International Journal of Advanced Science and Technology
Vol. 29, No. 8, (2020), pp.4086-4097

Urban Road Accident Evaluation And Road Accident Severity Prediction

Anant Ram[1], B.Sikander[2]
[1]*Department of Computer Engineering and Applications, GLA University, Mathura, India*
[1]*anant.ram@gla.ac.in*
[2]*Department of Computer Engineering and Applications, GLA University, Mathura, India*
[2]*sikander.gla_mt18@gla.ac.in*

Abstract

Road accidents have emerged into a world issue .Based on the latest World Health Organization (WHO) report, road accidents are the world's 10th major cause of death. This has become a major problem in many developed and developing countries due to the large number of road accidents each year. It is completely inadmissible and saddening to enable road accidents to destroy the residents. Therefore, a detailed analysis is needed to handle this overwhelmed situation. In this paper, the research study seeks to use machine-learning models to more closely investigate road accidents and to estimate the rate of accidents in India. In this article, we also describe some factors that specifically influence road accidents and provide some useful suggestions on the matter. Here we considered five popular classification methods to build an accurate prediction model, such as Naive Bayes, Logistic Regression, Decision tree. The result is given by these methods is somewhat the same so then we used the ensemble learning concept i.e. Random Forest, lastly, we used a support vector machine to classify road accidents into Fatal, serious & Slight accidents. Ultimately, Random Forest provides the best and most effective result.

Keywords: *Road accidents, accident severity, Urban roads, Road Engineering, Supervised learning feature analysis method.*

1 Introduction:

Road accident is an unanticipated event that has happened on road with user, even though very often it occurs. Sadly, we can see a growing spike in road accidents not only in a particular country but all over the planet, high-road accidents in years past. There is a compelling influence on community and also on the economies of any nation, as deaths & injuries cost enormous. According to World Health Organization reports, traffic accidents worldwide surpassed 1.35 million. Due to over-speed vehicles, in India road accidents has taken more than 1.5 lakh lives in 2018, and the rate of road accidents is mostly increasing. In 2018, the Ministry of Road and Highways submitted a report on Indian road accidents showing that the road accident rate has increased by 0.46 percent since last year. Increased highway accidents and rise in death toll every day, In 2018, as per the road accidents reported, out of 1,51,417, the number of people killed 4,67,044 and 4,69,418 people got injured. Over-speeding accounted for 64.4% of the persons killed. Figure 1 shows the graph of the number of accidents from 2001 to 2016.

In past years, the study of road accidents has drawn significant interest of the researchers to identify factors that have a direct impact on traffic accidents. However identification of such factors is based on analytical data, or simple interviews or questionnaire surveys. While using these kinds of simple methods, it is not possible to obtain a perfect and unerring answer. The key problem is that analyzing the behavioral traits of traffic accidents using such forms of traditional research methods is quite difficult. Since events are relatively spontaneous and extemporaneous, it is quite difficult to observe explicitly. It is therefore very impossible to obtain 100% accurate data. Implementation of an innovative approach that can produce better outcomes for research is required. Machine learning is among the most improved & with multiple functionality areas of Artificial Intelligence science that can be applied to achieve more desirable and effective result here [1].

This research paper aims primarily at estimating the rate of road accidents and assessing the expanse of an accident using various techniques of machine learning [1].

