Road Accident Prediction using Deep learning

Malavika Prasad(*Author*)
Dept.of computer science and engineering
Mangalam college of engineering,
Ettumanoor,India
malavikaprasad21@gmail.com

Neenu Joseph(*Author*)
Dept.of computer science and engineering
Mangalam college of engineering,
Ettumanoor,India
neenualphonsajoseph202@gmail.com

Nandana K Saji(*Author*)
Dept.of computer science and engineering
Mangalam college of engineering,
Ettumanoor,India
nandanaksaji2952002@gmail.com

Eldhose K Paul(*Author*)
Dept.of computer science and engineering
Mangalam college of engineering,
Ettumanoor,India
eldhose.paul@mangalam.in

***Abstract*—A database of the traffic accidents was organized and analyzed, and an intersection accident risk prediction model based on different mechanical learning methods was created to estimate the possible high accident risk locations for traffic management departments to use in planning countermeasures to reduce accident risk. Using Bayes' theorem to identify environmental variables at intersections that affect accident risk levels, this study found that road width, speed limit and roadside markings are the significant risk factors for traffic accidents. Meanwhile, Naïve Bayes, Decision tree C4.5, Bayesian Network, Multi-layer perceptron (MLP), Deep Neural Networks (DNN), Deep Belief Network (DBN) and Convolution Neural Network (CNN) were used to develop an accident risk prediction model. This model can also identify the key factors that affect the occurrence of high-risk intersections, and provide traffic management departments with a better basis for decision-making for intersection improvement.**

Keywords— Byes' theorem, Deep Neural Network (DNN) and Convolution Neural Network (CNN)

I. INTRODUCTION (*Heading 1*)

A high accident risk prediction model is developed to analyze traffic accident data and identify them priority intersections for improvement. A traffic accident database was organized analyzed. An Intersection Crash Risk Prediction Model Based on Different Machine Learning methods for estimating potential high accident risk locations for traffic management have been developed department to use in planning countermeasures to reduce the risk of accidents. Using Bayes theorem identify environmental variables at intersections that influence the level of crash risk, this study found that the width of the road, the speed limit and the markings along the road are significant risk factors for traffic accidents.Meanwhile, Naïve Bayes (NBD), Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN) were used to develop the accident risk prediction model.

This model can also identify key factors that influence the occurrence of high-risk intersections, and provide operations management departments with a better basis for decision-making intersection improvement. Using the same environmental characteristics as high risk intersections for model inputs to estimate the level of risk that may occur in the future, which can be used to prevent traffic accidents in the future. In addition, it can also be used as a reference for future intersection design and environmental improvements.In practical applications, our proposed model can be used to predict probability (or "risk") accidents at different intersections by identifying similar environmental variables, ie it enables authorities to take practical steps to effectively reduce incidence and severity accidents together with the costs associated with such accidents. In addition, research results identify important environmental factors that influence the occurrence of traffic accidents. To effectively reduce the risk of accidents, in recent years traffic accident management agencies in countries around the world not only have established standards and operating procedures for road surveys, but also sought to develop accident risk analysis and forecasting methods. The it hoped that longitudinal crash data would be used to identify and classify high-risk ones intersections, allowing efficient prioritization of scarce resources to minimize frequency and severity of traffic accidents.

II. LITERATURE SURVEY

SI.NO	TECHNOLOGY	AUTHOR	ADVANTAGES	DISADVANTAGES
1	A road accident prediction model using data mining techniques	Dhanya Viswanath,Preethi K,Nandini R, Bhuvaneshwari R	It helps to identify key factory of traffic injury severity	Requires large database, more expensive
2	Accident avoidance system using IR transmitter	Adnan Bin Faiz, Ahmed Imteaj, Mahfuzulhoq Chowdhury	Uses the alarm pulses and vibration system as the first level of safety	The smartphone's sensors could provide false data sometimes
3	Android application for automatic Accident detection	Dnyanesh Dalvi, Vinit Agrawal, Sagar Bansod, Apurv Jadhav, Prof. Minal Shahakar	It is integrated with multimodal alert dissemination	Slow in responding
4	Accident detection and reporting system using GPS,GPRS and GSM technology	Md. Syedul Amin, Jubayer Jalil, M. B. I. Reaz	Capture the location of vehicle accident	Limited rate of data transfer
5	Real time traffic accident detection system using wireless sensor network	Hossam M. Sherif, Hossam M. Sherif, Samah A. Senbel	Long distance data collection and transmission	It cannot be used for high speed

