International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V70I9P106

Volume 70 Issue 9, 43-49, September 2022

Original Article

Prediction of Severity of an Accident Based on the Extent of Injury using Machine Learning

Surendra Kumar Reddy Koduru

Business Intelligence & Reporting Lead, NC, USA.

Received: 12 August 2022   Revised: 18 September 2022   Accepted: 28 September 2022   Published: 10 October 2022

Abstract - *Accidents are currently regarded as the most disturbing cause in many countries. Several deaths have been recorded for generating massive deaths throughout numerous countries just as a result of road accidents that predominantly occur during traffic. Vehicle accidents are the leading cause of deception, distress, and fatality. The majority of accidents occur only over a long period from various countries and are referred to as unsafe or dangerous conditions associated with large volumes of traffic, particularly vehicle traffic. Exploring the causes of these incidents can help identify the most important features in determining the accident's severity.*

Almost all the repercussions, such as light conditions, speed zones, part of the injury, climate, and so on, are also participating in and closely linked to the cause of traffic accidents, of which only a few are emphasized and addressed in accident criticality rules. The overall goal of this study is to measure the severity of traffic accidents that occur. The key directing vectors are the accident attributes, which include the part of the slight, car allocation on the highway, and ecologically responsible properties, all of which help the output results about the strong levels of the accident criticality classes.

Keywords **-** *Severity prediction, Machine learning, Accident prediction, Road accidents, Traffic system design.*

1. Introduction

Traffic Safety" is now the government's top priority. Detecting the causes of road accidents has become a top concern in recent years, intending to lower the impact of accidents in congested areas. Data mining is a process that involves gathering data from disparate sources and encapsulating it into useful information. Data mining users can evaluate information from incompatible dimensions or orientations, segregate it, and encapsulate the correlation identities. Data mining is the process of identifying complements or patterns between the number of paths in a database, most frequently in relational databases.

The main goal in analyzing accident data is to identify the essential factors related to road and traffic accidents. The different scenarios involve mining association rules to discover the model and identify the cause of the crash. This work mainly focuses on how the machine learning algorithm is applied in detecting criticality in a traffic accident. This work deals with the development of discretion from a group of control methods and the development of a model for classifying injury criticality. All these models are developed using various machine-learning techniques. On a major role particularly related to traffic accidents, supervised machine learning algorithms, such as Naive Bayes (NB), Random Forests (RF) and k-nearest neighbors (KNN), were executed.

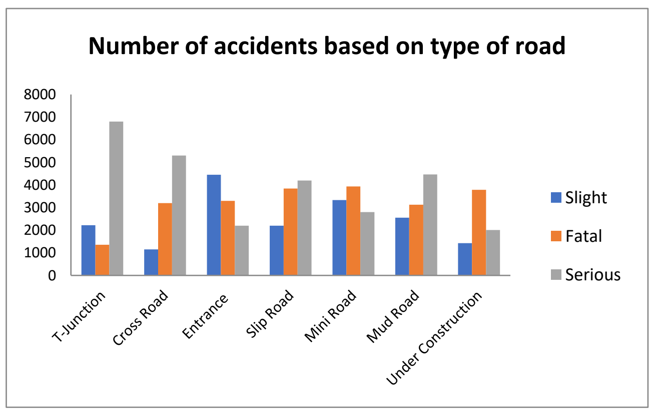

Fig. 1 Number of accidents based on the type of road

Fig. 1 shows the comparison to wars and diseases, and road accidents are the main source of the maximum death rate based on various types of roads. Nearly across the world, every year in a road accident, 12,45,000 people get injured or disabled, while 5,800 people lose their lives every day. To avoid such impacts, improving the transportation system and enhancing traffic safety that leads to functional administration of accidents is important to alleviate accident consequences. This work aims to detect the criticality of a road accident, which relies upon the environmental conditions and part of the injury.

Identity of the exactness of criticality of any twist of fate that had passed off affords a few critical facts for reacting all through the emergency crises to look at the criticality degree of the accident, evaluate the viable reasons and broaden systematic twist of directional fate methods. For example, assuming the preventive measures of the primary concepts that rule engineering, visitors and transportation making plans pertain to street transportation networks thinking about each city and rural areas. While the street twist of fate happens, there is probably a severe implication for the government. Moreover, the good-sized caution to the general public transportation is thru street accidents. The involved administrative departments of the shipping gadget must take important preventive protection measures and improve the information to the not unusual place humans take the preventive measures in ameliorating the street instances in reducing street accidents. It can simplest be finished by growing sufficient information to the drivers with crucial visitor engineering needs.

