ROAD ACCIDENT ANALYSIS AND PREDICTION OF ACCIDENT SEVERITY USING MACHINE LEARNING

AKANKSHA JADHAV[1], SHRUTI JADHAV[2], ARCHANA JALKE[3], KIRTI SURYAVANSHI[4]

[1-4]Smt. Indira Gandhi College of Engineering, Navi Mumbai (400701)

Abstract: In recent years, the road accident has become a global problem and marked as the ninth prominent cause of death in the world. Due to the enormous number of road accidents every year, it has become a major problem in our country. It is entirely inadmissible and saddening to allow its citizen to kill by road accidents. Consequently, to handle this overwhelmed situation, a precise analysis is required. In this it will be done to analyse traffic accidents more deeply to determine the intensity of accidents by using machine learning approaches in our country. We also figure out those significant factors that have a clear effect on road accidents and provide some beneficent suggestions regarding this issue. Analysis has been done, by using Deep Learning Neural Network and AdaBoost these two supervised learning techniques, to classify the severity of accidents into Fatal, Grievous, Simple Injury and Motor Collision these four categories.

Keywords: Traffic accident, machine learning, adaboost, deep neural network.

1. INTRODUCTION

The problem of deaths and injuries as a result of accidents is to be a global phenomenon. [1] Traffic safety has been a serious concern since the start of the automobile age, almost one hundred years ago. [2] It has been estimated that over 300,000 persons die and 10 to 15 million persons are injured every year in road accidents throughout the world. [3] Statistics have also shown that mortality in road accidents is very high among young adults that constitute the major part of the work force. [4] In order to overcome this problem there is need of various road safety strategies, methods and counter measures. The survey was conducted on different causes of death due to injury. World Health Organization (WHO) report tells a horrible story that, most of the deaths between the ages 15 to 29 years are occur due to road traffic accidents and per year, more than 1.25 million people lost their lives due to road crashes. A survey from WHO reported some common reasons like shortage of training institutes, poor condition of roads as well as poor traffic management are the root causes. So to overcome this issue a systematic approach and firmly based solution is required with efficient and effective measures. So our system encounters such parameters and gives a systematic and visualizes view to overcome and interpret the respective problem. Engineers and researchers in the automobile industry have tried to design and build safer automobiles, but traffic accidents are unavoidable. [5] Patterns involved in dangerous crashes could be detected by developing a prediction model that automatically classifies the type of injury severity of various traffic accidents. These behavioural and roadway patterns are useful in the development of traffic safety control policy. It is important that measures be based on scientific and objective surveys of the causes of accidents and severity of injuries. The system presents some models to predict the severity of injury that occurred during traffic accidents using machine-learning approaches. We considered networks trained using learning approaches. Experiment results reveal that among the machine learning paradigms considered various paradigms approaches.

1.1 Problem Statement:

To handle the enormous number of road accidents in a locality a precise analysis is required. This analysis will be done more deeply to determine the intensity of the road accidents by using supervised learning techniques like Deep Learning Neural Network and AdaBoost. This will classify the severity of the accidents as fatal, grievous, simple injury and motor collision. Many of agencies especially government agencies are identify the factors that contribute to the accident roads or highways. The measurements to prevent accident speed reduction, widen divider, or other else. These different types of the accident on the critical roads or highways for of agencies such as Royal Mal (JKR), Road Transport process, planning process or in remedy process for into a serious part when all the road users measurement of how the accident can be occur. There model that has been developed to analyze the accident can analyze all the variables. It also difficult for the model that has been developed due to hardly complicated mathematical model.

1.2 Project Objectives:

1. To study the causes of the accident by features extraction from the images given as the input.
2. To understand the severity of the accident based on these features.
3. To classify it as fatal, grievous, simple injury or motor collision.
4. To carry out the above mentioned algorithms to predict the performance and accuracy of each.
5. To conclude the fastest algorithm.
6. To understand the effect of each feature on the accident and conclude how much is it responsible for the accident.

2. Related Work

2.1 Image processing :

Though these given approaches keep an accurate track of motion of the vehicles but perform poorly in parametrizing the criteria for accident detection. They do not perform well in establishing standards for accident detection as they require specific forms of input and thereby cannot be implemented for a general scenario. The existing approaches are optimized for a single CCTV camera through parameter customization. However, the novelty of the proposed framework is in its ability to work with any CCTV camera footage.