Table 1: Literature survey of accident prediction model

III. RELATED WORKS

The rapid development and wide application of computer technologies, computer network technologies, multimedia and communication technologies, and the Internet of Things fields [1], has driven the recent development of intelligent road traffic management systems [2]. Li et al. The Internet of Things allows for the collection of various kinds of information through sensors [3], each of which represents an independent information source [4] from which data is collected at a certain frequency for categorization and analysis.Each independent information source would sense, measure,capture and transmit information anytime and anywhere.The development of advanced chip design and new materials have also increased the utility and longevity of such sensors [5], while also allowing for anti-interference, multi-mode, and self-adapting features [6].

These developments provide the technological basis for intelligent expressway management systems, integrating Internet of Things applications due to the introduction of mass information compatibility. High-speed wired and wireless networks have been integrated to create three-dimensional connections, ensuring the accuracy of data information, wider transmission bandwidth, higher spectrum utilization, more intelligent access, and more efficient network management [7]. The development of these advanced technologies mainly depends on NGN (Next Generation Network)communication network technologies and new wireless communication networks (3G, 4G, ZIGBEE) [8].Expressway construction and traffic is rapidly growing around the world, and the demand for social development is growing synchronously [9].

Improving the efficiency of existing expressway traffic infrastructure requires the effective collection and analysis of usage data [10]. As cars and individual drivers are increasingly linked to wireless transmissions,drivers demand increasingly sophisticated traffic information,allowing them to assess current local traffic and driving conditions, predict future conditions, and identify optimal driving routes [11]. Expressway traffic management agencies also need to effectively monitor highway conditions and coordinate timely emergency response including police, rescue and repair units [12].

The data to drive such coordination is sourced from sensor networks that monitor traffic and environmental conditions throughout the highway network. Such monitoring data can be used to improve and simplify signal control algorithms and traffic efficiency. Wireless sensor networks can be applied to control subsystems and guidance subsystems in the execution subsystem, and to improve signal controller function to implement the bus priority function of the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICCIDT - 2023 Conference Proceedings

intelligent transportation system [13]. Besides, the position sensor can help achieve functions such as energy-saving and emission reduction.

IV. METHOD

A. Baye's theorem

a) Bayes' theorem serves as the foundation for the Naive Bayes (NB) algorithm. Chiang (1995) suggested a whole data storage and analysis system for road traffic safety, including Bayes' theorem as the key analytical tool [7]. A known target variable's prior probability, which is frequently available through training samples, is assumed by NB. Furthermore, the participating attribute values are presumptively independent of one another given any target variable or dependent variable. Assuming that training materials have a set of attributes X = [[X1, X2,..., Xn]], X does not contain the attribute for the target variable, and C is the set of values for the target variable's attributes,[[C1,C2,...,Cm]].

P(C|X) denotes the likelihood that a given collection of X traits will be present for the target category C.

$$P(C/X)=(P(X/C)*P(C))/P(X) \qquad (1)$$

According to the Naive Bayes theory, if each feature is assumed to be independent of the others, then equation (1) becomes:

Where P(Xi/C) is the likelihood that feature Xi appears in a class Cm,Cm ∈ C,

$$P(C/X)=n_{i=1}P(X_i/C)P(C)n_{j=1}\ P(X_j) \qquad (2)$$

The prior probability of the class Cm,Cm C across the board is P(cm). For a given set of features, the classifier's output is the group with the highest probability. The denominator can be regarded as a constant because it is independent of C and the value of the features Xi is provided. The probability of each class Cm,Cm C is computed in equation (2) to yield the maximum class,which is argmax c $=P(C = c)\ n_{i=1}\ P(X = X_i \mid C = c)$ (3)
where argmax c is used to represent the function that provides the largest class.