Data mining is crucial in analyzing information in this evolving day-to-day engineering technique. In the past years, a drastic expansion of gathering and categorizing data has been examined nowadays, which led to extending the database volumes in the form of human convenience in understanding that storage of this numerous data can be done only through some powerful tool mechanism.

In addition to relying on the information, managers and end users serve as decision-makers in this scenario because they are not given powerful tools for extracting critical data. Some statistical models are extensively used for investigating road accidents and the comparative connection between collisions and environmental factors. Investigators' primary goal is to study data to create categories using data mining algorithms. Data mining is used to reach final findings that have not been determined and may involve a database information evaluation process.

The last drawn conclusions include data cleaning and initial processing, constructing a data warehouse, selecting target data set, finding used features and determining new features, visualizing data, selecting data mining operations, pattern extraction, evaluating and interpreting results and deleting inadequate patterns, calculating the outputs of the information, and terminating from important data from various methods that which are used to determine the accident information. Although, some applications that are often used in traffic prevention examination are association rules perusal and machine learning algorithms.

Machine learning (ML) is an artificial intelligence specialization (AI). Deep learning can be clearly comprehended to target consumers by the layout of Predictive analytics, or predictive modelling can also be referred to as Machine learning, according to American computer scientist Arthur Samuel in 1959. Two of the ML algorithms have received widespread approval. Unsupervised learning identifies the algorithm to which the user did not include the data to trace the conditions to create the data based on the inputs. In contrast, supervised learning builds the algorithm based on the assumptions of input and output information provided by the user. The following machine learning grouping was set up to advance the specific strategy for determining the critical importance of the accident.

2. Literature Survey

Sameen and Biswajeeth proposed their "Severity Prediction of Traffic Accidents with Recurrent Neural Networks" in 2017. The model predicts the levels of severity and injury parts of the patients in accidents using an RNN. It works better to correlate the sequential data among the accident records. One or more gates replace the hidden layers; these are used for data operations in memory blocks to control the internal layers. The input traffic gives the factors of levels of accident severity.

Andrew and Joseph, in the year of 2015, proposed their work on "Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data–Driven, Machine Learning Approach". The decision of clinical rules predicts the death mortality of hospital patients using a machine learning model with the help of scores. K-means are used to pre-process the data to divide the clusters with mean data points.

Pradhan and Sameen proposed their work on "Predicting Injury Severity of Road Traffic Accidents Using a Hybrid Extreme Gradient Boosting and Deep Neural Network Approach" in 2019. In the pre-processing data stage, the missing values get removed and extracted information works based on feature extraction and target variables. The hybrid model predicts 96% accuracy in terms of optimization parameters, whereas the XGBoost model also works the same as the traditional model with an accuracy of severity levels of 95%.

Xiaoyi and Pan Lu proposed their work on "Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree" in the year 2020. This model predicts the compared results of the decision tree and random forest. The unbalanced data can be eliminated by improving the characteristics.

Once the prognostication system is developed, and its usage is organized, the website is progressed for the end user. The website includes an initial filling form that involves various options to be determined. These include various kinds of climates and kinds of vehicles etc. Finally, after filling in the required details in the form, the algorithm is invoked with the details given by the user to the prediction system with the user-submitted form. The end user gets the accident criticality in the form of a percentage.

3. System Model

3.1. Classical Statistics versus Machine Learning

Data mining approaches such as categorization are used to examine road accidents. The different components which are the main cause of road accidents can be examined by two methods, namely Classification and Regression Tree (CART) model. The examination method explaining road safety precautions are Naïve Bayes and Decision Tree classification algorithms. Consequently, the testing factor leading to the accident is identified using classification algorithms. Pedestrian and two-wheeler accidents are examined through decision trees and association rule mining...

Many factors relating to terribly critical accidents were investigated. The environmental conditions for pedestrian accidents and two-wheeler accidents were incompatible.

