

Improved Accuracy of Calculation of Vehicle Crash Severity in Highways using Random Forest over Multi-Layer Perceptron Algorithm

Vignesh.S¹, Sashi rekha K²

¹Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.
²Project Guide, Corresponding Author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India. Pincode: 602105.

Abstract

Aim: To improve the accuracy rate of vehicle crash severity in highways using Random forest over MultiLayer Perceptron. **Materials and Methods:** Random forest and MultiLayer Perceptron with sample size of (N=10) is executed with varying training and testing splits for calculating the accuracy for accident crash severity with g power as 75%, threshold 0.000 and confidence interval 95%. The performance of the classifiers are evaluated based on their accuracy rate using accident severity dataset. **Results:** The accuracy for calculating accident crash severity in Random Forest(91%) and MultiLayer Perceptron (83%) is obtained(p<0.005). **Conclusion:** Prediction of accident crash severity using Novel Random Forest (RF) algorithm appears to be significantly better than MultiLayer Perceptron (MLP) with improved accuracy.

Keywords: Crash severity, Novel Random Forest Algorithm, Multi Layer Perceptron Algorithm, Machine Learning, Artificial Intelligence.

DOI:10.47750/pnr.2022.13.S04.181

INTRODUCTION

Accidents are incidents that happen randomly. The increase of vehicles in these years has also increased the number of accidents and deaths. To predict the severity of accidents and to prevent death and reduce injury level this method is implemented (Ahmadi et al. 2020). The importance of crash severity prediction in recent years has helped hospitals provide proper medical care as fast as possible when an accident occurs and also to amend road safety (Zhang et al. 2018). This research helps to predict the crash severity for transportation safety planners so that hospitals and agencies can provide emergency care with artificial intelligence (Iranitalab and Khattak 2017). Traffic accidents can be lessened by using it (Assi et al. 2020).

A lot of research has been performed on crash severity prediction using machine learning algorithm. 1761 research articles were published in Google Scholar and 37 articles were found in IEEE Xplore. In this work RF, Adaboost,GBDT,Logistic Regression is used to help the traffic management department to predict the accuracy of crashes with regarding to road infrastructure (Tang et al. 2019). This proposed system can extend the application for studies on the condition of environments in two way rural highways as it pertains to the impacts of crash severity using multiple Decision Trees (Abellán, López, and de Oña 2013). In this, various algorithms like ANN with 90% accuracy ,SVM with 89% and RF with 92%. During the journey we can check constantly for accidents happening via the probe vehicle based on location and speed, it can thus direct us to take secondary route to avoid furthermore accidents with help of artificial intelligence (Dogru and Subasi 2018).SVM model is used in this paper to evaluate the sensitive impact of explanatory variables to measure the severity of the crash (Li et al. 2012).The Random Forest Algorithm proved most efficient based on the accuracy and is used to predict the correct accident severity of the dataset (Geyik and Kara 2020)

Our institution is passionate about high quality evidence based research and has excelled in various fields (Parakh et al. 2020; Pham et al. 2021; Perumal, Antony, and Muthuramalingam 2021; Sathiyamoorthi et al. 2021; Devarajan et al. 2021; Dhanraj and Rajeshkumar 2021; Uganya, Radhika, and Vijayaraj 2021; Tesfaye Jule et al. 2021; Nandhini, Ezhilarasan, and Rajeshkumar 2020; Kamath et al. 2020). Based on the literature survey it can be concluded that the existing accident severity prediction model was not able to accurately determine severity of crash. In the proposed work the lack of accuracy in predicting the accident crash severity is analyzed and improved using machine learning algorithms such as Novel Random Forest algorithm and Multi Layer Perceptron Algorithm

Materials and Methods

The study setting of the proposed work is done in the Image Processing Lab, Department of Computer science and Engineering at Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (SIMATS). The number of groups used for the study is two group classification algorithms. The Group 1 is Novel Random Forest Algorithm and Group 2 is MultiLayer Perceptron Algorithm. Using clinical analysis (Kane, Phar, and BCPS n.d.), the analysis of sample size of $N = 10$ has been carried out with confidence of 95%. The input dataset is collected from kaggle.com (Accident Severity). It consists of two dataset namely test and train. Both the datasets have 17 attributes. Table 1 Contains the description of the dataset, the attribute which are present in the dataset incorporates like collision ref no, policing area, day, month, week, hour, weather conditions etc. Totally it consists of 1550 rows in the datasets which has null and duplicate values.