2.2 Image feature extraction :

Vision-based traffic accident detection system for automatically detecting, recording, and reporting traffic accidents at intersections. This model first extracts the vehicles from the video image of CCTV camera, tracks the moving vehicles, and extracts features such as the variation rate of the velocity, position, area, and direction of moving vehicles. The model then makes decisions on the traffic accident based on the extracted features. And we suggested and designed the metadata registry for the system to improve the interoperability.

2.3 Uploading of image to website:

Our biggest concern is when people have taken photos or videos of the scene and they get uploaded to social media before we get to identify those involved and contact the family. But by using this photos we can get to know how this accident happened . we get all the information about the accident. So we can predicate the rod accident. In website we upload this image then analysis this with the use of Adaboost and neural network. Where AdaBoost algorithm is more efficient to predicate accident.

3. Methodology

Models are created using accident data records which can help to understand the characteristics of many features like drivers behavior, roadway conditions, light condition, weather conditions and so on. This can help the users to compute the safety measures which is useful to avoid accidents. It can be illustrated how statistical method based on directed graphs, by comparing two scenarios based on out-of-sample forecasts. the model is performed to identify statistically significant factors which can be able to predict the probabilities of crashes and injury that can be used to perform a risk factor and reduce it.

Here the road accident study is done by analyzing some data by giving some queries which is relevant to the study. The queries like what is the most dangerous time to drive , what fractions of accidents occur in rural, urban and other areas. What is the trend in the number of accidents that occur each year,do accidents in high speed limit areas have more casualties and so on … These data can be accessed using Microsoft excel sheet and the required answer can be obtained. This analysis aims to highlight the data of the most importance in a road traffic accident and allow predictions to be made. The results from this methodology can be seen in the next section of the report.

The system provides a three-phase solution to analyze road accidents in India using machine learning and computer vision. The block diagram covers the whole approach in a broad sense. The central component of the block diagram represents the selected approach whereas its counterparts indicate the experimented techniques. After testing the accuracy and results of all the methods. We came up with an optimal approach for each and every task.

4. Algorithms

4.1 AdaBoost :

AdaBoost algorithm, short for Adaptive Boosting, is a Boosting technique that is used as an Ensemble Method in Machine Learning. It is called Adaptive Boosting as the weights are re-assigned to each instance, with higher weights to incorrectly classified instances. Boosting is used to reduce bias as well as the variance for supervised learning. It works on the principle where learners are grown sequentially. Except for the first, each subsequent learner is grown from previously grown learners. In simple words, weak learners are converted into strong ones. Adaboost algorithm also works on the same principle as boosting, but there is a slight difference in working.

Uses of AdaBoost in image processing:

1. Image data
2. Classification prediction problems
3. AdaBoost is increasingly being used in the industry and has found its place in Facial Recognition systems to detect if there is a face on the screen or not.

4.2 Deep neural network :

A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. There are different types of neural networks but they always consist of the same components: neurons, synapses, weights, biases, and functions.

Uses of DNN :

1. DNNs ability to take in a lot of inputs, process them to infer hidden as well as complex, non-linear relationships, DNNs are playing a big role in image and character recognition.
2. Image recognition is an ever-growing field with widespread applications from facial recognition in social media.
3. DNNs are powerful models that have a wide range of applications.

Artificial Neural Network(ANN) Unlike many other prediction techniques, ANN does not impose any restrictions on the input variables (like how they should be distributed). Additionally, many studies have shown that ANNs can better model heteroskedasticity i.e. data with high volatility and non-constant variance, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data. This is something very useful in financial time series forecasting (e.g. stock prices) where data volatility is very high .

5. Approaches of model

5.1 Vehicle Detection:

This phase of the framework detects vehicles in the video.

