Original Article

Improved Accuracy of Calculation of Vehicle Crash Severity in Highways using Random Forest over Decision Tree Algorithm

Vignesh.S [1], Sashi rekha K[2*]

[2]Research 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.
[2*]Project Guide, 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 Decision tree. **Materials and Methods:** Random forest and Decision tree 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 Decision Tree (85%) is obtained($P<0.005$).**Conclusion:** Prediction of accident crash severity using Random Forest (RF) algorithm appears to be significantly better than Decision Tree (DT) with improved accuracy.

Keywords: crash severity , Novel Random Forest Algorithm, Decision Tree Algorithm , Machine Learning, Artificial Intelligence,

DOI:10.47750/pnr.2022.13.S04.176

INTRODUCTION

Accidents are naturally unspecified happening incidents.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).It is useful in reducing the repercussion of traffic crashes.(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 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 be extend the application for studies on the condition of environments in two way rural highways because of the concerning 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% has been used. During travel we can check constantly for crashes occurring via the probe vehicle based on location and speed it can direct to take secondary route to avoid furthermore accidents with assist 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 terms of accuracy. 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 Decision Tree 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 Decision Tree 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

Decision Tree Algorithm

This algorithm is used in predictive modeling which can be used for statistics,data mining, artificial intelligence etc.It will be like branches where the target variables are done for observations of a pattern.If the outcome has target variable or any classification values are known as classification trees.If the values are uninterrupted then they are called regression trees.Then further this tree is dividing the subsets as rote node and then to establish successor child which will occur recursively.Thus the prediction is made.

$$\boldsymbol{Entropy} = \sum_{x \epsilon X} -\boldsymbol{p}(\boldsymbol{a})\boldsymbol{E}(\boldsymbol{a}) \qquad (\mathbf{1})$$

$$\boldsymbol{Information\ Gain} = \boldsymbol{Gain}(\boldsymbol{A},\boldsymbol{S}) = \boldsymbol{E}(\boldsymbol{S}) - \sum_{a \epsilon T} \boldsymbol{p}(\boldsymbol{a})\boldsymbol{E}(\boldsymbol{a}) \qquad (\mathbf{2})$$

Here in (1) and (2) S is the dataset, class set is given by X, ratio of elements in x class is represented by p, the subset of data created from database S is represented by T. Table 3 contains the pseudocode of Decision Tree Algorithm.

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 variable 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 Decision Tree. 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 DT 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 4contains the accuracy for Random Forest classifier with N = 5 and Table 5 contains the accuracy for Decision Tree. classifier with N = 5

Random Forest Classifier input is taken from the datasets as a class form and gives the output accuracy of 91%. Decision tree classifier input is taken as class form from the data set and gives the output accuracy of 85%.Table 8 shows the accuracy values for both Random forest and Decision tree algorithm.It can be seen that Random forest has given better results of accuracy than Decision Tree. From Table 7 It can be observed that Random Forest has better significance value than Decision Tree 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 Decision Tree with improved accuracy as referred in Table 6. The Novel Random Forest classifier shows some difference in terms of accuracy score speed and performance in comparison with Decision Tree.

Comparison between each algorithm's performance in terms of accuracy, and loss proved that Random Forest is the best overall. The outcome of this paper is similar to (Nour et al. 2020) where they found that Random Forest Algorithm shows better performance with 73% accuracy than Decision Tree. Hence Random Forest can be used to predict accident crash severity efficiently.(Elyassami, Hamid, and Habuza 2020)In this work the experiment was carried out using regressive testing of all possible variations that are there for the given models. The Random Forest method involved the use of various tree depth levels along with the use of confidence vote and Gini index criteria. It resulted in an accuracy of 72.98. In the proposed System it can be found that the staging framework has an enormous number of hyper parameters that are to be calibrated. The classification models can be enhanced inorder to improve the features of classification. (Wahab and Jiang 2019) In this work the Random Forest model showed the best prediction performance and surpassed the traditional MNLM methods. (Chen and Chen 2020) paper showed that the results produced by Random forest in terms of accuracy and specificity are better compared to that of other machine learning algorithms. The study was done using 15 original values and 10 significant variables. Similarly in (Ji and Levinson 2020) this work RF shows best performance for predicting the severity using its iterative improving algorithms.

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 Decision Tree.The accuracy of Random Forest classifier is 92% and Decision Tree with 85% accuracy.The accuracy of predicting the accident crash severity is more in Novel Random Forest than Decision Tree.