Random Forest Algorithm

Novel Random Forest Algorithm extensively used for both regression and classification problems. This algorithm can be used in various places such as banking, prediction works, health, stock markets, artificial intelligence etc. To get more accurate and stable estimates the random forest will create forest like structures and combines them. Subsets from both datasets and attributes are selected arbitrarily and gets trained. Using this method overfitting of data can be lowered. This algorithm takes lower training time than many other algorithms on large datasets with maintaining precision of the accuracy when a huge part of data is not present. Table 2 contains the pseudocode of Random Forest Algorithm

Multi Layer Perceptron Algorithm

It is a neural network model. It is used in classification and generalization. It works in only one direction like the data or the information goes through input node then hidden nodes and the exit node only this way thus it is an artificial feed forward network. Neural Network is just a mixture of perceptron joined in many ways with many activation functions. It produces output which forms a linear structure using the input height with real value inputs. It contains at least 3 layers of nodes. The output layer will become as the entrance layer for the next node thus it acts only as feed forward neural network. For training this uses a supervised learning technique known as backpropagation. First the perceptron is fitted and the classifier is used to test the scaled data. Thus the prediction is made. Table 3 contains the pseudocode of Multi Layer Perceptron

Statistical Analysis

The algorithms are run on a Personal Computer with 64-bit Operating System, 8GB RAM with steady internet connection and software such as Collab Notebook and python 3.9 for executing these algorithms. The independent variables are junction detail, junction control whereas the dependent variables are policing area, week day of collision, hour of collision etc. The software used here for statistics is SPSS version 26. The data set is prepared by using 5 iterations each of both Random Forest and Multi layer Perceptron. The testing variables are accuracy and loss whereas the group ID is given as grouping.

Results

The mean accuracy and loss values using T-test for both the algorithms along with the standard deviation is shown in Table 6. Here RF and MLP classifiers are used. The performances of the classifiers are measured by accuracy value. The Dataset is split into training and testing data to find accuracy. Where Table 4 contains the accuracy for Random Forest classifier with $N = 5$ and Table 5 contains the accuracy for Multilayer Perceptron classifier with $N = 5$. Novel Random Forest Classifier input is taken from the datasets as a class form and gives the output accuracy of 91%.

MultiLayer Perceptron classifier input is taken as class form from the data set and gives the output accuracy of 71%. Table 8 shows the accuracy values for both Random forest and MultiLayer Perceptron algorithm. It can be seen that Random forest has given better results of accuracy than MultiLayer Perceptron. From Table 7 It can be observed that Random Forest has better significance value than MultiLayer Perceptron with a value of $p = 0.000$. Accuracy and the Loss for both the algorithms are presented in a bar graph in Fig. 1.

Discussion

Random forest as convincingly appears better than Multilayer Perceptron with improved accuracy as referred in sTable 6. The Novel Random Forest classifier shows some difference in terms of accuracy score, speed and performance in comparison with MultiLayer Perceptron. Comparison between each algorithm's performance in

terms of accuracy and loss to prove that Random Forest is the best overall.

The outcome of this paper is similar to this (Al-Moqri et al. 2020) where they found that Random Forest Algorithm with 94% accuracy shows better performance than Multi Layer Perceptron with 89%. Seventeen Prediction attributes were used in this experiment.(Tasios, Tjortjis, and Gregoriades 2019) in this work one of the attributes is selected as class attribute and the attribute as predictive ones. Random Forest has shown more accuracy consistently than MultiLayer Perceptron.Hence Random Forest can be used to predict accident crash severity efficiently. (Wahab and Jiang 2019) In this work the Random Forest model showed the best prediction performance and surpassed the traditional MNLM methods. (Kumeda et al. 2019) carried out the work using Random Forest, Fuzzy-FARCHD algorithm, Multilayer Perceptron etc. The data was collected from UK road accidents in the year 2016. A total of 555 records were mined to get further insights. The experiment showed that the Fuzzy-FARCHD algorithm gave the best accuracy with 85%.

Most of the previous work done is based on the data from the specific dataset. In the proposed work the lack of accuracy in predicting the accident severity is improved by machine learning algorithms.In future more attributes in the dataset can be included and images can also be used as datasets for better accuracy. Artificial intelligence can also be used in future to predict the severity of the crash.

