International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V70I2P217

Volume 70 Issue 2, 143-150, February, 2022

Research Article

Prediction of Road Accident Using Artificial Neural Network

Mayura Yeole[1], Rakesh Kumar Jain[2], Radhika Menon[3]

[1]*Research Scholar, JSPM's RSCOE, Tathawade and Faculty at PCCOER, Pune, Maharashtra, India.*
[2] *Principal, JSPM's RSCOE, Tathawade, Pune, Maharashtra, India.*
[3] *Professor & associate Dean Research, Dr DYPIT, Pimpri, Pune, Maharashtra, India.*

[1]y.mayura@gmail.com, [2]jainrb20@gmail.com, [3]radhika.tharoor@gmail.com

Abstract *- A traffic accident is one of the biggest road safety concerns. Predictability models are used to explain the relationship between highway calamity and applicable parameters such as volume of traffic, conditions of roads and environmental concerns. For this work, a prediction model has been developed using ANN (Artificial Neural Network) and compared with the multiple linear regression techniques. The dataset collected and used for this study was for a mixed traffic flow of Pimpri Chinchwad Municipal Corporation (PCMC), Pune, Maharashtra, India. Weather condition, Vehicles loading condition, number of lanes and different time slots of accidents are considered as input parameters. Along with these, Traffic signals, speed breakers, the road intersects were also reflected as input parameters. For this, the accident data were collected for a period of six years ranging from 2014 to 2019. A recorded number of 887 major, as well as minor accidents, were considered for this study. For the ANN model, the available accident dataset was divided into three parts. In this study, 70% of data were used for model preparation, 15% for Training and Testing of the model and the remaining 15% for Validation of the model. Results show that the prediction model using ANN gives excellent accuracy. In this study, more emphasis has been given to the actual parameters which are responsible for accidents caused by mixed traffic flow.*

Keywords — *Regression, Performance, Road accident, Prediction, Artificial Neural Network.*

I. INTRODUCTION

The provision of safety during the transport of goods and passengers from origin to destination is an important function of the transportation system. A healthy transportation system always impacts the economic growth of the country directly. As per the ministry of the "Road and highway", the number of registered vehicles is increasing every year. This increasing number is the main cause of road accidents. Based on details available with a transport research department under the ministry of road transport & highway of India, more people were died because of road accidents during 2016 as judged against the previous year [9]. "Accident prediction model" can be applied to generate a relationship between accidents and various variables like traffic flow, number of registered vehicles, road condition, weather condition, etc. To improve road infrastructure safety, accident prediction tools are necessary as accident prediction models. In India, mixed traffic runs on road networks that move from one place to another [17, 21]. Figure 1 shows the road accident classified according to the type of Vehicle for Maharashtra in 2018 [6].

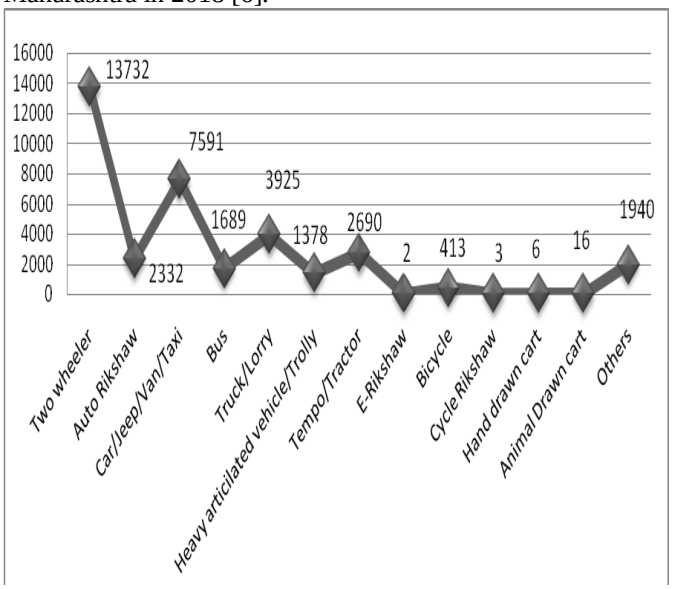


Figure 1: Road Accident Classified According to a Type of Vehicle for Maharashtra in 2018

Based on the statistics data available during the year 2014 to 2019 from the Ministry of Road Transportation Department, the registered motor vehicles ('000) in India are shown in Figure 2. Registered transport is increasing annually in India; therefore, road safety is a major concern since it affects the bulk of the public in the countryside who are heavily involved in road transport [9].

