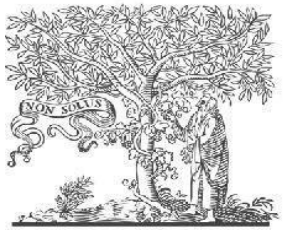
International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT



ELSEVIER SSRN

IJIEMR Transactions, online available on 8th Aug 2024. Link

https://ijiemr.org/downloads.php?vol=Volume-13& issue=issue08

DOI: 10.48047/IJIEMR/V13/ISSUE 08/11

Title A Structural Survey on Prediction of Traffic Accident Severity Using Machine Learning Models

Volume 13, ISSUE 08, Pages: 85 -91

Paper Authors
Dr. G.V. Ramesh Babu , Kampli Eekshitha

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper as Per UGC Guidelines We Are Providing A ElectronicBar code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

A Structural Survey on Prediction of Traffic Accident Severity Using Machine Learning Models

Dr. G.V. Ramesh Babu
Associate Professor, Department of Computer Science, Sri Venkateswara University, Tirupati
gvrameshbabu74@gmail.com
Kampli Eekshitha
Master of Computer Applications, Sri Venkateswara University, Tirupati.
eekshithak1203@gmail.com

Abstract
A leading cause of fatality is still traffic accidents on highways, despite improvements in traffic safety measures. The cost of the lives lost and the property damage caused by traffic accidents is very significant for developing countries. Many factors contribute to traffic accidents that are more crucial than others in determining how the collision will be. The primary elements that affect how bad an accident will be are distance, weather, wind chill, humidity, vision, and wind direction. This study suggests a combination of deep learning and machine learning to forecast the severity of traffic accidents.

Keywords: Road accidents severity, random forest, convolution neural network, feature importance, ensemble learning.

Introduction

Leading causes of injuries, fatalities, lifelong disability, and property damage include automobile accidents. It not only affects the economy but also affects the health care system because it puts a burden on the hospitals. Statistics shown by the ministry of public security of china from the years 2009 and 2011, traffic accidents caused an average of 65123 people to lose their life and 255540 got injuries annually. Identification of primary factors affecting road accident severity is required to minimize the level of accidental severity. Accidental Severity does not happen by chance; there are patterns that can be predicted and prevented. Accidental occurrences can be studied and prevented. Accident severity prediction is a significant topic in accident management**.** Accidental severity analysis involves the factors like the number of injuries, the number of casualties, and the destruction of property. The authors take the severity level independently and consider four options, light injury, severe injury, fatal, and property damage. The accidental severity level had classified as injury, possible injury, and property damage. In the last two decades, accidental severity has been the most popular research area. For the classification of traffic accidents, researchers used various statistical methodologies. These methods aid in determining the root cause of traffic accidents. The mixed logit modeling approach, the logit model, and the ordered probit model are a few of the traditional statistical-based studies. Nevertheless, these methods are unable to handle multidimensional datasets.

Prevention of road traffic accidents is a highly discussed issue that requires immediate action. Data related to severe accidents are needed, a comprehensive understanding of the circumstances that the cause of the accident is required, and systematic efforts are needed to to collect road traffic data. The health sector has the revocability to establish a better data-

collecting system and share the data with multiple audiences. Only reliable and authentic data can help to reduce traffic accidents and also help to find the severity level of the accident.

Scope:
In this research we can find important factors that affect the severity of an accident.

It will help to predict the traffic accident severity based on the factors influencing the traffic accident.

Purpose
One of the leading cause of deaths in any country is traffic accident. Traffic accidents creates a lot of socio-economic damages to a country. Traffic accidents can't be stopped but we can comprehend the traffic accidents severity based on the factors affecting the traffic accident.

Problem Statement: - The factors which cause traffic accidents, are bad road conditions, weather conditions, Vehicle Condition, Visibility and so on. Highways, Crossings and Junctions are becoming the accident spots these days. The other reasons for accidents are using mobiles, over speed, overtaking from wrong side and so on.

Comparative study about various Algorithms for Prediction of Traffic Accident Severity.