Their dominance was also evaluated, and it concluded that components were dominant, and the rule-based analysis solutions were consistent with the results of other manageable explorations using probabilistic models. Various decision tree algorithms and Naive Bayes algorithms were combined with distorted selection algorithms and implementations. The successes were examined, and the answers revealed that feature selection improved all models' classification performance.

In further examination, association rule mining is used after grouping the accidents determined by their frequency using a k-means clustering algorithm. Association rules can achieve the connection among the components which lead to the accident. By this, a contrasting method can be done to acquire the anticipated value. Comparisons between the Logistic regression model and the classification tree model can be done to conclude that the classification tree offers more data than the logistic regression model.

When estimated by combing different risk mechanisms, it is concluded that the Naïve Bayes classifier works efficiently when the risk factor for a provided critical level is liberated. Besides, the decision tree doesn't need any conjecture dependencies among the risk factors and works effectively without including the dependencies among different factors.

Classic statistical methods, however, may not be the best results because they have little limit constraints in determining the criticality of road accidents. Statistical regressions are used in many projects involving data form and construction and a sequential and functional form among dependent and explanatory variables. Identifying the institutions regarding linking linkages in the data gets complex and dense when these assumptions fail. Furthermore, because the globe fully relies on growing technologies, every operating firm backs up its historical data, which is useful in future technologies.

Regular statistical procedures should evolve with some proofs for having difficulties producing duplicates with bigger volumes of data, as Datasets are rapidly growing.

Regular statistical techniques should evolve with some proofs for having difficulty in making duplicates with the larger volumes of data for which the Datasets are rapidly increasing. From such a point of view, it is made simple using machine learning techniques in these complex data sets. It can ease the necessity of data and reduce the vulnerability of investigations. Considering these limitations of statistical modelling, many researchers have deployed (non-parametric) machine learning techniques for identifying the criticality of the injury.

In this case, utilising machine learning techniques to simplify these complicated data sets can lessen the need for data while also reducing the vulnerability of investigations. Many researchers have used (non-parametric) machine learning algorithms to determine the criticality of an injury, taking into account the limits of statistical modelling. Different machine learning paradigms, such as neural networks trained using hybrid learning approaches, support vector machines, decision trees, and concurrent mixed models, including decision trees and neural networks, are used to reproduce the criticality of the vehicle crash and its devastation. In machine learning paradigms, the hypothetical conclusions show that the hybrid decision tree neural network technique is superior to the single method.

3.2. Data Mining

The term Data mining is referred to the mechanism of deriving information from complex frames of data. In other words, it can also be termed as extracting knowledge from the data. In determining such problems, data pre-processing is a demonstrated method. This method uses the initial or raw data for additional processing. This additional processing includes attribute selection, data cleaning and data transformation. The information accumulated from different origins was synthesized, delineated, and inspected. The information that is not related to the data mining process is disregarded.

The process used to ascertain defective, deficient, or obstructive data and to upgrade the efficiency of improving identified errors and omissions is called data cleaning. The term indicates decreasing the faults and increasing the data quality efficiency. Few of the reported entries are clearly rejected either by human error or due to the development of the issue reporting system. If such errors can be corrected are rectified directly by the system. If a few errors identified for disclosure were not rectified, the entire study related to the disclosure was terminated.

The conversion of data values from the source to the destination data system of data format is called data transformation. During this transformation, some attributes were converted into necessary formats; for instance, the attribute "Minute" was converted to a 60-Minute format. The values of the attributes are hard-coded for recommended descriptions in the training dataset. The formal representation of the protocols and attributes of efficiency, maintenance and increase, which standardize a protocol, is mentioned below.

These represent the two deliberated estimations supporting the association rule in the mining process. The statistical significance of a protocol defined to carry and reliance is the confidence level of the identified associate rule that decreases the characteristic preference method in enhancing this work. In this, we suggested a feature selection method using different attributes for selecting, which are executed largely in road accidents. An apriority algorithm is introduced to remove the unnecessary rules by taking only a few main rules to select the attributes.

3.3. Data Set Selection

While working on the prediction systems, data acts as a major crucial role. It plays a key role in the entire work, i.e., the total system relies upon this data. The first and foremost step is to prefer the data on its selection which is to be initiated correctly. We obtain the data from government websites as everyone can access these datasets. Similarly, other possible ways of obtaining such data from numerous websites exist. The particular data which has been chosen depends on certain few restrictions and different factors and is used in our model for the prediction system.