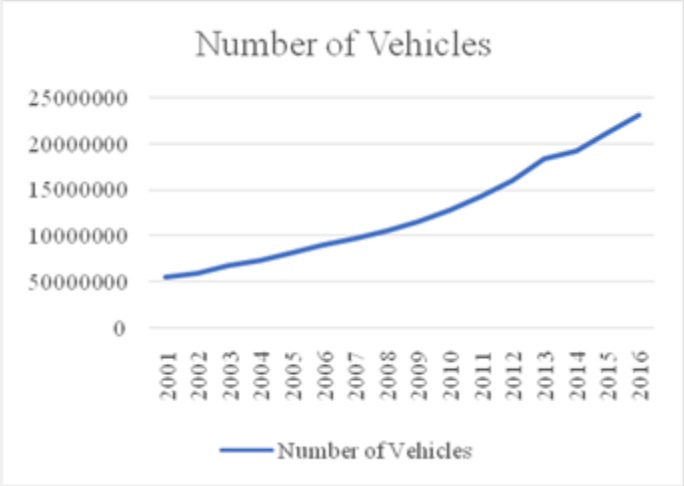

Figure 1 Number of Accidents in India from 2001 to 2016

In machine learning there are so many proven methods available for exploring this field. The research study conducted here to scrutinize road accidents, so that it forecast the prospect of road accidents based on road conditions by using the five advanced and most ubiquitous supervised machine learning approaches to demonstrate the accuracy of the prediction model in this field. Those approaches are Decision Tree, Naïve Bayes, Logistic Regression, Random Forest and Support Vector Machine (SVM).The researchers categorized the severity for an accident into 3 groups i.e. Fatal, Slight and Serious. In order to designate the seriousness in these three groups, the function has been chosen as fifty causes which can influence the overall no. of accidents. For our learning materials, 70% of road accident data comprising around 1 lakh report of a road accident from 2001 to 2019 were used. Random Forest obtained best performance among these five strategies and its accuracy was 82%.

Paper is organized as follows, Section 2 presents literature review, section 3 presents proposed model, section 4 presents the experimental evaluation and finally section 5 conclude the paper.

2 Related Work

Previous research study does not include Indian road and which does have not found any improve and satisfactory research work in this area in the form of Indian traffic. There are a few empirical works on the subject of Indian traffic accidents, but sadly such experiments were carried out either by using basic statistical methods or simply by conducting a survey. Unfortunately, the implementation of the modern machine learning model is in the formative stage here. With the vision-based techniques used in [2] presents that road accident identification can also be detected through visualization. In [2] they used roadside video data to train their models but their work only reached 85% precision in scenarios. The researchers in the paper [3] performed analysis on National Highway. The analysis involved 892 road accidents. The method used is multiple decision tree inductions algorithms. The analysis also involved traffic accident patterns. In this research authors figure out some rules and regulation for trees to decrease road accident on this highway. In this [4], authors used machine learning technique to conduct the accidents status on Istanbul. For such experiment they used CART algorithm for estimating accident probability and achieved accuracy more than 81.5%. The authors [5] used two mining algorithms named, k-means and association rule algorithm to classify various traffic accident-affiliated variables. The

authors [6] used mining algorithms such as association rule and k-means. Use of these strategies to classify the region most vulnerable to accidents and the key factors associated with it in India. There are various data mining techniques available but in [7] authors used Naive Bayes and compared their result with Decision Tree. This study also used KNN to determine a connection between road characteristics which was report in Ethiopia and their accident severity. Based on this analysis they set some rules regarding to the safety on road. In [8], the author used logistical regression to identify the accident cause because of road defects. The authors in [9], proposed an approach that is based on deep learning technique. The proposed method has better performance in terms of prediction power compared to the deep learning model without a regression layer and the Support Vector Machine (SVM) model. This model shows an 89.824% improvement in the deep learning model.

In paper [10] author proposed an algorithm based on a support vector machine which is able to predict driver's intention near a road intersection. The aim of this model is to classify whether the driver will stop, turn right, left or go straight according to traffic signal indicators. In paper [11] the author proposed three different models based on the type of road, namely Roundabout model, Junction model, and Straight section model. All three models function in their respective road type and the results of these models are not as quite good but not worst. As a start, these models illustrate road type and how it can be a cause of road accidents. In [12] presented an analysis of various artificial intelligence techniques. This article focuses on accident prediction analysis and also presents the unsafe driving patterns. According to this survey, it used Support Vector Machine (SVM), genetic algorithm and artificial neural networks.

In [13] authors proposed a technique that identify the secondary causes of distracted driving behavior. This study is based on driver behavior like speed, longitudinal acceleration etc.. It also uses the standard deviations to identify the different types of secondary tasks; drivers are engaged in while driving. Three secondary tasks which are to be considered like hand-held cell phone calling, texting, and interaction with an adjacent passenger.

