Road Accident Prediction Model Using Machine Learning: A Review

N P Shah[1], Prajakta Masal[2], Samina mulani[2], Shraddha Khandomalke[2], Mahadevi Mukadm[2]

[1]Asst. Prof., Department of Computer Engineering, Savitribai Phule Pune University, Pune, India.

[2]UG student, Department of Computer Engineering, Savitribai Phule Pune University, Pune, India.

Corresponding Author: prajaktamasal20@gmail.com

Abstract: - The number of daily accidents due to road conditions, vehicle speed, weather conditions, etc.is a very crucial part today. Mainly accidents involving due to some weather, environmental conditions and due to traffic also. The relation between occurrence of traffic accidents and some factors and identifying main factors that contributing accident severity. Using advanced predictive capabilities of technology accurately assessing accidents and finding main reason or factors those contributing accidents. A series of machine learning and deep learning models are presented by integrating random forests and convolutional neural networks (RFCNN) to predict traffic accident severity. The model optimized further in future and help us to better monitor accident prone areas and provide emergency services.

I. INTRODUCTION

The number of road deaths in India is very alarming. Accidents involving heavy vehicles such as tractors, trucks, and public transportation utility vehicles such as coaches are among the deadliest accidents that occur. Weather conditions such as rain, fog and snow also contribute to the risk of accidents. Accurately assessing accidents and knowing the factors that cause accidents can therefore help reduce accidents. This requires careful study of accidents and development of accident prediction models.

For machine learning, the advanced predictive capabilities of computer systems. Machine learning is a convenient way to make true, empirical decisions. The relationship between the occurrence of traffic accidents and contributing environmental factors. Identifying the main factors influencing traffic accident severity is necessary to minimize accident-related severity and to analyze and avoid accident events using this model.

Accurately predict traffic accidents using a suite of learning models, including machine learning and deep learning models. Random forest is a machine learning technique that is a type of supervised learning algorithm that takes a set of decision trees over different subsets of a dataset to improve the accuracy of predictions for that dataset. Convolutional neural networks are a type of deep learning. CNNs are commonly used for pattern recognition and image processing.

II. LITERATURE SURVEY

Analysis of Traffic Accident Severity based on Machine Learning and Deep learning decision-making levels [1] It including RFCNN as fusion of both machine and deep learning model to assess the frequency of traffic severity. A variety of features are identified by Random Forest, most of which determine the severity of an accident, such as distance, temperature, humidity, and wind direction and some other factors are identified. random forest is also used as inputs to the model to improve accuracy. The classification results of fusion of Random Forest and Convolutional Network are higher than those of another algorithm like Ranom forest Voting Classifier. Random forest achieved 0.74 accuracy, 0.78 accuracy, and 0.79 detection.

Location-Based Accident Prediction Using Machine Learning [2] Applications that influence traffic accidents by various conditions.

Therefore, as a result of various investigations, it was found that conditions such as the age of the car model and weather conditions affect the occurrence of traffic accidents. So, they developed an application that accurately predicts traffic accidents depending on these factors.

The dataset is from his Kaggle dataset on road accidents in Bangalore. The system uses a support vector machine algorithm and the model accuracy is 88.6%. The purpose of such models is to predict what situations will lead to accidents and take preventive measures. EDA is an approach used to analyze data using visualization techniques and make predictions using static summaries and graphs.

Analyzing Traffic Accidents in India Using Data Mining Classification Algorithms [3] The algorithm with the lowest average difference between real and predicted error as mean absolute error and high accuracy is selected as the useful algorithm. They used his KNN algorithm is supervised learning approach uses classification as well as regression concepts. This is an easy algorithm that's having all available cases and evaluates K neighbors to classify new cases. In presence of different factors accuracy and error rate is calculated.

ANN classification algorithm is detected to be the best algorithm compared to other classification algorithms with a highest accuracy.

Traffic accident prediction model using data mining techniques [4] Traffic severity prediction using apriori and SVM techniques an application that helps predict traffic and road accident severity on the basis of some criteria or factors that are some types of vehicles, driver age, humidity, temperature conditions, and road conditions. SVM and a priori methods are used for accident prediction. We use SVMs (Support Vector Machines) to analyze how accident risks is there and Association algorithm like Apriori used to apply some mining techniques. provide a set of common items based on a given dataset.

The model is divided into some modules: Rule mining and risk prediction, graph plotting, and new data input. The datasets used in this study is from India's Open Government Data (OGD) platform. The dataset is from his Bangalore region from 2014 to 2017 and was used to develop the model. It includes details such as the date, time, location of the accident, and type of accident.