3.4. Data Cleaning and Transformation

The further step after selecting the dataset is to immaculate the data and convert it into the required format, however probable. The dataset that we use might consist of various configurations. Also, it is feasible to use various datasets from various derivations, which may be present in various file formats. To represent them in user-friendly formats, we need to transform them to the required format or translate them into system support formats using type predictions. The purpose of this is to achieve the constraints of the datasets not required by the existing system; by combining them, the system becomes complex, increasing its processing time. Another cause of cleaning these datasets is that they include null and garbage values. The emulsion to this problem is by restoring the garbage values with the data modified by the various methods to accomplish.

3.5. Data Cleaning and Transformation

After the data has been updated and converted, it can be processed to the next stage after cleaning and considering the required limitations. The entire dataset has been partitioned into two equal partitions, each of which can store 60-40 or 75-25 percent of the data. The data is divided into the greatest divisions for processing. The algorithms we build will be done on this data, which will control the algorithm's understanding of constraints on its possession and represent predictions on future or unknown data. The necessary constraints derived from the cleaned data are only used during the algorithm's execution.

3.6. Random Forest Model (RF)

The Random Forest model (RF) is a collective method that develops linear decision trees at the priming period and results in the class which relies upon as approach of the class (Division) or termed as prognosis (reversion) of the independent trees. As mentioned earlier, the maximum of the selections needed to divide into a node is two sets, whereas the minimum may require are fixed to one for each leaf node. Higher division of the trees results in higher accuracy and avoids multi-collisions' difficulty.

It is a combination research method that develops a progression of decision trees at deployment time and returns the class that is the procedure of the classes (categorization) or means prognosis (reverting) of the particular trees. The samples are essential to break the node with the smallest number set to two, whereas the largest number of samples per leaf is set to one.

3.7. Random Forest Model (RF)

It specifies "probabilistic classifiers" constructed by the powerful individual assumptions among the characteristics of Bayes's theorem. However, Gaussian NB is involved because the characteristic contains groups of uninterrupted variables, also termed Normal distribution. A bell shape is obtained when marked on the graph, consistent with the mean of the predicted values mentioned.

$$p(accidents \mid Severity) = \frac{p(sevirity|accidents) * p(accidents)}{p(sevirity)}$$

Depending upon the derivation of Bayes' theorem with powerful (unaffected) individual expectations among the characteristics, the Naive Bayes classifiers are a group of simple "probabilistic classifiers". The Gaussian NB is preferred, such that the features contain linear variables.

3.8. K-Nearest neighbour (KNN)

In order to forecast the values of any current data point, the KNN algorithm uses 'feature similarity'. It explains the current data point and assigns its value depending on how nearly it represents the points in its training set.

$$Distance(x_i, y_i) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_m - y_m)^2}$$

The KNN algorithm can be compared to supervised ML, which is used under categorization and reversion predictive algorithms. This KNN algorithm is termed loafing learning since it is examined as non-parametric.

The term non-parametric is explained as no expectations. Since this model is mainly based upon the predictions of the user in finding the criticality compared to that of its structure, loaf learning explains no suggestions can be drawn due to which it doesn't identify from the input methodologies instantly; rather, it saves the dataset during the time of categorizing performing an activity on the dataset.

4. System Architecture

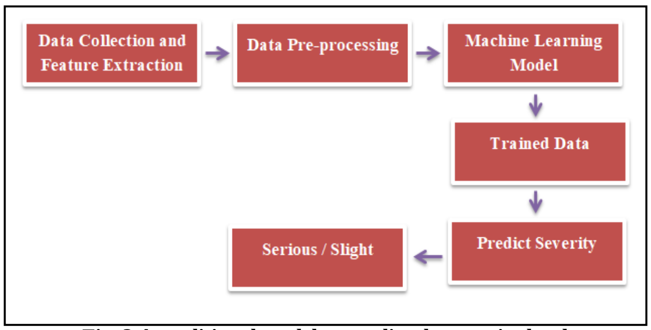

Fig. 2 A traditional model to predict the severity levels

Fig. 2 shows a traditional model only to predict the levels of accident severity, such as serious or slight. This model produces the results without considering the accident factors. So, this is insufficient for the current scenarios to calculate the severity level factors.