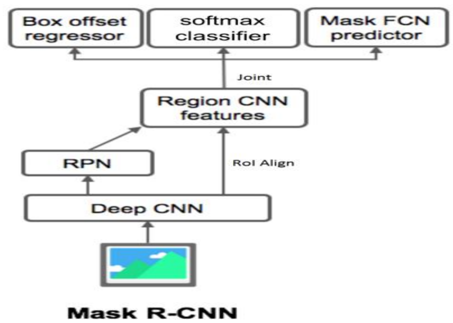


Fig. 1. The Mask R-CNN framework (from [13])

The object detection framework used here is Mask RCNN (Region-based Convolutional Neural Networks) as seen in Figure 1. Using Mask R-CNN we automatically segment and construct pixel-wise masks for every object in the video. Mask R-CNN is an instance segmentation algorithm that was introduced by He et al. [8]. Mask R-CNN improves upon Faster R-CNN [14] by using a new methodology named as RoI Align instead of using the existing RoI Pooling which provides 10% to 50% more accurate results for masks [8]. This is achieved with the help of RoI Align by overcoming the location misalignment issue suffered by RoI Pooling which attempts to fit the blocks of the input feature map. Mask R-CNN not only provides the advantages of Instance Segmentation but also improves the core accuracy by using RoI Align algorithm. The result of this phase is an output dictionary containing all the class IDs, detection scores, bounding boxes, and the generated masks for a given video frame.

5.2 Vehicle Tracking and Feature Extraction:

After the object detection phase, we filter out all the detected objects and only retain correctly detected vehicles on the basis of their class IDs and scores.

Once the vehicles have been detected in a given frame, the next imperative task of the framework is to keep track of each of the detected objects in subsequent time frames of the footage. This is accomplished by utilizing a simple yet highly efficient object tracking algorithm known as Centroid Tracking [15]. This algorithm relies on taking the Euclidean distance between centroids of detected vehicles over consecutive frames. From this point onwards, we will refer to vehicles and objects interchangeably. The centroid tracking mechanism used in this framework is a multi-step process which fulfills the aforementioned requirements. The following are the steps:

1) The centroid of the objects are determined by taking the intersection of the lines passing through the mid points of the boundary boxes of the detected vehicles.

2) Calculate the Euclidean distance between the centroids of newly detected objects and existing objects.

3) Update coordinates of existing objects based on the shortest Euclidean distance from the current set of centroids and the previously stored centroid.

4) Register new objects in the field of view by assigning a new unique ID and storing its centroid coordinates in a dictionary.

5) De-register objects which haven't been visible in the current field of view for a predefined number of frames in succession.

The primary assumption of the centroid tracking algorithm used is that although the object will move between subsequent frames of the footage, the distance between the centroid of the same object between two successive frames will be less than the distance to the centroid of any other object. Once the vehicles are assigned an individual centroid, the following criteria are used to predict the occurrence of a collision as depicted in. C1: The overlap of bounding boxes of vehicles C2: Determining Trajectory and their angle of intersection

C3: Determining Speed and their change in acceleration

5.3 Accident Detection:

This section describes the process of accident detection when the vehicle overlapping criteria (C1, discussed in Section III-B) we will introduce three new parameters (α, β, γ) to monitor anomalies for accident detections. The parameters are:

1) Acceleration Anomaly, α

2) Trajectory Anomaly, β

3) Change in Angle Anomaly, γ

When two vehicles are overlapping, we find the acceleration of the vehicles from their speeds captured in the dictionary. We find the average acceleration of the vehicles for 15 frames before the overlapping condition (C1) and the maximum acceleration of the vehicles 15 frames after C1. We find the change in accelerations of the individual vehicles by taking the difference of the maximum acceleration and average acceleration during overlapping condition (C1). The Acceleration Anomaly (α) is defined to detect collision based on this difference from a pre-defined set of conditions. This parameter captures the substantial change in speed during a collision thereby enabling the detection of accidents from its variation. Since in an accident, a vehicle undergoes a degree of rotation with respect to an axis, the trajectories then act as the tangential vector with respect to the axis. By taking the change in angles of the trajectories of a vehicle, we can determine this degree of rotation and hence understand the extent to which the vehicle has underwent an orientation change.

Lastly, we combine all the individually determined anomaly with the help of a function to determine whether or not an accident has occurred. This function f(α, β, γ) takes into account the weightages of each of the individual thresholds based on their values and generates a score between 0 and 1 A score which is greater than 0.5 is considered as a vehicular accident else it is discarded. This is the key principle for detecting an accident.