Conclusion

The Calculation of accuracy of accident crash severity in highways is carried out using the dataset obtained from Kaggle.It is done by Random Forest and MultiLayer Perceptron.The accuracy of Random Forest classifier is 92% and Multi Layer Perceptron 83% accuracy.The accuracy of predicting the accident crash severity is more in Random Forest than Multi Layer Perceptron.

Declarations

Conflict of interests

No conflict of interest in this manuscript.

Authors Contributions

Author SV was involved in data collection, data analysis, manuscript writing and Author SRK was involved in conceptualization, data validation, and critical review of manuscript.

Acknowledgment

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding: We thank the following organizations for providing financial support that enabled us to complete the study.

1. Soft Square Solutions, Palavakkam, Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

REFERENCES

1. Abellán, Joaquín, Griselda López, and Juan de Oña. 2013. "Analysis of Traffic Accident Severity Using Decision Rules via Decision Trees." *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2013.05.027>.
2. Ahmadi, Alidad, Arash Jahangiri, Vincent Berardi, and Sahar Ghanipoor Machiani. 2020. "Crash Severity Analysis of Rear-End Crashes in California Using Statistical and Machine Learning Classification Methods." *Journal of Transportation Safety & Security*. <https://doi.org/10.1080/19439962.2018.1505793>.
3. Al-Moqri, Tariq, Xiao Haijun, Jean Pierre Namahoro, Eshrak Naji Alfalahi, and Ibrahim Alwesabi. 2020. "Exploiting Machine Learning Algorithms for Predicting Crash Injury Severity in Yemen: Hospital Case Study." *Applied and Computational Mathematics*. <https://doi.org/10.11648/j.acm.20200905.12>.
4. Assi, Khaled, Syed Masiur Rahman, Umer Mansoor, and Nedat Ratrouf. 2020. "Predicting Crash Injury Severity with Machine Learning Algorithm Synergized with Clustering Technique: A Promising Protocol." *International Journal of Environmental Research and Public Health* 17 (15). <https://doi.org/10.3390/ijerph17155497>.
5. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. "Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine." *Energy & Fuels: An American Chemical Society Journal* 35 (12): 9930–36.
6. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. "Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea." *Journal of Nanomaterials* 2021 (July). <https://doi.org/10.1155/2021/8115585>.
7. Dogru, Nejd, and Abdulhamit Subasi. 2018. "Traffic Accident Detection Using Random Forest Classifier." 2018 15th Learning and