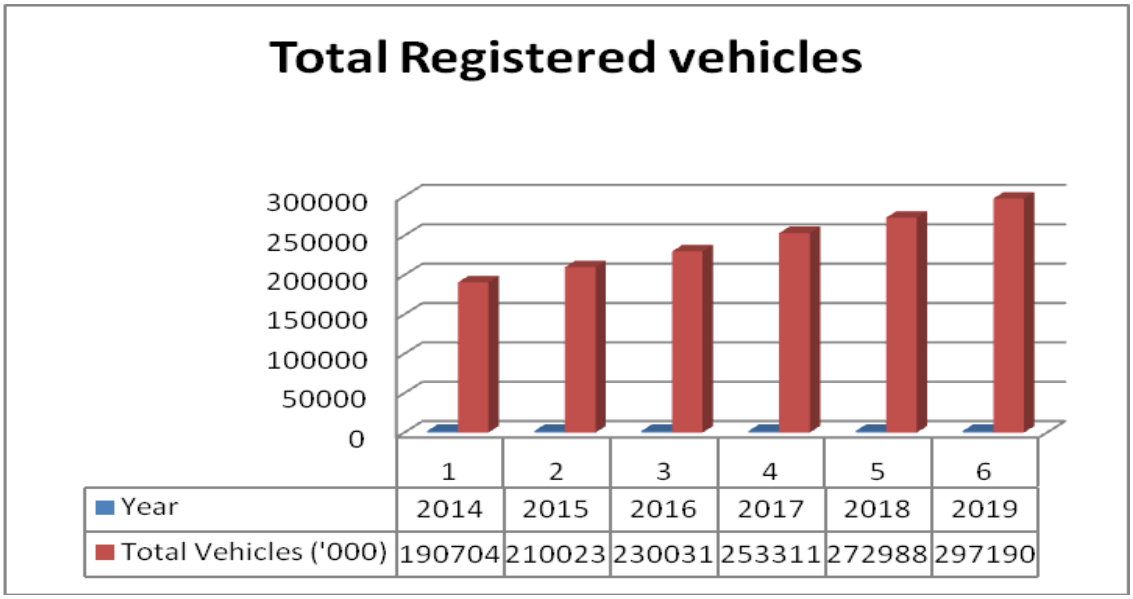
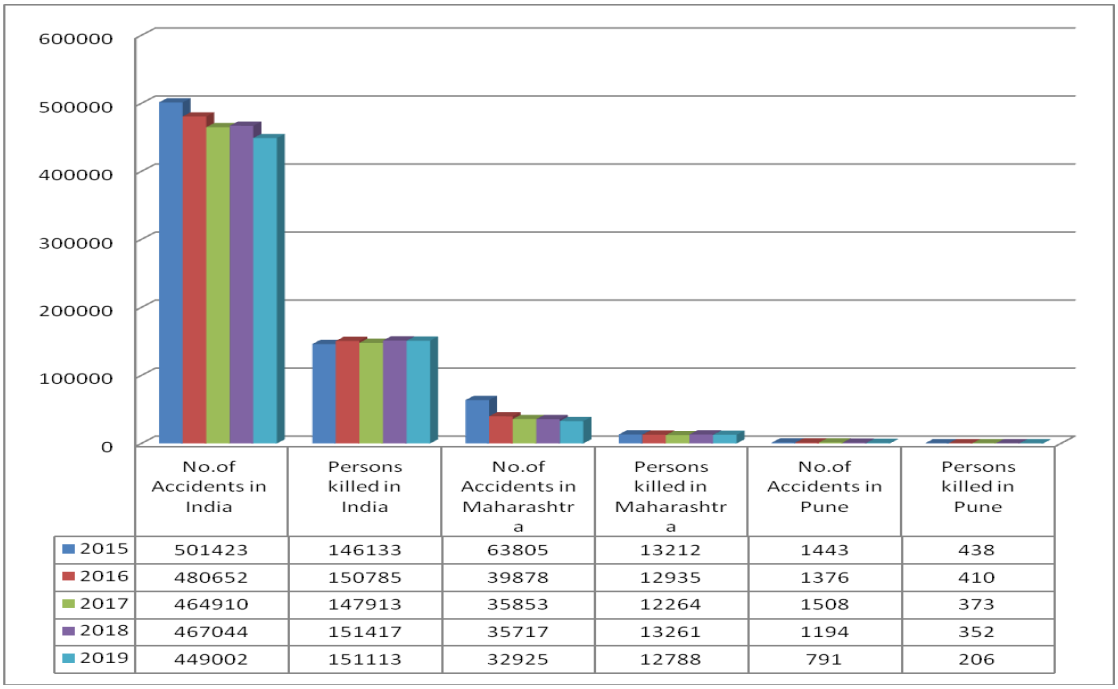