Table.1. Survey Table

Sr. No	Literature (Year)	Models/Algorithms	Accuracy
1.	[1](2021)	RFCNN Model	99.43%
2.	[2](2018)	Decision Tree	96.778%
		SVM	86.63%
		Naive Bayesian	85%
3.	[3](2021)	CNN	97.82%
4.	[4](2020)	Logistic Regression	97.6%

III. ACCIDENT PREDICTION DESCRIPTION

3.1 Materials

3.1.1. Dataset

Common datasets used for experiments are from Kaggle and various research institutes. This is a national car accident record covering his 49 states in the United States. Accident data includes records from February 2016 to June 2021. This dataset has approximately 4.2 million accident records.

3.1.2 Dataset Description

Table.2. Dataset Description

Features	Description
City	Shows the city in address field.
Temperature	Shows the temperature (in Fahrenheit).
Weather	Shows Weather condition (rain, snow, humidity, fog, etc.)
Distance(mi)	The length of the road extent affected by the accident.

Accidents prediction including some variety including real prediction investigation of accident hotspots, accident analysis, predicting accidents through causality extraction, or investigating the effect of humidity, fog, precipitation and other environmental factors on then analyzing and predict the which area is comes under red zone.

3.2 Algorithm

3.2.1 RFCNN

Random Forest is a model containing multiple decision trees. Build each tree using a random subset of the training data. The samples that make up the subset are drawn from the sorted training data. Therefore, some samples in the tree can be used multiple times. It also splits tree nodes considering random subsets of features. During the test phase, each decision tree in the forest makes predictions for a specific test sample. By using multiple trees, random forests achieve superior prediction accuracy by avoiding the overfitting to which single decision trees tend. Each decision tree is built using a random subset of features from the data. A convolutional neural network is a deep learning approach neural network model that processes complex data during computation.

3.3 Logistic Regression

It is commonly used for prediction and classification problems and it is used to detect categorical dependent variable with the help of independent variables. The output of logistic regression problem is between 0 and 1. It gives discrete values and binary such as 0 and 1, yes or no, true or false. Simply put, it detects the probability of an event occurring by filtering the data into a logical function.it is easier to implement and can efficiently train the model.

3.4 Decision Trees

This is the most commonly used supervised learning algorithm for prediction problems. A decision trees is like a pictorial representation of tree where each internal node represents its feature of dataset. They break down complex data into more manageable parts. This algorithm is based on the most important attributes/independent variables to form the most unique groupings possible. A decision tree starts with one main idea and then branches out based on consequences of decisions. On each decision step or node of this tree, used for classification. The main aim of creating this tree is to formed a training model that can use to detect efficiently any target values by applying some decision and rules from prior data.

3.4 SVM

A SVM is use labeled data known as supervised learning methods used for classification.

It uses a subset of training points in the decision function.it having two category one is linear and another is nonlinear SVM. when a simple linear SVM classifier applied the it works by making a straight line between two classes. It separates into two classes that means it differentiate one sides. number of features with a particular coordinate value is there. It is efficient with multi featured datasets and also useful as some decision function so that it makes memory efficient using some support vectors.

IV. MODELING AND ANALYSIS

The model is formed using the dataset having some features. This helps to understand the characteristics of driver behavior, road conditions, lighting conditions, weather conditions and many other characteristics. This helps the user to calculate safety measures to help avoid accidents. The way statistical procedures are based on directed graphs can be illustrated by comparing his two scenarios based on out-of-sample predictions. The model is run to identify statistically significant factors that can predict the probability of accidents and injuries that can be used to enforce and mitigate risk factors.

Here a study on traffic accidents is done by identifying some criteria most dangerous time to drive and percentages of traffic severity and accident in areas or other areas.
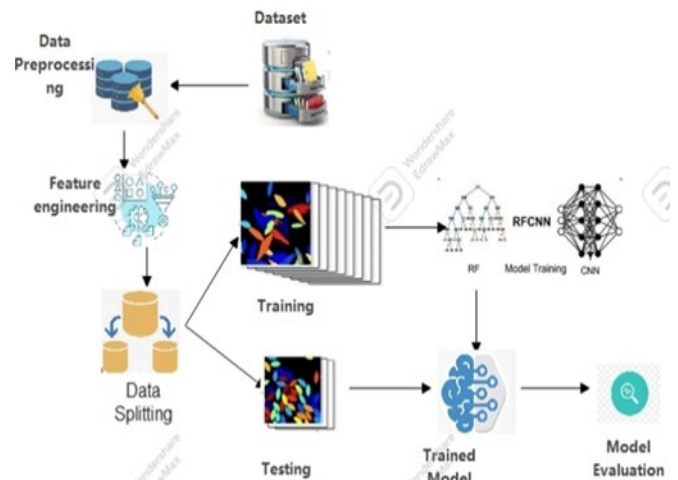