Decision Tree classifier outperform with an accuracy of about 99.8% in classifying secondary tasks which distract drivers. The authors in [14] proposed a LiDAR-Video Driving benchmark dataset. It uses point clouds to help driving policy learning. The supervised end to end segmentation improves the performance. Machine learning approaches are classified into different classes like supervised, unsupervised, semi-supervised & reinforcement learning [15]. The researcher used five most popular machine learning methodology for road analysis. Those techniques are Naïve Bayes, Logistics Regression, Decision Tree, Random Forest and Support Vector Machine. In this paper we are using supervised machine learning approach.

3 Proposed Model

As we are aware that machine learning algorithms are applied on data to achieve the accurate result and it requires first to preprocess data. Therefore we proposed the model, presented in the following figure 2. This framework presents the sequence and flow of the experimental process followed in this work.

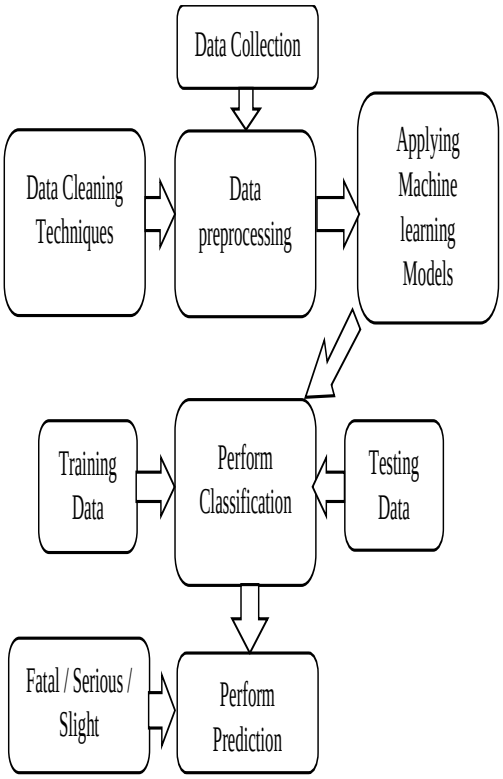

Figure 2: Proposed framework

Reliable and accurate incident data reports are the most important and primary need for improved performance through the implementation of machine learning approaches. But it is quite challenging to get a perfect and 100 % accurate data set. Therefore, following flow process shown in figure 3 has been used to achieve the objective.

Data Collection: The training data set for accurate prediction of the severity of injuries, a considerable number of traffic accident reports with full information are needed. The data set has collected from the Ministry of Road Transport & Highways ' Transport Research Wing (TRW). In this research work, It consists of a total of 1,22,636 traffic accidents reported in India from 2000-2018.We divide the entire data set into two parts, the training Data set and Test Data set. The training data set sample size is around 70% of the entire data set and remaining 30% data set is used as test data set.