Synod	Author	Methodologies used	Merits	Demerits
1	Zhibin Li andZiyuan Pu	OP model, MNL model, Decision Tree, K-Nearest Neighbour, Random Forest and Support Vector machine	Crash severity models help hospitals provide fast medical care.	Machine learning methods estimate variable importance differently on ex-planatory variables.
2	Buket Geyik, Medine Kara	Multi-layer perceptron, Decision Tree, Random Forest Classifier and Naive Bayes	The accuracy scores of the algorithms can be improved by using a balanced dataset from another nation.	Accident rates increase, making data sets difficult to handle using conventional statistical methods.
3	Umer mansoor, Nedal t. Ratrout, Syed masiur rahman and khaled	KNN, DT, AdaBoost, SVM, and FNN	Models can be applied globally to other crash databases to improve crash severity prediction accuracy.	The two-layer ensemble model has higher running time than individual models.
4	Rabia Emhamed Al Mamlook, Tiba Zaki Abdulhameed,	Logistic Regression Algorithm, Decision Tree Algorithm, Light-	Age is the most important factor in determining	Aged drivers are more likely to be involved in

	Raed Hasan, Hasnaa Imad Al-Shaikhli, Ihab Mohammed, Shadha Tabatabai.	GBM Algorithm, Random Forest Algorithm and Naive Bayesian Algorithm	severity of fixed object crashes among older drivers.	crashes with severe injuries.
5	Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine	Decision Tree, AdaBoost, KNN, Naïve Byes	Ada-iterative Boost's classification on decision trees produces the best results.	Predictions are made using static techniques, so if any data is missed, the model won't produce the expected results.
6	Shakil Ahmed, Md Akbar Hossain, Md Mafijul Islam Bhuiyan	Logistic Regression, K-Nearest Neighbor, Naive Bayes , Random Forest, Adaptive Boosting and Extreme Gradient Boosting	Ensemble mode ML algorithms are better than single mode algorithms for reducing risk	This information can be used to train a model.
7	Zhibin Li, Pan Liu, Wei Wang, Chengcheng Xu	Support Vector Machine, OP model	The SVM model outperforms the OP model in predicting crash injury severity.	The freeway highway dataset was used to assess its potential, not to compare it to other statistical models.
8	Fang Zong, Hanggou Xu and Huiyng Zang	Bayesian Network, Regression Models	Bayesian networks are better suited for predicting accident severity than regression models.	Lack of data limits consideration of driver characteristics and traffic conditions.
9	Raj Biswas, John Oram, Dinesh Rao, Mehdi	XG Boost models, random forest (RF), and neural networks (NN).	Enabling safer route planning.	Two expressways in Shanghai are used to evaluate this methodology.
10	Miaomiao and Yindong Shen	Random Forest and Bayesian Optimization	models were compared with respect to the testing accuracy	Traffic accident data exhibit high dimensions, nonlinearity, and collinearity.
11	Maher Ibrahim Sameen and Biswajeet Pradhan	RNN, Long Short-Term Memory algorithm	RNN model to determine	More research to

			effects of factors.	get optimal result
12	Sivabrahmam Talloju, Suneel Kumar Duvvuri	KNN, Logistic Regression, Random Forest, Support Vector Machine, Multilayer Perceptron, Gradient Boost, and AdaBoost.	The Gradient Boost algorithm has surpassed all the other algorithms in comparison	The data in the excel form cant be run with out code
13	Puneeth raj N, Dr. Girish	KNN, Decision Tree, and Naive Bayes ,multi-layer perceptrons	will be able to gauge both the severity of traffic accidents and the contributing causes	Only two algorithms got high accuracy.
14	Mohammed Akour, Osama Al Qasem, Firas Hanandeh	Balanced Random Forest, XGBoost, and decision tree	voting appeared to perform better than stacking models and other individual classifier	Not all parameters may be present in dataset, limiting impact on accident severity.
15	Shiva Prasad Satla, Manchala Sadanandam, Buradagunta Suvarna	Logistic regression, Random forest and K-nearest neighbour	logistic regression had a strong accuracy score	Only provides instructions, how to drive

Literature Survey:-

1:- This work tested different machine learning and statistical models for their ability to accurately predict collision injury severity. The results showed that the RF and KNN were the best models with the highest overall forecasting accuracy, in line with Irani Talab and Khattak's conclusions that KNN and RF are the two most effective models.

2:- This study used machine learning methods to accurately fore-cast the levels of accident severity from the STATS19 dataset, using decision trees, random forests, Naive Bayes, and MLP algorithms. They discovered that the NAive Bayes algorithm performs better than others, and their work can be applied to motorcycle accidents.