6. Experimental Results

6.1 Dataset Used :

This work is evaluated on vehicular collision footage From different geographical regions, compiled from YouTube. The surveillance videos at 30 frames per second (FPS) are considered. The video clips are trimmed down to approximately 20 seconds to include the frames with accidents. All the data samples that are tested by this model are CCTV videos recorded at road intersections from different parts of the world. The dataset includes accidents in various ambient Conditions such as harsh sunlight, daylight hours, snow and night hours. A sample of the dataset is illustrated in Figure 3.

6.2 Results, Statistics and Comparison with Existing models:

We estimate the collision between two vehicles and visually represent the collision region of interest in the frame with a circle as show in Figure 4. We can observe that each car is encompassed by its bounding boxes and a mask. The magenta line protruding from a vehicle depicts its trajectory along the direction. In the event of a collision, a circle encompasses the vehicles that collided is shown. The existing video-based accident detection approaches use limited number of surveillance cameras compared to the dataset in this work. Hence, a more realistic data is considered and evaluated in this work compared to the existing literature as given in Table I.

6.3 detection of a vehicular accident:

The proposed framework achieved a detection rate of 71 % calculated using Eq. 8 and a false alarm rate of 0.53 % calculated using Eq. 9. The efficacy of the proposed approach is due to consideration of the diverse factors that could result in a collision.

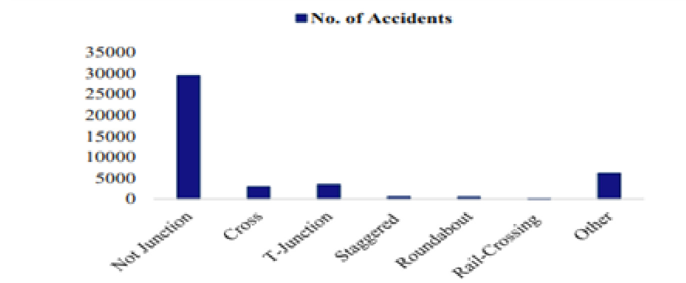
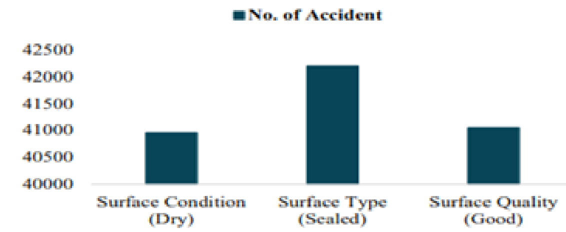


Figure 6. Effect of junction type on accident



Detection Ratio = Detected accident cases /Total accident cases in the dataset × 100

False Alarm Rate = Patterns where false alarm occurs/ Total number of patterns × 100

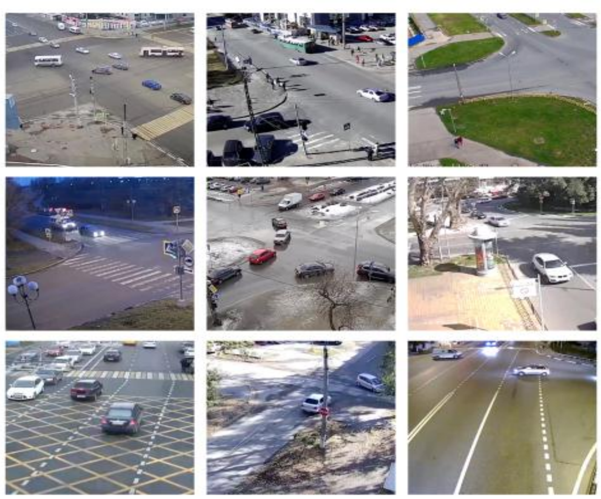


Fig. 3. Videos depicting various traffic and environmental conditions in the dataset used to evaluate the performance of the proposed framework.

In this research paper, to evaluate the performance of the proposed approaches, we performed two different. experiments based on the accident severity class. In our first experiment, we have determined the performance of each algorithm, for four accident severity classes (Fatal / Grievous /Simple Injury/ Motor Collision). Naïve Bayes and Ada-Boost both of them, achieve the high accuracy among these four approaches, and their accuracy is 80% (Table II).

By overall performance, Ada-Boost gives the best result because of its iterative classification on decision tree.

No. of Class / Algorithms	Precision (%)		Accuracy (%)		F1 Score (%)	
	Four Class	Two Class	Four Class	Two Class	Four Class	Two Class
Decision Tree	68	70	71	73	71	71
KNN	68	70	67	69	67	69
Naïve Bayes	63	63	80	80	71	71
Ada-Boost	68	75	80	80	73	74

Table II: Severity prediction results of algorithms

The performance of AdaBoost is much better than the previous experiment as precision and F1 score increased here in a noticeable way.