Fig. 2 Registered Public Vehicles in India

Along with the various latest manufacturing vehicles on the road, the level of accidents is also increasing. These circumstances require to be inspected, and research needs to be done to identify the major risk factors. In addition, it is important to ensure that the present transport structure in India is in a position to provide a wide range of vehicle users in the country [9]. Elements affecting increased accident threat include personal or behavioural characteristics like age and gender of the driver, use of seat belt, drug consumption during driving; atmospheric condition; condition of roads, illumination conditions, impact path, the position of a vehicle, in addition to the mechanical features of the vehicle itself like age of car and type of body [1,3,11].

Day by day, people are getting aware regarding traffic safety, and the efforts taken by the transport department result in less number of accidents in the year 2016 as compared to 2015 [9]. But till the number of accident is more including more fatalities. The road crash scenario in India is becoming worse year after year as the number of fatal road accident cases worsens. As a result, India's fatality rate is one of the deadliest among other budding countries around the globe [9]. Figure 1.3 shows the comparative accident count and fatalities in India, Maharashtra and Pune for the years 2015 to 2019.



	No.of Accidents in India	Persons killed in India	No.of Accidents in Maharashtra	Persons killed in Maharashtra	No.of Accidents in Pune	Persons killed in Pune
2015	501423	146133	63805	13212	1443	438
2016	480652	150785	39878	12935	1376	410
2017	464910	147913	35853	12264	1508	373
2018	467044	151417	35717	13261	1194	352
2019	449002	151113	32925	12788	791	206

Fig. 3 Accident and fatalities for the year 2015 to 2019

A. Overview of Road Transport In India

Heterogeneous traffic with cars and non-autos is commonly available in various budding countries and territories, for instance, China, India and Indonesia [22]. India, over the past two decades, has seen the expansion of its road network by almost 40%. At the same time, however, the number of its vehicles grew by an astonishing 158%. This is a very shocking condition when it comes to traffic management, which promotes not only an increase in the digit of accidents and fatalities but also unnecessary messes up a situation on the sides of the road [8]. One more time, it is clear that the impact of these accidents and fatalities are on the weakest part of society. Most of them are pedestrians, bicyclists, and bikers [8]. When these groups of people face such situations, they are made extra helpless due to further extension of health emergencies and loss of income and prospects to earn.

Maharashtra, India's most developed state, holds a supreme road network. This road network joins more than 99% of villages. Maharashtra has more than 7,500 km of National Highways, 35,000 km of State Highways and an amazing network of Expressways [8]. The road network is important for the economic development of any state and nation. Various local agencies like Public Works Department (PWD), Cantonment Boards, Municipal Council / Municipal Corporation / Nagar Panchayats, Maharashtra State Road Development Corporation (MSRDC), Maharashtra Industrial Development Corporation, MIDC), City and Industrial Development Corporation (CIDCO) etc., take care of road infrastructure in Maharashtra (Highway Police, Maharashtra). Pimpri Chinchwad region has a higher number of mixed vehicles, which leads to a higher risk of highway accidents and fatalities. The price of these victims and injuries has a large effect on the socio-economic development of one of the biggest industrial belts of India as well as Asia.

B. Literature Review

Many factors are responsible for any road accident, but the major affecting factors are a behaviour of a driver, unsafe driving practices, use of mobile phone, and consumption of drugs by the driver during driving with this very few and negligent driver's dangerous driving [1,3,11]. Motor vehicle failures like tire failure, brake failure, or steering failure contributed to a small proportion of accidents [10]. Vehicle failure results in a severe road accident; therefore, maintenance of the vehicle is mainly vital to become confident that the vehicle is in a safe and sound situation. For preparing the accident prediction model, several methods like Statistical analysis using regression, Artificial Neural Network & Fuzzy techniques are used.