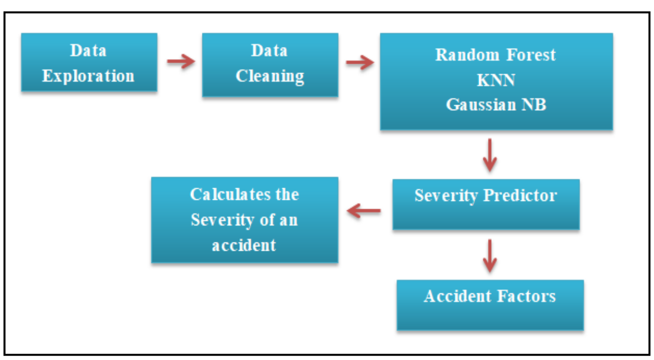

Fig. 3 Accident severity model using factors calculator

Fig. 3 shows it is a hybrid model than the traditional one, which combines machine learning techniques with predictor calculators through accident factors to determine the severity. The severity predictor works on two factors: a) Reading accident details and b) Predicting the severity. The reading module calculates the severity of an accident, such as slight, fatal and serious. The severity module reads the accident details such as junction control, light condition, weather condition, speed limit etc. The above two modules help to get accurate results for slight and serious.

5. Results and Discussion

The comparative results applied on the factors of serious and slight on the table-1 to table-4 show that the accuracy for the SVM is more efficient than the random forest, KNN and Gaussian NB. In table 1, the factors of slight are more compared to the serious accidents. Due to this, random forest produces efficient results as the record of serious accidents is low.

The table-2 produces moderate results by comparing with random forest and SVM as the weighted factors are high compared to the remaining. The table-3 produces the low results by comparing with random forest, KNN and SVM due to more slight accidents are recorded with the factor of seriousness.

Table 1. Accuracy comparison by random forest classifier

Factor	precision	recall	f1-score	Support	Accuracy
Serious	0.40	0.14	0.21	161	83.78%
Slight	0.86	0.96	0.91	906	
Macro Avg	0.63	0.55	0.56	1067	
Weighted Avg	0.79	0.84	0.80	1067	

Table 2. Accuracy comparison by K Neighbours classifier

Factor	precision	recall	f1-score	Support	Accuracy
Serious	0.31	0.19	0.24	153	81.25%
Slight	0.87	0.92	0.89	897	
Macro Avg	0.59	0.56	0.56	1025	
Weighted Avg	0.78	0.81	0.79	1033	

Table 3. Accuracy comparison by Gaussian NB

Factor	precision	recall	f1-score	Support	Accuracy
Serious	0.29	0.43	0.35	142	75.44%
Slight	0.89	0.81	0.85	753	
Macro Avg	0.59	0.62	0.60	982	
Weighted Avg	0.80	0.75	0.77	998	