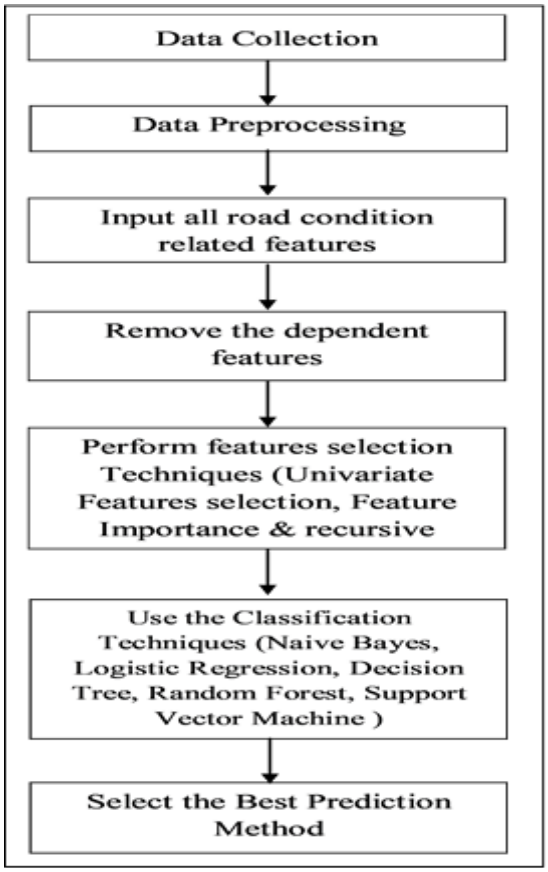

Figure 3: Process flow diagram

Data Pre-Processing: All reports of the incident were published using structured language in this data set. We organize this total data set according to the feature properly. In total, we found around thirty four factors that have one or other way an impact on previous accidents. Firstly, we used these thirty four features to methodize all accident records. After that, we find 8.7% missing values in the total data collection for many incident reports and 1.65% missing value for the limited 11 apps (Table 1).Since these missed values will impact efficiency, we have implemented a solution by using that column's mean value to provide a quantity where it is required. We are using this form because there is no excessive value that can influence the mean.

Feature Selection: Working with a large number of features will affect performance, as the number of features increases exponentially with training time. Even with the growing number of features it also has the possibility of over fitting. So, here collection of features is a critical factor for making a more accurate prediction. Sklearn (a python machine learning library) was used to diverge the less important feature. To acquire the most essential features, we performed an experiment utilizing three separate function selection algorithms. Such algorithms are Uni- variate Selection function, Recursive Elimination feature, and Importance feature. Uni-variate selection of features examines each feature extensively and chooses the best features based on uni-variate statistical testing. To choose the most essential features; we have introduced the Chi-Squared statistical test in non-negative features for this research work. Recursive Function Elimination operates by recursively eliminating the features and using the model's precision to select the features that help to determine the attribute.

Table 1 Important features obtained with the use of these three algorithms

Univariate Feature Selection	**Recursive Feature Elimination**	**Feature Importance**
District	Road geometry	Traffic control
Traffic control	Light	Weather
Weather	Weather	Junction Type
Time	Vehicle type	Time
Junction Type	Traffic control	District
Thana	Time	Thana
Light	Movement	Light
Road Geometry	divider	Road Geometry
Vehicle type	Road class	Vehicle type
Movement	Surface condition	Movement
Surface condition	Vehicle defect	Location type
Vehicle defect	Road Feature	Divider
Road class	Location Type	Vehicle loading
Location type	Surface type	Vehicle defect
Divider	Surface Quality	Road class

We implemented these three strategies of filtering of features and got the top fifteen features for each strategy (Table 1). We then tried to figure out these three experiments ' standard features and get eleven common features (Table 1).

Classification Techniques:

In this paper, we consider around 50 features which can cause a road accident, and based on our criterion we filter out those required features in which we build our prediction model.

Assume D as our raw dataset with around 1 lakh instances and X as input features which are 50 and T as our target values (Slight, Fatal, and Serious). We used Jypter notebook as our developing environment and sklearn library for machine learning methods.

In our proposed work, we first remove the null value and noise instance from our dataset, and then we filter out useful and valid data to build the prediction model. The steps are as follows:

1) First, we filter out features on basic of road environment condition which is involved in road accidents (directly and indirectly).

$$X = \{x_1, x_2, x_3 \dots, x_n\} \quad where\ n = 34$$

2) We again performed manual filtration in which we remove useless features on the basics of involvement in road accidents.

$$X = \{x_1, x_2, x_3 \dots, x_n\} \quad where\ n = 18$$

3) Now we performed machine learning features selection technique (Univariate features selection, recursive features elimination, and features importance) which we already stated above and the features we get through those techniques are given in table 1.

$$X = \{x_1, x_2, x_3 \dots, x_n\} \quad where\ n = 11$$

After that we applied the following technique to achieve the level of road accident prediction accuracy and we observed that random forest outperform with respect to the rest of the methods.