Using "logistic regression", the relationships between type of road, the flow of traffic, and road accidents were identified and observed that the road type has a significant impact on the number of accidents [23]. The parameters like lane width, Speed of vehicles, Road connectors, traffic flow, and the number of the entrance are focused on by various researchers [2]. Different developed countries were considered for model preparation with different parameters such as geometrics characteristics of the road, characteristics of traffic and environmental parameters [2]. At Tehran-Qom throughway, accident data for a span of three years with parameters such as "average daily traffic volume, percentage of heavy vehicle, average speed and road condition index" are considered as input variables for ANN and log-normal models, which were used to determine the number of crashes [5]. Researchers concluded that for freeways, the ANN model is more proficient at determining the count of road crashes, and normal Speed of vehicles and normal daily traffic is the most efficient constraints [5]. Ogwueleka used a Multi-layer perceptron ANN model having a linear activation function an algorithm of backpropagation with a number of vehicles, accidents, and population as model parameters [4]. Different parameters such as years, the extent of horizontal curvature, highway sections, annual average daily traffic (AADT), section length (km), the extent of bend, traffic crashes with intense vehicles, and traffic accidents that occurred in a different season were considered for model preparation [14]. Data mining software like "Waikato Environment for Knowledge Analysis" (WEKA) was used by a few researchers to construct the ANN model [19]. Poisson regression was reflected as a typical and capable probabilistic model. For over-dispersed data, it fails and requires an alternate solution. During the examination of the suitability of the Negative binomial, it was noticed that extra scattering gives incorrect results if fundamental errors were coincided [20]. The controlling aspects like type of road, day, daylight situation, type of work zone, and climate functioned for mixed linear regression with a lognormal distribution model [18]. A fuzzy neural network has been used to predict road accident occurrence as one of the most common ways to predict accidents. The used fuzzy neural network model was established for Harbin city of China using data set with annual average daily traffic (AADT), a width of the lane (LW), Section length (SL) as input parameters [12]. For nonlinear independent and dependant variable correlation, the fuzzy logic system is preferable [15]. To evaluate accident statistics for the National Freeway 1 in Taiwan, a "negative binomial regression" model and an "artificial neural network" model were developed. The results give an idea that the ANN is reliable for accidents happening [13].

II. MATERIALS AND METHODS

A. Study Area Selection

As per the news published in Hindustan times on Friday, Feb.15, 2019, and Times of India dated Dec.29, 2017, Pune records the highest road accident emergencies and highest road fatalities across the state. For this study, data consider from India – Maharashtra State –Pune-Pimpri Chinchwad region, and on NH 48. Route of 17.5 km starting from Chinchwad station towards old Mumbai highway to Somatane phata near Talegaon. Figure 4, retrieved from Google map, shows the study area.

B. Features of the study area

Pimpri Chinchwad, an identical city, situated in the periphery of Pune, occurs to be one of the major industrial areas in Asia. Chinchwad railway station and Begadewadi railway stations are very near to the study area. This route is 17.5 KM. Maharashtra Industrial Development Corporation (MIDC) of Bhosari is situated at 4.9KM from Chinchwad railway station, and Talegaon MIDC is at 6.6.KM from Somatane phata. On the route or nearby the route, there exists a famous Bhakti Shakti Udyan, many schools, Ghorwadeshwar a historic place, A Gahunje cricket stadium, Prati Shirdi Sirgaon. There are many automobile companies and educational institutes located in this area resulting in a huge mixed traffic flow.

Fig. 4 Study area from Chinchwad station to Somatane phata NH 48 (Google, 2021)

C. Various characteristics/parameters considered

Few road intersections are there for the selected route. The parameters considered for this study are the driver's age and his eyesight, weather condition, time of the accident, vehicle loading condition, number of lanes and speed limit. The accident statistics data was collected from governing authority for the year 2014 to 2019, and a total of 887 major as well as minor accidents were considered. In this paper, a "multiple regression analysis" and "Artificial Neural network" were used to predict an accident. Table 1 gives information about Input (independent) and output (dependent) parameters along with their descriptions.