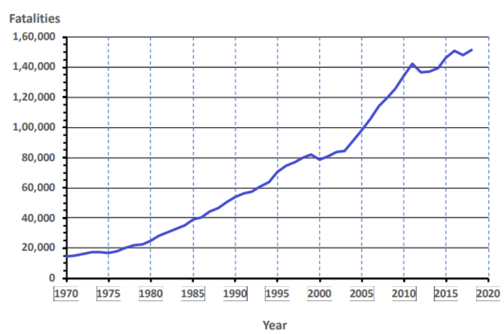


Figure 1 : Road traffic deaths in India 1970 through 2018 (Source: NCRB 2015 & Transport Research Wing, 2019).

the growth of personal motor vehicles registered in India by year according to official data (Transport Research Wing 2018). dental deaths and injuries in India varies according to age, gender, month and time. Age group 30- 59 years is the most vulnerable population group, though males face higher level of fatalities and injuries than their female accident scenario at state and city level shows that there is a huge variation in fatality risk across states and cities.

In general, while in many developed and developing countries including China, road safety situation is generally improving, India faces a worsening situation. Without increased efforts and new initiatives, the total number of road traffic deaths in India is likely to cross the mark of 250,000 by the year 2025.

CONCLUSION

In this paper, we propose a road accident analysis and prediction of accident severity using machine learning .Losses in road accidents are unbearable, to the society as well as a developing country like us. So, it has become an essential requirement to control and arrange traffic with an advanced system to decrease the number of road accidents in our country. By taking simple precautions, based on prediction or warnings of a sophisticated system may prevent traffic accidents. Moreover, it's a primary need for our country now, to tackle this situation where every day so many people were killed in a traffic accident and day by day this rate is getting increased. The implementation of machine learning is a functional and a great approach to take an accurate decision with the experience to manage the current situation and the findings of the analysis part can be suggested to traffic authorities for reducing the number of accidents. We can use proposed approaches to implement machine learning here because of their proven and higher accuracy to predict traffic accident severity. Moreover, to make it more feasible, we will try to make a recommender system by using these approaches that can give a prediction to the traffic accident and can warn the road user. In the future, it will be our try to create a mobile application by implementing this methodology to provide an accurate prediction to the user and make it very useful and beneficial also.

REFERENCES

1. T. Rahman, "Road Accidents in Bangladesh: An Alarming Issue", the World Bank, 2012. [Online].

2. 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.

3. 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), Dhaka, 2014, pp. 1-6.

4. 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), Dhaka, 2017, pp. 722-725.

5. 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), Anaheim, CA, 2016, pp. 426-430.

6. 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, 2017, pp. 604-610.

7. S. Kumar and D. Toshniwal, "A data mining approach to characterize road accident locations," Journal of Modern Transportation(2016), vol. 24, issue no. 1, pp. 62-72.

8. 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)

9. 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), Langkawi, 2012, pp. 1-4.

10. 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.

11. T. Adhikary, A. K. Das, M. A. Razzaque, A. Almogren, M. Alrubaian, and M. M. Hassan, "Quality of Service Aware Reliable Task Scheduling in Vehicular Cloud Computing," Mobile Networks and Applications, Volume 21, Issue 3, pp 482-493, June 2016.

12. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research (2011), vol. 12, pp. 2825-2830.

13. H. I. Bulbul and Ö. Unsal, "Comparison of Classification Techniques used in Machine Learning as Applied on Vocational Guidance Data," 2011 10th International Conference on Machine Learning and Applications and Workshops, Honolulu, HI, 2011, pp. 298-301.

14. R. Wang, "AdaBoost for Feature Selection, Classification and Its Relation with SVM", A Review. Physics Procedia (2012), vol. 25, pp. 800-807.

15. Z. Zhang, "Introduction to machine learning: k-nearest neighbors," Annals of Translational Medicine (2016), vol 4, issue no. 11, pp.218- 218.

16. I. Rish, "An Empirical Study of the Naïve Bayes Classifier," IJCAI 2001 Work Empir Methods Artif Intell (2001), vol 3.