Logistic Regression: Logistic Regression Model: A Statistical model that is used to explain data and to describe the relationship between one dependent variable and at least one ordinal, nominal or ratio-level independent variable. This approach performs well when the test is based on two values.

$$n(t) = \frac{1}{1 + e^{-t}}$$

Support Vector Classifier: It is a supervised machine learning algorithm that examines data used for regression and classification analysis. In this, every data items is plotted as a point in n-dimension (where n indicates the number of features) along with each feature value. After that, to separate the two classes very well we find hyper-plane by performing classification.

Gaussian Naïve Bayes: It is classification model based on Bayes Theorem. It is not a single algorithm but the collection of different algorithms where the entire algorithm shares same objective. It requires that each feature of dataset must be independent and not related with one another. The main advantage of this model is that it provides feasible solution on large dataset.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Random Forest: A classification algorithm containing more no. of decision trees. It is like a bootstrapping algorithm. On data sample it creates an individual decision trees and then gets the forecast from every decision tree and then finally select the best feasible solution by means of voting.

Decision Tree: The supervised algorithm is commonly used with decision tree for classification problems. It is like structure in which each node represents the test on an attribute and each branch represents the outcome of the test. Lastly the leaf node of the decision tree identifies a class.

4 Result Analysıs And Dıscussıon

In this analysis we conducted two different experiments focused on the level of incident occurrence to examine the efficiency of the solutions suggested. We evaluated the efficiency of each algorithm in our first analysis, with three incident seriousness levels (Fatal / Serious / Slight).Among these five techniques, Random Forest achieves high accuracy and their accuracy is 82 percent (Table 2).The Random Forest gives the best output by overall performance which needs more training time than others.

Table 2 severity prediction results of algorithms

Algorithms	**Precision (%)**	**Accuracy (%)**
Naïve Bayes	79.81%	79.80%
Logistic Regression	79.81%	79.82%
Decision Tree	79.81%	79.80%
Random Forest	79.87%	82%

Support Vector Machine	79.82%	79.80%

We observed that for the other two groups, most of the incidents in our data set are Serious and are very low in value. For that cause we integrated these two accident occurrence levels into one class in our second study i.e. Slight and Serious. We have therefore achieved the efficiency of the solutions suggested for two classes of accident severity (Fatal/Serious). We also tested with the technologies in our dataset and tried to determine their affects on a road accident. We also systematically observed, in Figure 4, that the no. of accidents increased based on the condition of certain devices. It is a significant noteworthy thing to take effective measures to limit the number of accidents.

Figure 4Amount of accidents per feature status

We also observed the affect of vehicle types on a traffic accident and it has been found that 82 per cent of the taxis are liable for traffic accidents as shown in Figure 5. Figure 6 indicates that most of the incidents occurred on National Route.