3:- This work examined the effectiveness of multiple machine learning models for predicting the severity of traffic collisions using data from the Department of Transport, Great Britain. Three performance indicators (precision, recall, and F1 score) were developed to provide insight into the performance of models.

4:- The main factor causing crashes in elderly injury severity can be found utilising the suggested predictive model light-GBM. The Michigan Traffic Agencies will be able to tackle the increased risk of traffic injuries to elderly drivers with greater proactivity

thanks to this model. The most important factors, according to the results, are age and traffic volume.

5:- Machine learning techniques can be used to create a traffic accident prediction model with cutting-edge technology to reduce the number of accidents. Ada boost and Naive bayes were shown to have the best accuracy in this study. The findings showed that rush hour has a high accident rate, and accidents were even found to rise de-pending on the state of surface impact features.

6:- This paper compares three single mode classification algorithms and three ensemble mode classification algorithms to estimate the binary and multiclass injury severities. RF is the most suitable ML algorithm in this situation, as driving characteristics such as age, gender, alcohol and drug use have a substantial impact on the severity of road accident injuries. Additionally, ensemble mode ML algorithms function more effec-tively overall.

7:- Xu An SVM model was created to predict the severity of injuries related to specific collisions using crash data gathered at 326 motorway diverging zones. The same dataset was also used to create an OP model. They compared the SVM model's and the OP model's performance in terms of prediction. It was
discovered that the SVM model outperformed the OP model.

8:- In this , they applied bayesian network and regression model on records of traffic accidents reported to the police for the Chinese province of Jilin in 2010 dataset . According to the results Bayesian network outperformed the regression model.

9:- In this, Author has used tabular, several feature data types as input, to structure the road accident risk prediction as a binary classification problem. Here, they used a baseline model and trained XG Boost to get the best results, but it was unable to generalise to a new city without that city's data. To get around this, they used a shaply additive explanation (SHAP) value graph, which gives th significant features that cause traffic accidents on the data they had collected, and were able to predict the probability of accidents.

10:- In this, authors have used the hybrid between the Bayesian optimization and Random forest on traffic accidents in the US from February 2016 to March 2019 dataset. In terms of precision, F1 score and Accuracy it outperformed the SVM, KNN and ANN.

11:- Accident severity can be estimated as a pattern identification problem that can be resolved using deep learning, statistical methods, and occasionally physical modelling techniques.In this paper,they used the RNN model(Recurrent Neural Network).Input, LSTM, two dense layers, and a surtax layer are the five major layers of the RNN model. In comparison to the MLP and BLR models, the RNN model's validation accuracy score was the highest.

12:- This paper's major goal is to create a machine learning model that will help the RTA authority foresee the severity of incidents.They conducted research to determine what causes accidents to result in more fatalities. Utilized techniques include KNN, Logistic Regression, Random Forest, Support Vector Machine, Multilayer Perceptron, Gradient Boost, and AdaBoost. The Gradient Boost algorithm has surpassed all the other algorithms in comparison.

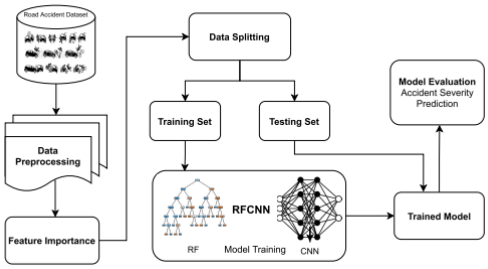
13:- The primary goal of this research is to assess auto accidents and rank their seriousness using a combination of deep learning and machine learning techniques. In this work, four of
the most sophisticated and well-known supervised learning algorithms in machine learning are employed to investigate traffic incidents. They are Decision Tree, KNN, Naive Bayes, and MLP techniques used in deep learning.Decision trees and MLP both outperformed the other model in this comparative model with high accuracy.

14:- In this research, machine learning (ML) approaches are used to forecast accident
severity.This study's primary objective is to assess ensemble algorithms' performance and compare it to that of individual classifiers.For building stacking and (hard and soft) voting models, the algorithms Balanced Random Forest, XGBoost, and decision tree were chosen. Out of all the methods, the voting (hard) ensemble technique yields the highest TNR.