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. Assi, Khaled, Syed Masiur Rahman, Umer Mansoor, and Nedal Ratrout. 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.
4. Chen, Mu-Ming, and Mu-Chen Chen. 2020. "Modeling Road Accident Severity with Comparisons of Logistic Regression, Decision Tree and Random Forest." *Information*. https://doi.org/10.3390/info11050270.
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, Nejdet, 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. Elyassami, Sanaa, Yasir Hamid, and Tetiana Habuza. 2020. "Road Crashes Analysis and Prediction Using Gradient Boosted and Random Forest Trees." *2020 6th IEEE Congress on Information Science and Technology (CiSt)*. https://doi.org/10.1109/cist49399.2021.9357298.
9. 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.
10. Iranitalab, Amirfarrokh, and Aemal Khattak. 2017. "Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction." *Accident; Analysis and Prevention* 108 (November): 27–36.
11. Ji, Ang, and David Levinson. 2020. "Injury Severity Prediction From Two-Vehicle Crash Mechanisms With Machine Learning and Ensemble Models." *IEEE Open Journal of Intelligent Transportation Systems*. https://doi.org/10.1109/ojits.2020.3033523.
12. 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.
13. Kane, Sean P., Phar, and BCPS. n.d. "Clinical Tools and Calculators for Medical Professionals - ClinCalc." Accessed October 10, 2021. https://clincalc.com/.
14. 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.
15. 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.
16. Nour, Mohamed K., Atif Naseer, Basem Alkazemi, and Muhammad Abid. 2020. "Road Traffic Accidents Injury Data Analytics." *International Journal of Advanced Computer Science and Applications*. https://doi.org/10.14569/ijacsa.2020.0111287.
17. Parakh, Mayank K., Shriraam 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.
18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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 Decision Tree Algorithm

Input - Accident-severity (train and Test) Dataset
1. Initialization // input_attributes
2. Import Decision classifier

3. TheDecision tree Classifier is fit on the training set
4. Predict the records based on scaled values
5. precision values scores are predicted
Output: Prediction of Accuracy

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. Decision Tree Accuracy and Loss for N = 5

Iterations	Accuracy(%)	Loss(%)
1	85.49	14.51
2	85.68	14.32
3	85.31	14.69
4	84.45	14.55
5	85.50	14.55

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	91.8680	.28084	.12559
	Decision Tree	5	85.2860	.48531	.21704
Loss	Random Forest	5	8.1320	.28084	.12559
	Decision Tree	5	14.5240	.13297	.05946

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	.424	.533	26.248	8	.000	6.58200	.25076	6.00375	7.1602
	Equal Variances not assumed			26.248	6.409	.000	6.58200	.25076	5.9778	7.1862
Loss	Equal variances assumed	8.464	.020	-45.99	8	.000	-6.39200	.13896	-6.71244	-6.0715
	Equal variances not assumed			-45.99	5.708	.000	-6.39200	.13896	-6.73629	-6.0471

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

CLASSIFIER	**ACCURACY(%)**	**LOSS(%)**
Random Forest	91.70%	8.30%
Decision Tree	85.49%	14.51%

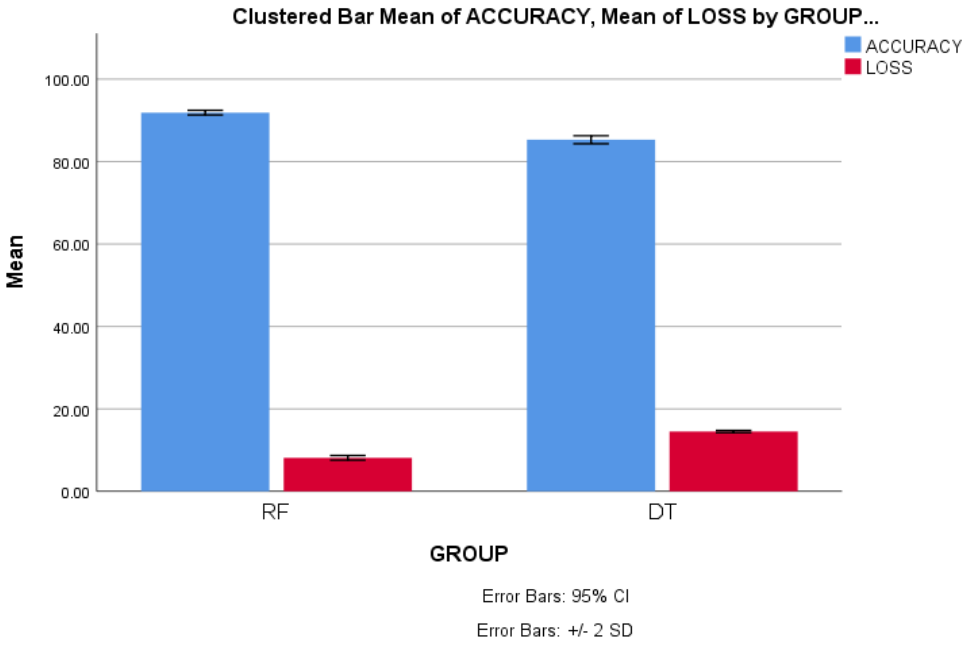

Fig. 1. Comparison of Random Forest algorithm And Decision Tree in terms of mean accuracy. The mean accuracy of Random Forest is better than Decision Tree. X axis(Groups): RF VS KNN algorithm, Y axis: Mean accuracy of prediction +/- 2SD