Table 1: Input and output parameters

Input & Output parameters	Descriptions
X1	A number of accidents are caused in sunny/ clear weather. Conditions.
X2	A number of accidents are caused by rainy weather conditions.
X3	A number of accidents are caused by a normally loaded vehicle.
X4	A number of accidents are caused by an overloaded Vehicle.
X5	A number of accidents caused by Empty Vehicle.
X6	A number of accidents occur due to improper signals.
X7	A number of accidents occur due to bad conditions on the road.
X8	A number of accidents occur due to improper speed breakers.
X9	The number of accidents occurs at road intersect.
X10	Number of accidents occurs in time 6:00 to 12:00 hrs(Day)
X11	Number of accidents occurs in time 12:00 to 18:00 hrs(Day)
X12	Number of accidents occurs in time 18:00 to 24:00 hrs (Night)
X13	Number of accidents occurs in time 00:00 to 06:00 hrs (Night)
X14	Two lanes
X15	More than two lanes
Y1	Total number of accidents
Y2	Deaths

III. RESULTS AND DISCUSSION

A. Model Development

The intention of this study is to extend a precise accident predictive model for future prediction of accident occurrences in the Pimpri Chinchwad region of Maharashtra State, India, using crash data from 2014 to 2019 using regression and artificial neural network models. Here, the available accident dataset was divided into three parts, 70% of data was used for model preparation, 15% for Training and Testing of the model and the remaining 15% for Validation of the model. The obtained results from regression and ANN models are presented in this section.

a) Multiple Linear regression Model

Multiple linear regressions (MLR) is a statistical technique that uses more descriptive parameters to forecast the result of a dependent variable. The purpose of this is to form the linear relationship between the independent and dependant parameters. For this model preparation, the independent and dependent parameters, along with their descriptions, have been mentioned in table 1.

Following results were noted by accident prediction model prepared using multiple linear regression,

Table 2. Results received from Regression Statistics

Regression Statistics	
Multiple R	**0.938546**
R Square	0.880568
Adjusted R Square	0.404342
Standard Error	29.41889
Observations	18

Analysis of variance (ANOVA) is a statistical technique that is used to check if the means of two or more groups are significantly different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples.

Table 3. Analysis of variance (ANOVA)

	df	***SS***	***MS***	***F***	***Significance F***
Regression	15	4789.27	319.28	44.554	0.022
Residual	2	14.33	7.166		
Total	17	4803.61			

Where,
df = degree of freedom,
SS= Sum of Squares,
MS= Means of Squares,
F= Overall F test for the null hypothesis,
Significance F = the significance associated with P-Value.

Table 4. Coefficients and P values

	Coefficients	**Standard Error**	**t Stat**	**P-value**
Intercept	-13.0563	3.849	-3.392	0.0770
Number of accidents caused in sunny/ clear weather conditions	0.1093	0.0891	1.226	0.344
Number of accidents caused in rainy weather conditions	-0.257	0.1705	-1.506	0.270
Number of accidents caused by normally loaded Vehicle	1.192	0.1131	10.541	0.008
Number of accidents caused by overloaded Vehicle	0.921	0.207	4.432	0.047
Number of accidents caused by Empty Vehicle	4.207	0.445	9.443	0.011
Number of accidents occurs due to improper signals	1.854	0.233	7.941	0.015
Number of accidents occurs due to bad condition of the road	0.491	0.115	4.251	0.051
Number of accidents occurs due to improper speed breaker	-0.562	0.404	-1.391	0.298
Number of accidents occurs at road intersect	2.728	0.572	4.767	0.041
Number of accidents occurs in time 6 :00 to 12:00 hrs(Day)	-1.773	0.296	-5.983	0.026
Number of accidents occurs in time 12:00 to 18:00 hrs(Day)	-0.971	0.140	-6.920	0.020
Number of accidents occurs in time 18:00 to 24:00 hrs(Night)	-3.827	0.521	-7.334	0.018
Number of accidents occurs in time 00:00 to 06:00 hrs(Night)	-2.015	0.360	-5.585	0.030
Two number of lanes	0.610	0.148	4.100	0.054
More than two number of lanes	0.494	0.108	4.576	0.044

The multiple regression model for a number of accident predictions has been developed passing through the set of points as,

$Y=-13.05+0.109X_1-0.257\ X_2 + 1.192X_3 +0.921X_4 +4.207X_5 +1.854X_6+0.491X_7-0.562X_8+2.728X_9-1.773X_{10}-0.971X_{11}-3.827X_{12} -2.015X_{13}+0.610X_{14}+0.494X_{15}$

Here, Y is the output (dependant) variable, and X is the input (independent) variable.