Fig.1. System Architecture of accident prediction model

V. CONCLUSION

This model is used further in future and its analysis is useful for various organization and provide safety as soon as possible.it is effectively used by governments to lower rate of traffic severity to stop accidents and provide road safety policies. Another area of this work is the development of mobile apps that allow drivers to choose driving routes. Calls to drivers through map services can also be implemented. In addition, it will inform you of the danger probability of the selected route with some hint about which area undergoes in danger.

This may be implemented in the future by service providers such as Uber, Ola, etc. This will also help us better monitor red zone areas and provide services as early. Also, given the risks arising from this model, better traffic safety instructions can be set along highways, reducing the workload of the human resources department.

REFERENCES

[1]. Viswanath, Dhanya and K, Preethi and R, Nandini and R, Bhuvaneshwari. (2021). "A Road Accident Prediction Model Using Data Mining Techniques.

[2]. P.Shivkumar,K.Shivkumar,.S.Vijay,B.R.Naveenkumar,Dr. A.Sathiya Sofia "Location based accident prediction model using machine learning",may-2022.

[3]. Suganya, E. and Vijayarani, S. (2017). "Analysis of road accidents in India using data mining classification algorithms.

[4]. K. Haleem, P. Alluri, and A. Gan, ''Analyzing pedestrian crash injury severity at signalized and non-signalized locations,''

[5]. Al Mamlook, Rabia and Kwayu, Keneth and Frefer, Abdulbaset. (2019). "Comparison of Machine Learning Algorithms for Predicting Traffic Accident''.

[6]. A. Thaduri, V. Polepally and S. Vodithala, "Traffic Accident Prediction based on CNN Model,

[7]. B. Sharma, V. K. Katiyar, and K. Kumar, ''Traffic accident prediction model using support vector machines with Gaussian kernel,''.

[8]. R. E. AlMamlook, K. M. Kwayu, M. R. Alkasisbeh, and A. A. Frefer, ''Comparison of machine learning algorithms for predicting traffic accident severity,'' in Proc. IEEE Jordan Int. Joint Conf. Electr. Eng. Inf. Technol. (JEEIT), Apr. 2019, pp. 272–276.

[9]. T. Beshah and S. Hill, ''Mining Road traffic accident data to improve safety: Role of road-related factors on accident severity in Ethiopia,'' in Proc. AAAI Spring Symp., Artif. Intell. Develop., vol. 24. Princeton, NJ, USA: Citeseer, 2010, pp. 1173–1181.

[10]. 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.

[11]. S. Ameratunga, M. Hijar, and R. Norton, ''Road-traffic injuries: Confronting disparities to address a global-health problem,'' Lancet, vol. 367, no. 9521, pp. 1533–1540, May 2006.

[12]. T. Beshah and S. Hill, ''Mining Road traffic accident data to improve safety: Role of road-related factors on accident severity in Ethiopia,'' in Proc. AAAI Spring Symp., Artif. Intell. Develop., vol. 24. Princeton, NJ, USA: Citeseer, 2010, pp. 1173–1181.

[13]. B. Yu, Y. T. Wang, J. B. Yao, and J. Y. Wang, ''A comparison of the performance of ANN and SVM for the prediction of traffic accident duration,''.

[14]. L. Breiman, ''Bagging predictors,'' Mach. Learn., vol. 24, no. 2, pp. 123–140, 1996. [43] R. E. Schapire, ''A brief introduction to boosting,'' in Proc. IJCAI, vol. 99, 1999, pp. 1401–1406.

[15]. L. Breiman, ''Random forests,'' Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[16]. C. R. Sekhar, S. Minal, and E. Madhu, ''Mode choice analysis using random Forrest decision trees,'' Transp. Res. Proc., vol. 17, pp. 644–652, Jan. 2016.

[17]. G. Biau and E. Scornet, ''A random forest guided tour,'' Test, vol. 25, no. 2, pp. 197–227, 2016.

[18]. Y. Zhang, H. Zhang, J. Cai, and B. Yang, ''A weighted voting classifier based on differential evolution,'' Abstract Appl. Anal., vol. 2014, May 2014, Art. no. 376950.