B. Deep Neural Networks

A deep learning framework called a deep neural network (DNN) can be thought of as a neural network with numerous hidden layers (Neural Networks). In a neural network, artificial neurons are used to create a mathematical model that resembles a biological neural network. Neurons are typically arranged in layers, and connections are only made between neurons in adjacent levels. The first layer receives the input low-order feature vector, which is then transformed into a high-order feature vector by advancing the neurons over time. The number of categories is the same as the number of neurons in the output layer.In order to represent the likelihood that the input vector falls into the appropriate category, the output vector is a probability vector. The predicted calculation of one neuron and its output description are presented in Eq. (8), where a_{ij} is the jth neuron in the ith layer and W_i is the weight of the neuron's synapse, which connects the jth neuron in the ith layer with the kth neuron in the layer below (i.e. layer i-1) Neurons.

$$a^i_j = \sigma(\textstyle\sum_k a^{i-1}_k w^i_{jk})$$

Three layers make up the neural network layer: an input layer, an output layer, and a hidden layer sandwiched in the middle. The neural network was designed with the intention of simulating how human neurons function. The output of this layer (matrix multiplication) is the linear combination of the inputs from the previous layer(s), which cannot be separated from the linear connection if the activation function is not utilised. In order for neural networks to express real complex models, the non-linear activation function is employed to raise the non-linear factor of neurons [8].

Sigmoid, Softmax, tanh, ReLU, and ELU are frequentlyused activation functions. An event (an element in a sample space) is mapped by the loss function to a real number that indicates the event's opportunity cost or economic cost. The reduction of the loss function is the optimization objective. As a result, the loss function determines how well the neural network model performs and what the optimization's objective is.

C. Convolutional Neural Networks

Deep learning has recently piqued the curiosity of academics and researchers across all disciplines. As a deep learning technique, the convolutional neural network has grown in popularity across many scientific disciplines. In the domains of computer vision, image recognition, and speech recognition, CNN is a rapid and effective feed forward neural network that has shown great results.

In recent years, the CNN model was created as a road traffic accident prediction model for accurately predicting highway road traffic accidents, hence promoting the efficacy of prediction. In this study, the CNN model outperformed the classic back propagation neural network model in terms of accuracy and efficiency, with a prediction accuracy of 78.5%, 7.7%.

By converting the gradient of the accident data into a grey image that represents the weight of the traffic accident's characteristics, a deep learning strategy with a CNN model was proposed for predicting the severity oftraffic accidents. The grey image was then fed into a CNN model that predicted severity. The Leeds City Council examined the performance of this suggested CNN model using data on traffic accidents from 2009 to 2016 and found that it performed better than the K-nearest neighbour technique, logistic regression, gradient boosting neural network, and support vector machines.

V. SYSTEM MODEL

The degree of information clutter reduction (benefit degree) can be determined by dividing the "expected information entropy before being partitioned by the target variable" by the "expected information entropy before being partitioned by an attribute," and choosing the node attribute that can produce the greatest benefit. s1,s2,...sm: A finite set of samples Category "C:" (c1, c2,...,cm)the quantity of samples falling into a particular category Thenumber of samples that fall under a specific attribute value's (av) range (Ak)

S_{ij} : The number of samples for a particular category under a particular attribute value (av) of a particular attribute (Ak) (ci)

P_i : The percentage of the sample (si/S) that falls within a specific category

P_{ij}: The percentage of samples that match a particular attribute value (av) of a particular attribute (Ak) for a particular category.