15:- They predict the dangerous roads in this paper. The data from the previous dsecade is used to identify the two categories in this: dangerous and non-dangerous. They

used various machine learning models to make predictions, and compared to all other models, logistic regression had a strong accuracy score .

Proposed Model



AdaBoost is a short form of adaptive boosting, AdaBoost Classifier (AC) is an ensemble model based on decision trees, Extra Tree Classifier (ETC) uses a random subset of features to split nodes of trees, Gradient Boosting Machine (GBM) is based on boosting and a powerful ensemble model to perform classification, Convolutional Neural Network (CNN) is a deep neural network model that handles data complexity during computation. The working of RF is shown as follows: p = mode {T1(y), T2(y), Tm(y)} (1) p = mode(Xmm=1Tm(y)) (2) p is the final prediction, calculate by majority votes. The probability of accident is represented by p and given by Y = logit = lnP1 − P= βX. Stochastic gradient descent (SGD) is an iterative approach and optimizes objective function by selecting smoothness in terms of properties. An ensemble model is a machine learning model that works on the merging of two or more models and gets better performance than individual classifiers. This research utilizes the voting ensemble of two machine learning models that are LR and SGD to predict road accident severity. Hyperparameters values are presented.

References: -

[1] World Health Organization, Global Status Report on Road Safety 2015, World Health Org., Geneva, Switzerland, 2015.
[2] W. Gissane, "Accidents—A modern epidemic," J. Inst. Health Educ., vol. 3, no. 1, pp. 16–18, 1965.
[3] H. Manner and L. Wünsch-Ziegler, "Analyzing the severity of accidents on the German autobahn," Accident Anal. Prevention, vol. 57, pp. 40–48, Aug. 2013.
[4] J. C. Milton, V. N. Shankar, and F. L. Mannering, "Highway accident severities and the mixed logit model: An exploratory empirical analysis," Accident Anal. Prevention, vol. 40, no. 1, pp. 260–266, 2008
[5] N. V. Malyshkina and F. L. Mannering, "Markov switching multinomial logit model: An application to accident-injury severities," Accident Anal. Prevention, vol. 41, no. 4, pp. 829–838, Jul. 2009.
[6] K. Haleem, P. Alluri, and A. Gan, "Analyzing pedestrian crash injury severity at signalized and non-signalized locations," Accident Anal. Prevention, vol. 81, pp. 14–23, Aug. 2015.
[7] M. Bedard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, "The independent contribution of driver, crash, and vehicle characteristics to driver fatalities," Accid Anal. Prevention, vol. 34, no. 6, pp. 717–727, 2002.
[8] S. S. Zajac and J. N. Ivan, "Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural connecticut," Accident Anal. Prevention, vol. 35, no. 3, pp. 369–379, May 2003.
[9] W.-H. Chen and P. P. Jovanis, "Method for identifying factors contributing to driver-injury severity in traffic crashes," Transp. Res. Rec., J. Transp. Res. Board, vol. 1717, no. 1, pp. 1–9, Jan. 2000.
[10] S. Sarkar, S. Vinay, R. Raj, J. Maiti, and P. Mitra, "Application of optimized machine learning techniques for prediction of occupational accidents," Comput. Oper. Res., vol. 106, pp. 210–224, Jun. 2019.
[11] K. Rumar, "Transport safety visions, targets and strategies: Beyond 2000," First Eur. Transp. Saf. Lecture, Brussels, Eur. Transp. Saf. Council, pp. 6–8, 1999.
[12] T. Hummel, "Land use planning in safer transportation network planning," Leidschendam, Inst. Road Saf. Res., Citeseer, 2001.
[13] A. S. Edlin, Per-Mile Premiums for Auto Insurance. Cambridge, MA, USA: National Bureau of Economic Research, Dec. 1998. [Online]. Available: https://ssrn.com/abstract=142667, doi: 10.2139/ssrn.142667.

[14] R. Wittink, "Promotion of mobility and safety of vulnerable road users: Final report of the European research project promising (promotion of measures for vulnerable road users)," Inst. Road Saf. Res., Tech. Rep., 2001.
[15] D. Mohan and G. Tiwari, "Traffic safety in low-income countries: Issues and concerns regarding technology transfer from high-income countries," in Reflections of the Transfer of Traffic Safety Knowledge to Motorising Nations, vol. 56. Melbourne, Global Traffic Safety Trust, 1998