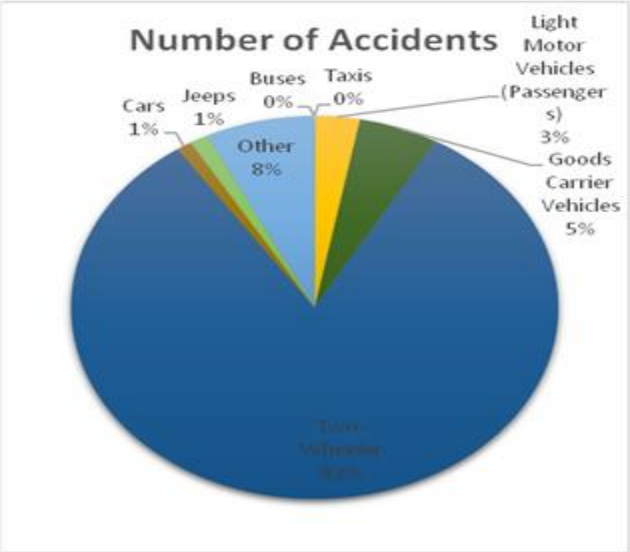

Figure 5 Effect of Vehicles on road accidents

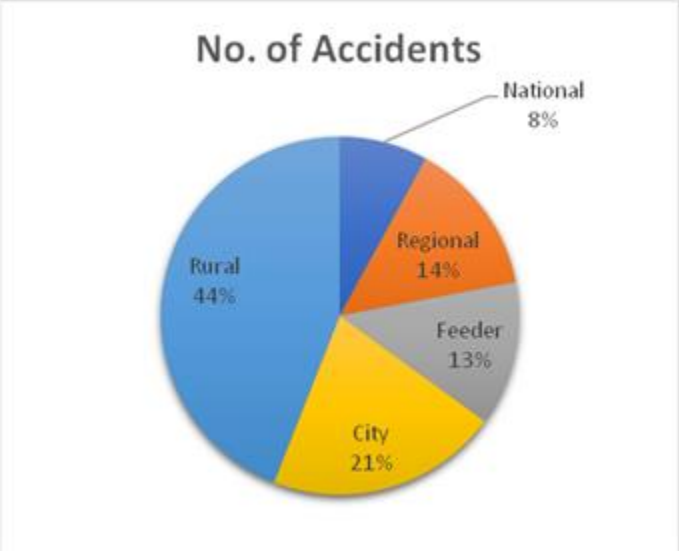

Figure 6 Effect of Road class on road accidents

So, we also performed the analysis to find out road accidents based on the traffic rush time. Based on the result, rush hour (06-18) accident rates are ex-cogitated to be very high relative to the other time as shown in figure 7.

Figure 7 Time impact on road crashes

Then we measured the casualties according to the type of junction. Figure 8 shows that the rate of accidents is higher where there is no junction, and at the T-Junction. Finally, we find in Figure 9 that the no. of accidents is that, depending on the state of the surface effect characteristics. Most of the accidents happened at the period when the surface layer is clear, surface layer sealed, and surface quality is good.

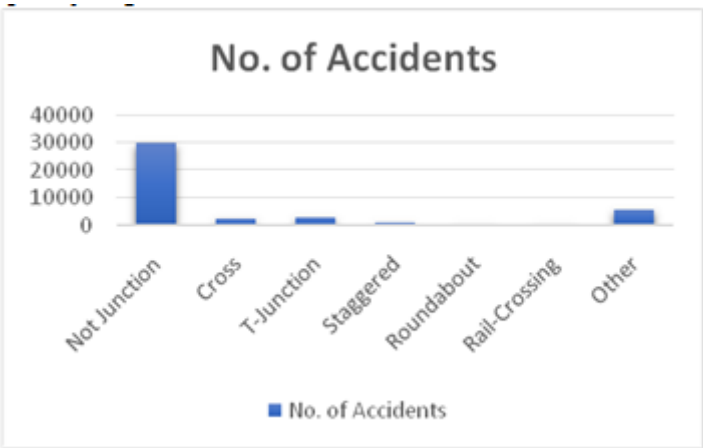

Figure 8 Effect of junction type on road accidents

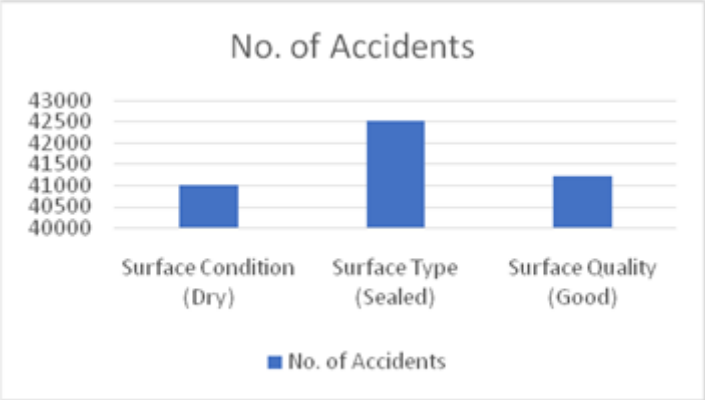

Figure 9 Effect of Surface condition on road accidents

5 Conclusıon

With civilization as well as a developing country like us, casualties on road accidents are intolerable. Therefore, reducing the frequency of accidents in any nation and to provide the traffic shape became essential. By taking basic steps, traffic accidents may be avoided based on sophisticated device forecasts or warnings. In reality, fixing this condition where so many people have died every day in a traffic accident is now a primary need for our world. This rate is increasing with each day. Implementing machine learning is a realistic and effective solution to informed decision taking with the awareness of managing the current situation. The findings of the study portion (Figure 4- Figure 9) can be suggested to traffic authorities to reduce the number of incidents that arise. Because of their verified and higher precision, we used suggested approaches for implementing machine learning to forecast traffic accident severity here. In addition, to make it more feasible, we can try to make the suggested framework by using these methods that can provide traffic accident prediction and can alert the road user. For the future, we try to enhance the accuracy of the model by a different method which gives more accurate value and also gives the right prediction so that the safety of life on the road can secure.