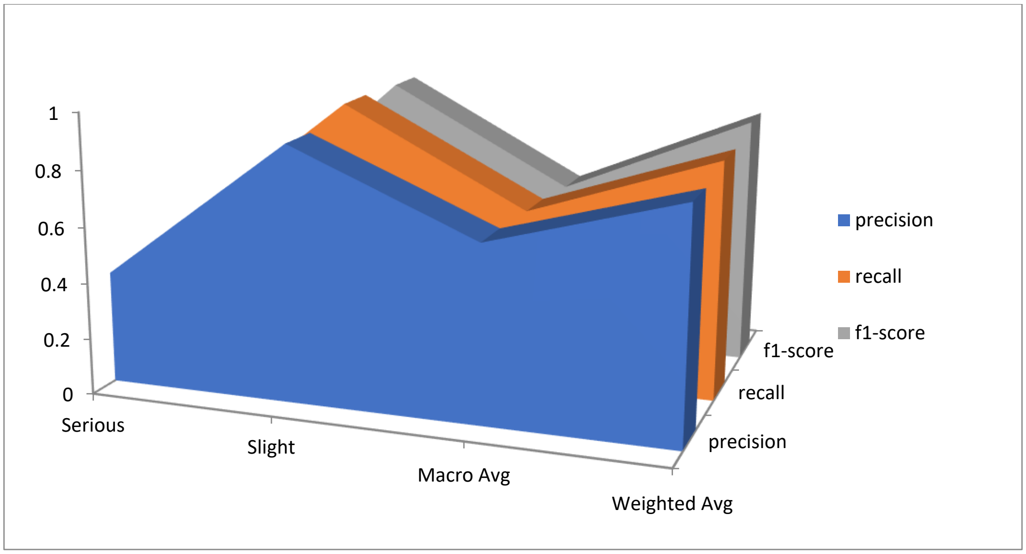

Fig. 4 Comparison of factors

Table 4. Accuracy comparison by SVM

Factor	preci sion	recall	f1-score	Support	Accuracy
Serious	0.35	0.21	0.28	161	84.6%
Slight	0.91	0.94	0.95	906	
Macro Avg	0.63	0.61	0.59	1067	
Weight ed Avg	0.82	0.85	0.89	1067	

On comparing the highest accuracy recorded results of various factors shown in figure 4, it is observed that the patients who met accidents under slight or moderate factors have the highest recovery rate by comparing the serious accident case factor. Serious accident cases have the lowest recovery rate as the patient's body parts' injured levels are high.

Table 5. Accident Severity is predicated as "SLIGHT"

Factor	Parameter
Carriageway Hazards	None
Light Conditions	Darkness
Day of the week	Saturday
Special conditions at the site	None
Road Type	A
Junction Control	Uncontrolled
Junction Details	Not a junction
Road Surface Condition	Frost
Road Type	Single Carriageway
Area	Urban
Weather Conditions	Fine, no high winds
Speed Limit	70
Part-of-injury	Hand
Time	Night

Table 6. Accident Severity is predicated as "SERIOUS"

Factor	Parameter
Carriageway Hazards	Present
Light Conditions	Light Present
Day of the week	Tuesday
Special conditions at the site	None
Road Type	A
Junction Control	Uncontrolled
Junction Details	Not a junction
Road Surface Condition	Dry
Road Type	Single Carriageway
Area	Urban
Weather Conditions	Fine, no high winds
Speed Limit	85
Part-of-injury	Spinal Cord
Time	Night

6. Conclusion

Every year, traffic accidents are a major disaster and productivity loss worldwide. It may be safer to design traffic system preventive measures with the help of these computational techniques. These study and inspect the control of machine learning algorithms to deploy effective and consistent classifiers, such as Random Forest (RF) and Nave Bayesian Classifier (NB) techniques. Among these, the matrixF1-Score, which is precise and based on the test, demonstrates that the Random Forest is thought to produce superior outcomes than the other learning methods. According to the research, support vector machines can make predictions with an accuracy of 84.6 cents. The data utilised to construct safe roadways could be valuable for highway engineers and transportation designers based on the developed algorithms. Additional research should be done to add more relevant data and evaluate the effects of these components. When analysing malignant and serious injuries, random forest is highly recommended. The predictive model plays a critical role in forecasting the outcomes of traffic accidents. The main limitation of this study is that it relies on essential components, including the driver, passenger, and pedestrian characteristics, as well as traffic conditions, to estimate accident criticality and time scale.

References

[1] Pradhan and Sameen, "Predicting Injury Severity of Road Traffic Accidents using a Hybrid Extreme Gradient Boosting and Deep Neural Network Approach," vol. 10, pp. 119-127, 2020.

[2] Xiaoyi and Pan Lu, "Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings using Random Forest Algorithm Compared with Decision Tree," vol. 200, pp. 106931, 2020.

[3] Laura Ebolia and Carmen Forciniti, "Factors Influencing Accident Severity: An Analysis by Road Accident Type", vol. 47, pp. 449-456, 2020.

[4] Md. Farhan and Ahmed, "Road Accident Analysis and Prediction of Accident Severity by using Machine Learning in Bangladesh," pp. 1-5, 2019.

[5] Jaspreet and Gurvinder Singh, "Evaluation and Classification of Road Accidents using Machine Learning Techniques," pp. 193-204, 2019.

[6] Lukuman and Haobin, "Evaluation and Classification of Road Accidents using Machine Learning Techniques," pp. 1-8, 2019.

[7] Bulbula and Fengli, "Classification of Road Traffic Accident Data using Machine Learning Algorithms," vol. 12, pp. 682-687, 2019.