- Technology Conference (L&T). <https://doi.org/10.1109/lt.2018.8368509>.
8. Geyik, Buket, and Medine Kara. 2020. "Severity Prediction with Machine Learning Methods." 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA). <https://doi.org/10.1109/hora49412.2020.9152601>.
 9. Iranitalab, Amirfarokh, and Aemal Khattak. 2017. "Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction." *Accident; Analysis and Prevention* 108 (November): 27–36.
 10. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. "Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile." *Scientific Reports* 10 (1): 18179.
 11. Kane, Sean P., Phar, and BCPS. n.d. "Clinical Tools and Calculators for Medical Professionals - ClinCalc." Accessed October 10, 2021. <https://clincalc.com/>.
 12. Kumeda, Bulbula, Fengli Zhang, Fan Zhou, Sadiq Hussain, Ammar Almasri, and Maregu Assefa. 2019. "Classification of Road Traffic Accident Data Using Machine Learning Algorithms." 2019 IEEE 11th International Conference on Communication Software and Networks (ICCSN). <https://doi.org/10.1109/iccsn.2019.8905362>.
 13. Li, Zhibin, Pan Liu, Wei Wang, and Chengcheng Xu. 2012. "Using Support Vector Machine Models for Crash Injury Severity Analysis." *Accident; Analysis and Prevention* 45 (March): 478–86.
 14. Nandhini, Joseph T., Devaraj Ezhilarasan, and Shanmugam Rajeshkumar. 2020. "An Ecofriendly Synthesized Gold Nanoparticles Induces Cytotoxicity via Apoptosis in HepG2 Cells." *Environmental Toxicology*, August. <https://doi.org/10.1002/tox.23007>.
 15. Parakh, Mayank K., Shriram Ulaganambi, Nisha Ashifa, Reshma Premkumar, and Amit L. Jain. 2020. "Oral Potentially Malignant Disorders: Clinical Diagnosis and Current Screening Aids: A Narrative Review." *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation* 29 (1): 65–72.
 16. Perumal, Karthikeyan, Joseph Antony, and Subagunasekar Muthuramalingam. 2021. "Heavy Metal Pollutants and Their Spatial Distribution in Surface Sediments from Thondi Coast, Palk Bay, South India." *Environmental Sciences Europe* 33 (1). <https://doi.org/10.1186/s12302-021-00501-2>.
 17. Pham, Quoc Hoa, Supat Chupradit, Gunawan Widjaja, Muataz S. Alhassan, Rustem Magizov, Yasser Fakri Mustafa, Aravindhan Surendar, Amirzhan Kassenov, Zeinab Arzehgar, and Wanich Suksatan. 2021. "The Effects of Ni or Nb Additions on the Relaxation Behavior of Zr55Cu35Al10 Metallic Glass." *Materials Today Communications* 29 (December): 102909.
 18. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." *Environmental Progress & Sustainable Energy* 40 (6). <https://doi.org/10.1002/ep.13696>.
 19. Tang, Jinjun, Jian Liang, Chunyang Han, Zhibin Li, and Helai Huang. 2019. "Crash Injury Severity Analysis Using a Two-Layer Stacking Framework." *Accident; Analysis and Prevention* 122 (January): 226–38.
 20. Tasios, Dimitrios, Christos Tjortjis, and Andreas Gregoriades. 2019. "Mining Traffic Accident Data for Hazard Causality Analysis." 2019 4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM). <https://doi.org/10.1109/seeda-cecnsm.2019.8908346>.
 21. Tesfaye Jule, Leta, Krishnaraj Ramaswamy, Nagaraj Nagaprasad, Vigneshwaran Shanmugam, and Venkataraman Vignesh. 2021. "Design and Analysis of Serial Drilled Hole in Composite Material." *Materials Today: Proceedings* 45 (January): 5759–63.
 22. Uganya, G., Radhika, and N. Vijayaraj. 2021. "A Survey on Internet of Things: Applications, Recent Issues, Attacks, and Security Mechanisms." *Journal of Circuits Systems and Computers* 30 (05): 2130006.
 23. Wahab, Lukuman, and Haobin Jiang. 2019. "A Comparative Study on Machine Learning Based Algorithms for Prediction of Motorcycle Crash Severity." *PloS One* 14 (4): e0214966.
 24. Zhang, Jian, Zhibin Li, Ziyuan Pu, and Chengcheng Xu. 2018. "Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods." *IEEE Access*. <https://doi.org/10.1109/access.2018.2874979>.

Tables and Figures

Table 1. Dataset Description

Column	Values(For categorical variables)	Type
Collision_Ref_No	Multiple Values Present	String, Categorical
Weekday_of_Collision	Multiple days present	String, Categorical
Special_conditoin_at_site	1 (Yes), 0 (No)	Numerical, Categorical
Ped_Crossing_HC	1 (Yes), 0 (No)	Numeric, Categorical
Policing_Area	Multiple Values present	String, Categorical
Hour_of_Collision	Multiple Values	Numeric, Categorical

Table 2. Pseudocode for Random Forest Algorithm

Input
- Accident-severity (train and Test) Dataset
1. Initialization // input_attributes
2. The Random Forest Classifier is fit on the training set
3. Predict the records based on scaled values.
4. The features are then trained, tested and summarized
5. Predict the precision score
Output: Prediction of Accuracy

Table 3. Pseudocode for MultiLayer Perceptron

Input
- Accident-severity (train and Test) Dataset
1. Initialization // input_attributes
2. Import perceptron classifier
3. The Multi Layer Perceptron Classifier is fit on the training set
4. Predict the records based on scaled values
5. precision values scores are predicted

Table 4. Random Forest Accuracy and Loss for N = 5

Iterations	Accuracy(%)	Loss(%)
1	91.70	8.30
2	92.10	7.90
3	91.65	8.35
4	91.65	8.35
5	92.24	7.76

Table 5. MultiLayer Perceptron Accuracy and Loss for N = 5

Iterations	Accuracy(%)	Loss(%)
1	83.01	16.99
2	83.50	16.5
3	82.71	17.29
4	83.07	16.93
5	83.71	16.29