Table 5. Actual Versus Predicted Number of Accidents

Observation	*Actual No. of Accidents*	*Predicted Total no. of accidents*	*Residuals*
1	42	42.25795	-0.257
2	65	64.0921	0.907
3	58	57.73278	0.267
4	75	75.51077	-0.510
5	46	46.46861	-0.468
6	52	52.35287	-0.352
7	75	74.62118	0.378
8	65	65.1654	-0.165
9	45	43.73169	1.268
10	63	62.77193	0.228
11	22	22.08852	-0.088
12	54	55.16048	-1.160
13	58	57.99492	0.005
14	40	39.9867	0.013
15	49	49.74013	-0.740
16	28	25.95482	2.045
17	31	30.28615	0.713
18	19	21.08301	-2.083

The multiple regression prediction models for the number of fatalities in a road accident has been developed passing through the set of points as,

$Y= 7.36 + 0.033\ X_1- 0.291\ X_2 -0.064\ X_3- 0.254\ X_4 - 0.392\ X_5 -0.314\ X_6 -0.072\ X_7 + 0.052\ X_8 + 0.108\ X_9 -0.191\ X_{10}+0.104\ X_{11} -0.016\ X_{12}+0.186X_{13}+0.168X_{14}+0.026\ X_{15}$

Here, Y is the output (dependant) variable, and X is the input (independent) variable.

Table 6. Actual Verses Predicted Deaths

Observation	*Actual Deaths*	*Predicted Deaths*	*Residuals*
1	5	4.685	0.314
2	10	8.893	1.106
3	6	6.363	-0.363
4	5	5.133	-0.133
5	3	3.465	-0.465
6	6	6.230	-0.230
7	3	2.819	0.180
8	4	3.828	0.171
9	6	5.216	0.783
10	5	5.268	-0.268
11	6	6.561	-0.561
12	5	6.188	-1.188
13	5	5.227	-0.227
14	12	11.963	0.036
15	8	7.870	0.129
16	5	5.288	-0.288
17	8	6.154	1.845
18	6	6.840	-0.840

b) Model Using ANN

Using an "artificial neural network", the results obtained for an accident prediction model are mentioned below. For preparing this prediction model, 24 hidden layers were considered based upon trial & error to get the best fit R-value. Here 16 are total input parameters & 2 output parameters were considered. The sigmoid function is used as an activation function.