Equation (4) calculates the expected information entropyprior to partitioning by the target variable I(s1,s2,...,sm), which represents the post-segmentation degree of entropy of the target variable of the training set (Target variable, Dependent variable). Ak: An attribute that contains the attribute values "a1, a2,..., am"

$$I(S_1, S_2,..., S_m) = -\sum_{i=1}^{m} p_i (p_i) \quad (4)$$

The sample ratio of the training sample divided by the attribute Ak is calculated by equation (5). As an illustration, the attribute "gender" has the following two attribute values: "7male, 3female," and the sample ratio is {7/10, 3/10}.

$$E(A_k)=\sum^{v}(S_{1j}+...+S_{mj})/S*I(S_1,S_2,...,S_m) \quad (5)$$

For a specific attribute value (av) (female) for attribute (Ak), equation (8) generates the information entropy, which is I (s1,s2,...,sm). Prior to being divided by an attribute variable, we multiply the respective sample ratios to obtain the anticipated information entropy.

$$I(S_{1j},S_{2j},..., S_{mj} = -\sum_{i=1}^{m} p_{ij} (p_{ij}) \quad (6)$$

Gain (Ak) for a specific attribute node is obtained bydeducting E (Ak) from I (s1,s2,...,sm).

$$Gain (A_k) = I (S_1, S_2,..., S_m) - E (A) \quad (7)$$

VI. ARCHITECTURE

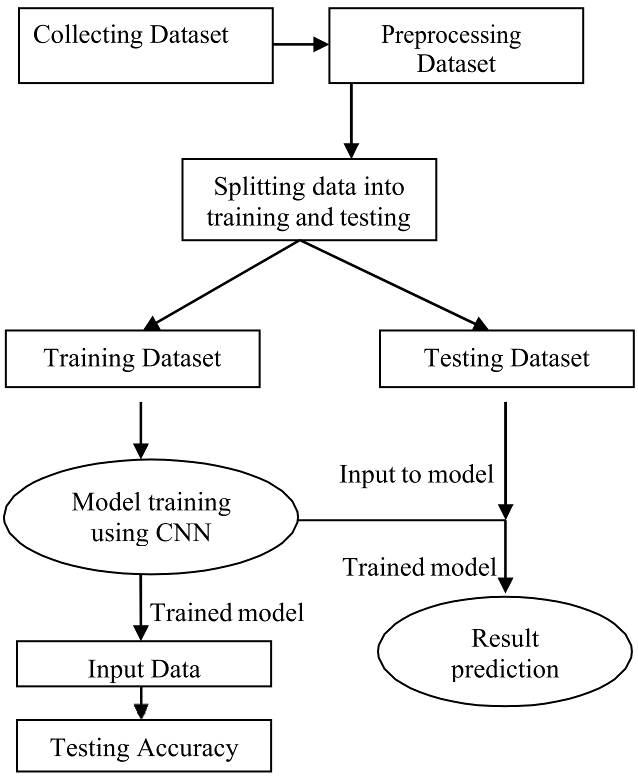

Fig1: Architecture of road accident prediction model

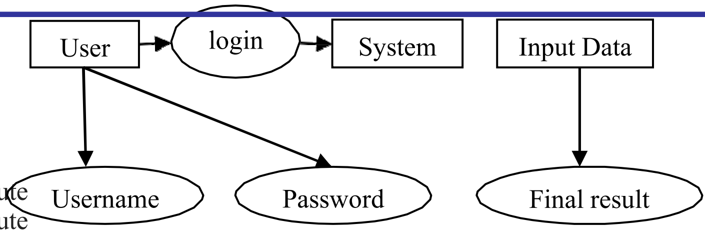

Fig 2: User Data flow diagram

B. *Level 0-DFD*

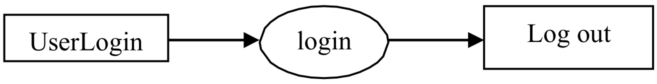

Fig 3: User-login Data flow diagram

VII. RESULT

Predicting the likelihood of accidents at particular junctions is the goal of accident risk analysis. The danger level of each intersection is determined based on the number of accidents and fatalities in the past. In order to estimate the level of accident risk at crossings when accidents have not yet happened, a risk prediction model for intersections is built by identifying the important environmental elements that influence the occurrence of accidents at crossroads.