Table 6. T-Test Group Statistics with Mean, Std.Deviation, Std.Error Mean and Confidence = 95%

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Random Forest	5	91.9220	.21557	.09641
	Multi Layer Perceptron	5	83.2000	.40100	.17933
Loss	Random Forest	5	8.0780	.21557	.09641
	Multi Layer Perceptron	5	16.8000	.40100	.17933

Table 7. Independent Sample T-Test is applied for the data set fixing confidence interval as 95% and level of significance as 0.05

		Levene's test for equality of variances		T-test for equality means with 95% confidence interval						
		f	Sig.	t	df	Sig. (2-tailed)	Mean difference	Std.Error difference	Lower	Upper
Accuracy	Equal variances assumed	3.202	.111	42.838	8	.000	8.7220	.20360	8.25259	9.19151
	Equal Variances not assumed			42.838	6.134	.000	8.7220	.20360	8.22642	9.19151
Loss	Equal variances assumed	3.202	.111	-42.838	8	.000	-8.72200	.20360	-9.19151	-8.25259
	Equal variances not assumed			-42.838	6.134	.000	-8.72200	.20360	-9.21758	-8.25259

Table 8. Comparison of the Random Tree Algorithm and Multilayer Perceptron Algorithm with their accuracy

CLASSIFIER	ACCURACY(%)	LOSS(%)
Random Forest	91.70%	8.30%

Multi Layer Perceptron	83.01%	16.99%
------------------------	--------	--------

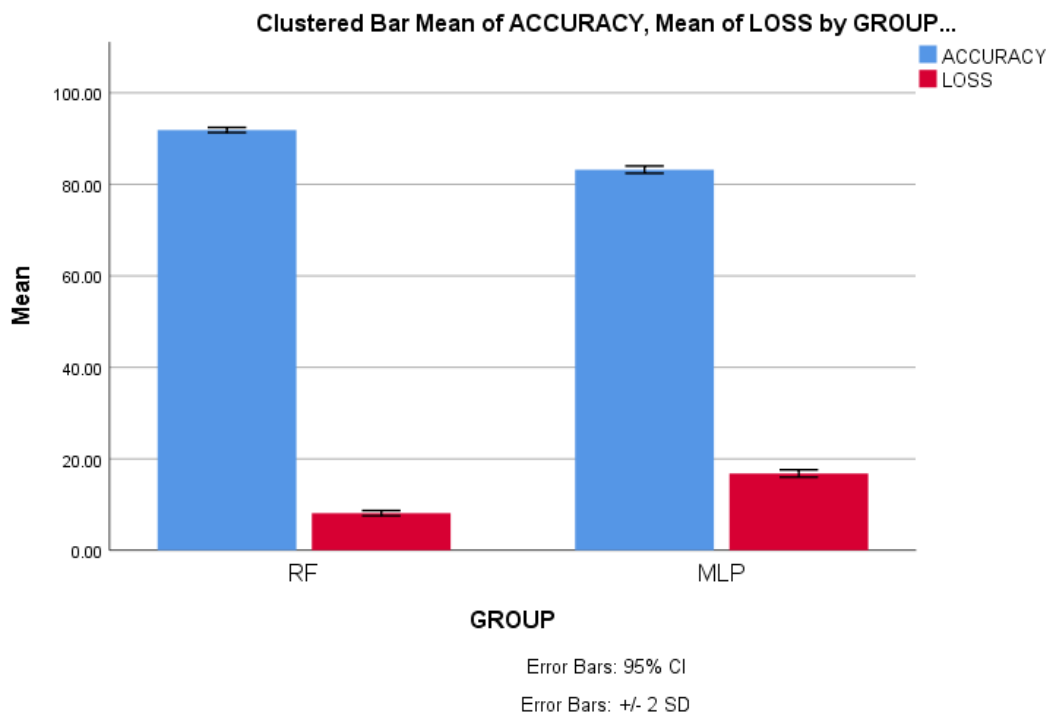


Fig. 1. Comparison of Random Forest algorithm And MultiLayer Perceptron in terms of mean accuracy. The mean accuracy of Random Forest is better than MultiLayer Perceptron. X axis(Groups): RF VS MIP algorithm, Y axis: Mean accuracy of prediction +/- 2SD