as Logistic Regression (LR), Support Vector Machines (SVM), and Artificial Neural Networks (ANN), was assessed using Convolutional Neural Networks (CNN). According to the research conducted by (Zheng et al., 2019), the CNN model demonstrated outstanding performance compared to all other alternative methods.

Trauma centers can employ this information to precisely evaluate the severity of injuries, deploy a specialized emergency vehicle, and direct the vehicle to the closest hospital. The primary objective of modeling traffic collision injury severity is to consider the characteristics of the vehicle, the specifications of the route, and the surrounding environment. The objective of this research is to elucidate the potential improvement in the efficacy of machine learning models, specifically Artificial Neural Networks (ANN) and Support Vector Machines (SVM), through their integration with clustering techniques. It will set a new standard for building machine-learning models.

1.3. Literature Review

A review of traffic crash severity on a worldwide scale identifies numerous crucial characteristics that contribute to the seriousness of driving injuries. According to research conducted by (Casado-Sanz et al., 2020), the severity of driving injuries in Spain has risen as a result of reduced traffic volume and an increasing number of larger vehicles, especially on crosstown highways. The arrangement of roads also has a substantial influence. For instance,the introduction of diverging diamond interchanges (DDIs) at highway terminals in Missouri has led to a significant reduction of 55 percent in occurrences that simply caused damage to the property. Additionally, there has been a notable decrease of 31.4 percent in crashes resulting in fatalities and injuries (Claros et al., 2017). On the other hand, the number

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632574
http://magellanes.com/

of lanes,lack of lane markings, dangerous driving, and excessive speed have made driver injuries worse on cross-town roads in Spain (Casado-Sanz et al., 2020). Furthermore, the Lahore-Islamabad Highway M-2 has identified road sections with straight segments as significant elements that contribute to crashes (Iqbal et al., 2020).
The state of the road surface is another crucial aspect. As an illustration, the rates of single- vehicle and multi-vehicle collisions have risen by 8.87 and 2.82 times, respectively, when thehighways in Florida are wet (Zhang et al., 2013). The degree of crashes is highly associated with the condition of the road surface, as seen in Colombo, Sri Lanka (Dhananjaya & Alibuhtto, 2016). Weather conditions directly influence the severity of crashes. According to (Zhang et al., 2013), the presence of a barrier on urban expressways in Central Shanghai Cityis minimally affected by temperature, precipitation, and snowfall. In crash-prone areas like the Southern Expressway in Sri Lanka, rainy weather has a significant impact on crash frequency,according to (Edirisinghe & Edirisinghe, 2017). Conversely, dry weather has been identified as a primary contributing factor to the Lahore-Islamabad Highway M-2, as highlighted by (Iqbal et al., 2020).
The primary factor contributing to traffic crashes is driver behavior, with driver error accounting for the majority of incidents (77.1%) on Indian expressways (Shantajit et al., 2018).This comprehensive analysis underscores the importance of addressing factors such as road design, surface conditions, weather, and driver conduct to mitigate the severity of driving injuries on a global scale, highlighting the intricate and diverse aspects of traffic incident severity.

2.0. Conclusion

This literature review aims to provide a comprehensive analysis of the machine learning methods used to predict the severity of injuries sustained in collisions. The discussionemphasizes the benefits, drawbacks, and approaches to implementing each idea. The investigation reveals that the simplicity and interpretability of traditional algorithms like decision trees and logistic regression define them. While using them, one can achieve a considerably lower level of accuracy compared to the usage of more complex methods. Neuralnetworks, SVM, and random forest are some of the examples of machine learning that cannoteasily detected complex non-linear patterns due to which they are more accurate and durable.From the analysis of the results obtained from the study, the researchers failed to establish a machine learning method that excelled in the accuracy of predicting the severity of the crashesthat involve injuries. Some of the most effective prediction models are developed using multiple algorithms and domain specialists with rigorous validation. Future studies should focus on enhancing algorithm interpretability for the real-world application of these techniquesin the analysis and formulation of road safety policies indicating that further research should be directed toward the development and complexity of the algorithms, the analysis of the combined models, and the quality of the data.