VIII. CONCLUSION

This study analyses traffic accident data and identifies priority intersections for improvement using a high accident risk prediction model. There has been a significant increase in pedestrian injuries, as well as fatalities, over the past few years. For accident data for provincial highway intersections, risk grouping in terms of CBI was carried out. Different mechanical learning techniques were then employed to create a prediction model for high-risk intersections. The findings indicate that environmental factors including road width, the posted speed limit, and the existence of roadside markings are important indicators of the likelihood of an accident. It was simpler to pinpoint the environmental characteristics of low- and medium-risk crossings based on the frequency of accidents there. The relative lack of data hurt prediction accuracy for high-risk accident intersections, while decision tree rules and detection models were shown to offer respectable prediction accuracy for clusters of high-low and high-medium risk intersections. Additionally, it was discovered that the DBM model performs best for model training with unbalanceddata, whereas NB performs best for intersection risk prediction. The findings of this study can serve as a guide for traffic management organisations to reduce the probability of accidents at intersections. This study will aim

A. *Level 1-DFD*

to provide a platform for high-risk accident analysis and prediction based on intersecting environmental elements. The following objectives will be accomplished by data collecting and analysis based on the locations of traffic accidents:

1. This platform's system can combine and analyse traffic information about the GIS layer and accident data, thus understanding the site of the accident as a whole. Afterward, by examining the impact of environmental elements at the scene of the accident and its causation multiple intersections allow us to create useful enhancement approaches as a guide for upcoming intersection design and improvements to the environment.

2. Use predictive models to estimate the likely locations of high-risk accidents to allow traffic management authorities to better prevent high-risk road accidents or serious casualties.

REFERENCES

[1] Intelligent Traffic Accident Prediction Model for Internet of Vehicles With Deep Learning Approach," by Da-Jie Lin, Mu-Yen Chen, Hsiu-
Sen Chiang, and Pradip Kumar Sharma, Senior Member, IEEE, 2021

[2] A road accident prediction model employing data mining technique, Dhanya Viswanath, Preethi K, Nandini R, and Bhuvaneshwari R, 2021

[3] "Accident Avoidance System Using IR Transmitter," IJRASET, Volume 5 Issue IV, April 2017, by Miss Priyanka M. Sankpal and Prof. P. P.

[4] Android Application for Automatic Accident Detection, 2017; Dnyanesh Dalvi, Prof. Minal Shahakar, Vinit Agrawal, Sagar Bansod,Apurv Jadhav.

[5] "Real time traffic accident detection system employing wireless sensor network," Md. Syedul Amin, Jubayer Jalil, and M. B. I. Reaz (2014).

[6] "Real time traffic accident detection system employing wireless sensor network," Md. Syedul Amin, Jubayer Jalil, and M. B. I. Reaz (2014).

[7] Q. Cai, M. Abdel-Aty, J. Yuan, J. Lee, and Y. Wu, "Real-time crash prediction on expressways using deep generative models," Transp. Res. C, Emerg. Technol.,

[8] Data Mining Concepts and Techniques, J. Han and M. Kamber Morgan Kaufmann, San Mateo, California, USA, 2001, pp. 284–287

[9] Deep spatiotemporal graph convolutional network for traffic accident prediction, L. Yu, B. Du, X. Hu, L. Sun, L. Han, and W. Lv, Neurocomputing, vol. 423, pp. 135–147, Jan. 2021.

[10] L. Zheng and "A unique method for real-time crash prediction,"T. Sayed at intersections with signals," Transp. Res. Emerg, C. Technology, volume 117, 2020 August, Art. No. 102683.

[11] "Analysis of real-time crash risk for expressway ramps using traffic, geometry, trip generation, and socio-demographic variables," Accident Analysis and Prevention, vol. 122, no. 1, January 2019, pp. 378–384. L. Wang, M. Abdel-Aty, J. Lee, and Q. Shi.

[12] "Crash prediction based on traffic platoon features utilising floating car trajectory data and the machine learning approach," Accident Analysis and Prevention, vol. 133, Dec. 2019, Art. no. 105320, by J. Wang, T. Luo, and T. Fu.

[13] Key feature selection and risk prediction for lane-changing behaviours based on vehicles' trajectory data: T. Chen, X. Shi, and Y.
D. Wong, Accident Analysis and Prevention, vol. 129, pp. 156–169, August 2019.