[8] Rabia and Keneth, "Comparison of Machine Learning Algorithms for Predicting Traffic Accident Severity," vol. 1, pp. 272-276, 2019.

[9] Mayura Yeole, Rakesh Kumar Jain, Radhika Menon, "Prediction of Road Accident using Artificial Neural Network," *International Journal of Engineering Trends and Technology*, vol. 70, no. 3, pp. 151-161, 2022. Crossref, https://doi.org/10.14445/22315381/IJETT-V70I2P217

[10] Mpho and Dr. Vukosi, "Predicting Road Traffic Accident Severity using Accident Report Data in South Africa," pp. 11-17, 2019.

[11] S Vasavi, "Extracting Hidden Patterns within Road Accident Data using Machine Learning Techniques," vol. 6, pp. 13-22, 2018.

[12] Cigdem, "Predicting the Severity of Motor Vehicle Accident Injuries in Adana-Turkey using Machine Learning Methods and Detailed Meteorological Data," vol. 6, no. 1, pp. 72-79, 2018.

[13] Amirfarrokh and Aemal, "Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction," vol. 108, pp. 27-36, 2018.

[14] Sameen and Biswajeeth, "Severity Prediction of Traffic Accidents with Recurrent Neural Networks," vol. 7, pp. 476-493, 2017.

[15] Bahar and Blanca, "The Identification of Patterns of Interurban Road Accident Frequency and Severity using Road Geometry and Traffic Indicators," vol.14, pp. 4122-4129, 2016.

[16] Andrew, Joseph, "Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data–Driven, Machine Learning Approach," vol. 23, pp. 269-278, 2015.

[17] Akguuml A. P, & Doğan E, "An Application of Modified Smeed Adapted Andreassen and Artificial Neural Network Accident Models to Three Metropolitan Cities of Turkey," *Scientific Research and Essays*, vol. 4, no. 9, pp. 906-913, 2009.

[18] Dr. Surendiran R, Dr. Thangamani M, Monisha S, Rajesh P, "Exploring the Cervical Cancer Prediction by Machine Learning and Deep Learning with Artificial Intelligence Approaches," *International Journal of Engineering Trends and Technology*, vol. 70, no. 7, pp. 94-107, 2022. Crossref, https://doi.org/10.14445/22315381/IJETT-V70I7P211

[19] Delen D, Sharda R, & Bessonov M, "Identifying Significant Predictors of Injury Severity in Traffic Accidents using A Series of Artificial Neural Networks," *Accident Analysis and Prevention*, vol. 38, no. 3, pp. 434–444, 2006.

[20] Polson N. G, & Sokolov V. O, "Deep Learning for Short-Term Traffic Flow Prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.

[21] Abdel-Aty M. A, & Abdelwahab H. T, "Predicting Injury Severity Levels in Traffic Crashes a Modeling Comparison," *Journal of Transportation Engineering*, vol. 130, no. 2, pp. 204–210, 2004.

[22] Crammer K, & Singer Y, "On the Algorithmic Implementation of Multiclass Kernel-Based Vector Machines," *Journal of Machine Learning Research*, vol. 2, no. 12, pp. 265–292, 2001.

[23] Xie Y, Lord D, & Zhang Y, "Predicting Motor Vehicle Collisions using Bayesian Neural Network Models: An Empirical Analysis," *Accident Analysis and Prevention*, vol. 39, no. 5, pp. 922–933, 2007.

[24] Chong M, Abraham A, & Paprzycki M, "Traffic Accident Analysis using Machine Learning Paradigms," *Informatica*, vol. 29, no. 1, 2005.

[25] Kunt M. M, Aghayan I, & Noii N, "Prediction for Traffic Accident Severity: Comparing the Artificial Neural Network," *Genetic Algorithm Combined Genetic Algorithm and Pattern Search Methods, Transport,* vol. 26, no. 4, pp. 353-366, 2011.

[26] Abellán J, López G, & De OñA J, "Analysis of Traffic Accident Severity using Decision Rules Via Decision Trees," *Expert Systems with Applications*, vol. 40, no. 15, pp. 6047–6054, 2013.

[27] Friedman J. H, "Stochastic Gradient Boosting," *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367-378, 2002.