References

[1] WHO.2023. Road traffic injuries [Online]. Available: https://www.who.int/news-room/ fact sheets/detail/road-traffic-injuries [Accessed 11 January 2023].
[2] Santos, K., Dias, J. P., & Amado, C. (2022). A literature review of machine learning algorithmsfor

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632574
http://magellanes.com/

crash injury severity prediction. Journal of Safety Research, 80, 254–269.
[3] Casado-Sanz, N., Guirao, B., & Attard, M. (2020). Analysis of the risk factors affecting the severity of traffic accidents on Spanish crosstown roads: The driver's perspective.
[4] Iqbal, A., Rehman, Z. U., Ali, S., Ullah, K., & Ghani, U. (2020). Road traffic accident analysisand identification of black spot locations on highway. Civil Engineering Journal, 6(12), 2448–2456.
[5] Zheng, M., Li, T., Zhu, R., Chen, J., Ma, Z., Tang, M., Cui, Z., & Wang, Z. (2019). Traffic accident's severity prediction: A deep-learning approach-based CNN network. IEEE Access, 7, 39897–39910.
[6] Li, Y., Ma, D., Zhu, M., Zeng, Z., & Wang, Y. (2018). Identification of significant factors in fatal-injury highway crashes using genetic algorithm and neural network. Accident Analysis & Prevention, 111, 354–363.
[7] Shantajit, T., Kumar, C. R., & Zahiruddin, Q. S. (2018). Road traffic accidents in India: an overview. International Journal of Clinical and Biomedical Research, 36–38.
[8] Wang, Y., & Zhang, W. (2017). Analysis of roadway and environmental factors affecting traffic crash severities. Transportation Research Procedia, 25, 2119–2125.
[9] Sameen, M. I., & Pradhan, B. (2017). Severity prediction of traffic accidents with recurrent neural networks. Applied Sciences, 7(6), 476.
[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] Claros, B., Edara, P., & Sun, C. (2017). When driving on the left side is safe: Safety of the diverging diamond interchange ramp terminals. Accident Analysis & Prevention, 100, 133– 142.
[12] Edirisinghe, E., & Edirisinghe, A. (2017). Analysis of accidents on the Southern Expressway.
[13] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 785–794.
[14] Dhananjaya, S. A. T., & Alibuhtto, M. C. (2016). Factors influencing road accidents in Sri Lanka: a logistic regression approach.
[15] Zhang, J., Chen, X., & Tu, Y. (2013). Environmental and traffic effects on incident frequency occurred on urban expressways. Procedia-Social and Behavioral Sciences, 96, 1366 1377.
[16] Li, Z., Liu, P., Wang, W., & Xu, C. (2012). Using support vector machine models for crash injury severity analysis. Accident Analysis & Prevention, 45, 478–486.
[17] Kashani, A. T., & Mohaymany, A. S. (2011). Analysis of the traffic injury severity on two lane, two-way rural roads based on classification tree models. Safety Science, 49(10), 1314– 1320.
[18] Kunt, M. M., Aghayan, I., & Noii, N. (2011). Prediction for traffic accident severity: comparing the artificial neural network, genetic algorithm, combined genetic algorithm and pattern search methods. Transport, 26(4), 353–366.
[19] Abdel-Aty, M. A., & Abdelwahab, H. T. (2004). Predicting injury severity levels in traffic crashes: a modeling comparison. Journal of Transportation Engineering, 130(2), 204 210.
[20] Kockelman, K. M., & Kweon, Y.-J. (2002). Driver injury severity: an application of ordered probit models. Accident Analysis & Prevention, 34(3), 313–321.
[21] Karlaftis, M. G., & Golias, I. (2002). Effects of road geometry and traffic volumes on rural

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632574
http://magellanes.com/

roadway accident rates. Accident Analysis & Prevention, 34(3), 357–365.

[22] Al-Ghamdi, A. S. (2002). Using logistic regression to estimate the influence of accident factorson accident severity. Accident Analysis & Prevention, 34(6), 729–741.

[23] Duncan, C. S., Khattak, A. J., & Council, F. M. (1998). Applying the ordered probit model to injury severity in truck-passenger car rear-end collisions. Transportation Research Record, 1635(1), 63–71.

[24] O'donnell, C. J., & Connor, D. H. (1996). Predicting the severity of motor vehicle accident injuries using models of ordered multiple choice. Accident Analysis & Prevention, 28(6), 739–753.