References

1. K. M. Habibullah, A. Alam, S. Saha, A. Amin and A. K. Das, " A Driver-Centric Carpooling: Optimal Route-Finding Model using Heuristic Multi–Objective Search",2019 4th International

Conference on Computer and Communication Systems (ICCCS), Singapore, 2019.

2. M. M. L. Elahi, R. Yasir, M. A. Syrus, M. S. Q. Z. Nine, I. Hossain and N. Ahmed, "Computer vision based road traffic accident and anomaly detection in the context of Bangladesh", 2014 International Conference on Informatics, Electronics & Vision (ICIEV), pp. 1-6, Dhaka, 2014.
3. M. S. Satu, S. Ahamed, F. Hossain, T. Akter and D. M. Farid, "Mining traffic accident data of N5 national highway in Bangladesh employing decision trees", 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), pp. 722-725, Dhaka, 2017.
4. H. İ. Bülbül, T. Kaya and Y. Tulgar, "Analysis for Status of the Road Accident Occurrence and Determination of the Risk of Accident by Machine Learning in Istanbul", 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 426-430 Anaheim, CA, 2016.
5. P. A. Nandurge and N. V. Dharwadkar, "Analyzing road accident data using machine learning paradigms", 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC), Palladam, pp. 604-610 ,2017.
6. S. Kumar and D. Toshniwal, "A data mining approach to characterize road accident locations", Journal of Modern Transportation, vol. 24, issue no. 1, pp. 62-72, 2016.
7. Beshah, Tibebe, and Shawndra Hill, "Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia", AAAI Spring Symposium: Artificial Intelligence for Development (2010).
8. A. Esmaeili, M. Khalili and A. Pakgohar, "Determining the road defects impact on accident severity; based on vehicle situation after accident, an approach of logistic regression," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), pp. 1-4, Langkawi, 2012.
9. Dong, Chunjiao, Chunfu Shao, Juan Li, and ZhihuaXiong. "An improved deep learning model for traffic crash prediction." Journal of Advanced Transportation, 2018.
10. Amsalu, Seifemichael B., AbdollahHomaifar, Fatemeh Afghah, SainaRamyar, and Arda Kurt. "Driver behavior modeling near intersections using support vector machines based on statistical feature extraction." In 2015 IEEE intelligent vehicles symposium (IV), pp. 1270-1275. IEEE, 2015.
11. Gianfranco, Fancello, Stefano Soddu, and Paolo Fadda. "An accident prediction model for urban road networks." Journal of Transportation Safety & Security 10, no. 4 , pp. 387-405, 2018.
12. Halim, Zahid, RizwanaKalsoom, Shariq Bashir, and Ghulam Abbas. "Artificial intelligence techniques for driving safety and vehicle crash prediction." Artificial Intelligence Review 46, no. 3, pp. 351-387, 2016.
13. Osman, Osama A., Mustafa Hajij, SogandKarbalaieali, and Sherif Ishak. "A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data." Accident Analysis & Prevention, No. 123, pp. 274-281 ,2019.
14. Chen, Yiping, Jingkang Wang, Jonathan Li, Cewu Lu, Zhipeng Luo, Han Xue, and Cheng Wang. "Lidar-video driving dataset: Learning driving policies effectively." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5870-5878. 2018.
15. M. A. A. Mamun, J. A. Puspo and A. K. Das, "An intelligent smartphone based approach using IoT for ensuring safe driving," 2017 International Conference on Electrical Engineering and Computer Science (ICECOS), Palembang, 2017, pp. 217-223