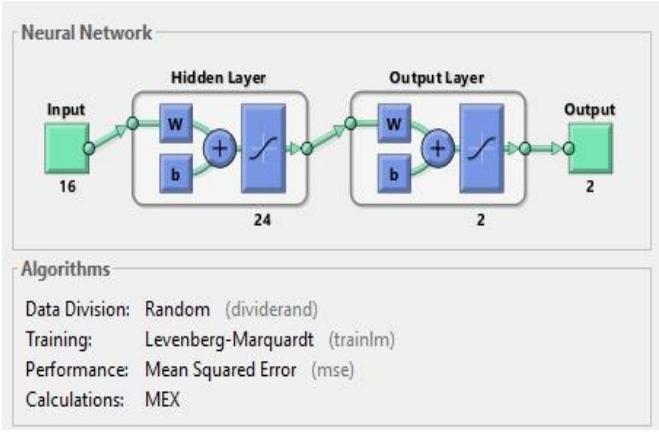


Fig. 5 Neural network diagram

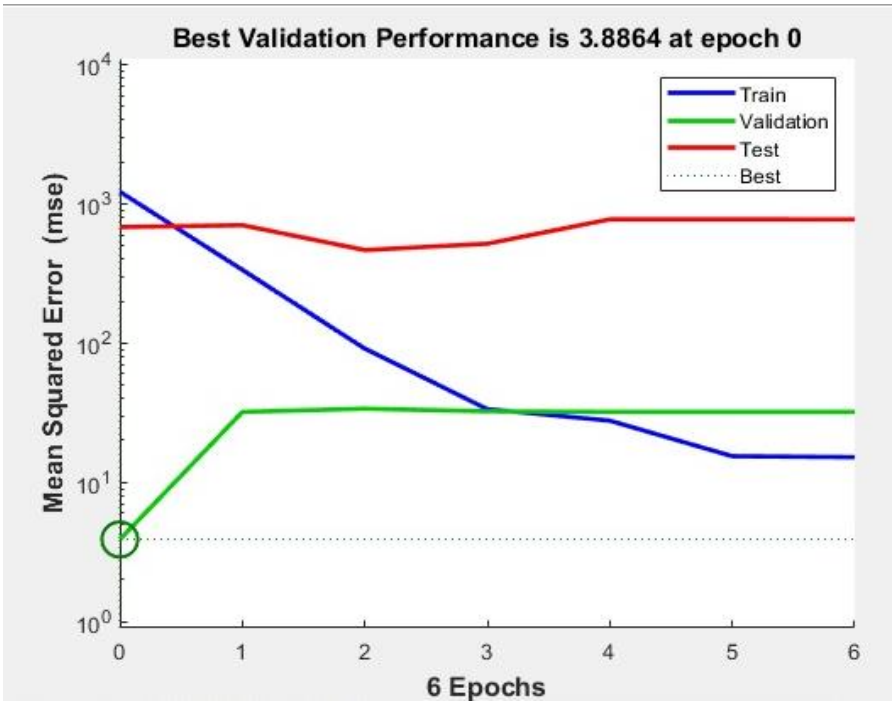


Fig. 6 Validation Performance

From this graph, the best validation performance is 3.8864 at epoch 0

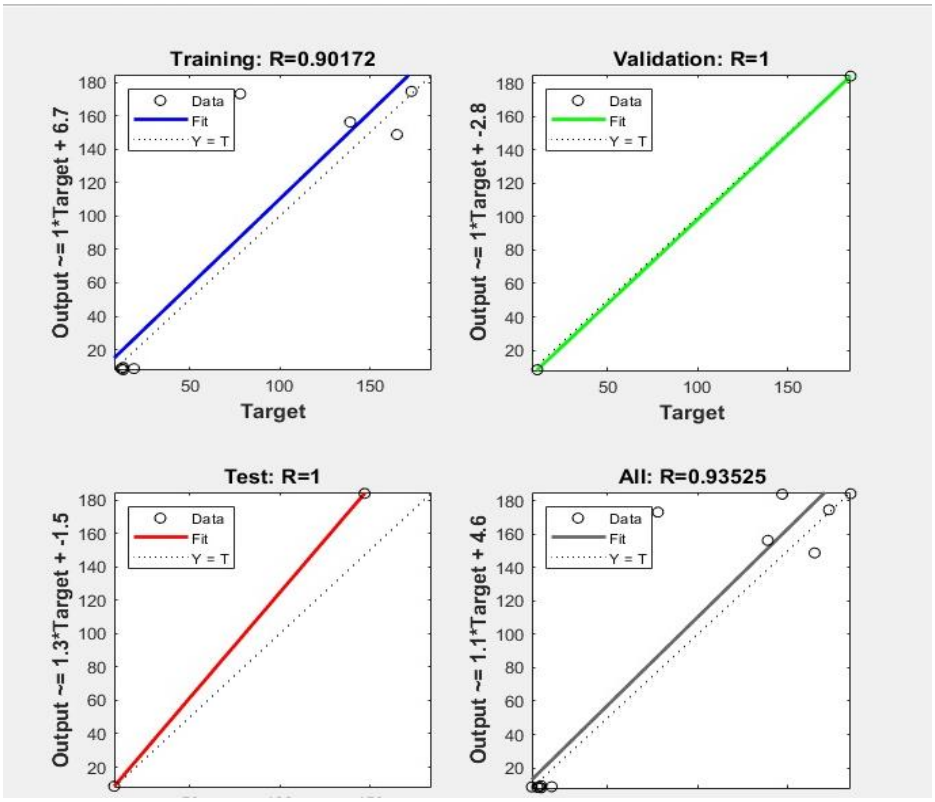


Fig. 7 Overall Performance of a Model

The available accident dataset was divided into three parts in this study. 70% of data was used for model preparation, 15% for Training and Testing of the model and the remaining 15% for Validation of the model. From the following values obtained,

The root mean square (R) for Training of a network was 0.9012,

Root mean square (R)) for the Testing of a network was 1,

The root mean square (R) for the Validation of a network was 1,

For the overall performance of the model, the Root mean square (R) was 0.93525

IV. CONCLUSION

In this study, the ANN model was compared with the multiple linear regression model. The relative performance of both models showed that the proposed ANN model performed better than the Multiple linear regression model in terms of accuracy. The accident and fatality prediction models are developed using multiple linear regression as well as ANN techniques. The comparative study shows that the performance of said prediction model, prepared using multiple regressions, is 88%, while using ANN, the model gives an overall 93% performance. So ANN Model gives very good accuracy for the prediction of an accident in the case of mixed traffic flow with the given input variables. By using this model, precautionary measures will be implemented at that location which will avoid accidents or reduce the impact of an accident. Also, using this model, the parameter which is actually responsible for an accident is identified easily. ANN took less time and was an excellent tool for the preparation of the prediction model.

The study reveals that a maximum number of accidents occurs during day time. The road intersections road conditions along the inviolacy of traffic signals are other major causes of the accident. During the study, it has been observed that maximum patches were there on the road, during the rainy season, potholes developed were impacted; the Speedy vehicles are also one of the major causes of accidents. This study is useful for a similar type of industrial region with mixed traffic flow.

This prediction model can be further enhanced on the real-time streaming data from the entire transportation medium with the Onboard diagnostics (OBD) devices. Also, more data frames regarding the environmental condition, traffic data, characteristics of different types of vehicles and road sensor data for mixed traffic flow shall be included for better accuracy in the prediction relevant to the exact time and surrounding condition.

REFERENCES

[1] Aworemi, Joshua, R., Abdul, A., Ibraheem, A., Olabode, & Segun, O., Analytical study of the causal factors of road traffic crashes in southwestern, Educational Research ., 1(4) (2010) 118-124.

[2] Bangaram Naga Kiran et al., A Review of Road Crash Prediction Models for Developed Countries , American Journal of Traffic and Transportation Engineering; 2(2) (2017) 10-25.

[3] Emmanuel Kofi Adanu and Steven Jones ., Effects of Human-Centered Factors on Crash Injury Severities, Journal of Advanced Transportation, Article ID 1208170, (2017) 11 pages https://doi.org/10.1155/2017/1208170.

[4] F. N. Ogwueleka et al., An Artificial Neural Network Model for Road Accident Prediction: A Case Study of a Developing Country" Acta Polytechnica Hungarica., 11(5) (2014) 177 -197.

[5] Fatemeh Bagheri Khalili et al., Variable Efficiency Appraisal in Freeway Accidents Using Artificial Neural Networks -Case Study", CICTP 2012 © ASCE., (2012).

[6] Highway traffic book

[7] https://data.gov.in/search/site?query=road+accident

[8] Report Published by accident research cell, Maharashtra. https://highway police.maharashtra.gov.in/en/

[9] https://morth.nic.in › sites › files › Road_Accidednt

[10] J. R. Treat, N. S. Tumbas, S. T. McDonald et al.., Tri-level study of the causes of traffic accidents: final report, vol. 1: causal factor tabulations and assessments, DOT HS-805 085, Indiana University: Institute for Research in Public Safety, Bloomington, Ind, USA., (1979).

[11] Jonathan J.Rolison et al., What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records, Elsevier Accident Analysis & Prevention, 115 (2018) 11-24.

[12] Lai Zheng & Xianghai Meng., An Approach To Predict Road Accident Frequencies: Application of Fuzzy Neural Network" Submitted to the 3rd International Conference on Road Safety and Simulation, September , Indianapolis, USA., (2011) 14-16.

[13] Li-Yen Chang., Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network" Safety Science, Elsevier 43 (2005) 541–557.

[14] M. Y. Çodur et al., An Artificial Neural Network Model for Highway Accident Prediction: A Case Study of Erzurum, Turkey" Promet – Traffic & Transportation, 27(3) (2015) 217-225

[15] Mohammed Gaber et al., Traffic Accidents Prediction Model Using Fuzzy Logic: Aswan Desert Road Case Study" Journal of Engineering Sciences Assiut University Faculty of Engineering ,45(1) (2017) 28 – 44.

[16] N. R. Vatti, P. L. Vatti, R. Vatti and C. Garde., Smart Road Accident Detection and communication System, 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), (2018) 1-4 doi: 10.1109/ICCTCT.2018.8551179.

[17] Nabanita Roy et al.,Effect of Mixed Traffic on Capacity of Two-Lane Roads: Case Study on Indian Highways 10th International Scientific Conference Transbaltica 2017, Procedia Engineering 187 (2017) 53 – 58.

[18] Shan XUE and Jinxian WENG., Analysis of Uncertainty Associated with Response Time in Work Zone Traffic Accidents" CICTP: Safe, Smart, and Sustainable Multimodal Transportation Systems © ASCE, (2014) 3710-3722.

[19] Sharaf Alkheder et al., Severity Prediction of Traffic Accident Using an Artificial Neural Network, Journal of Forecasting, J. Forecast Published online in Wiley Online Library, (2016). DOI: 10.1002/for.2425.

[20] Sujata Basu and Pritam Saha., Regression Models of Highway Traffic Crashes: A Review of Recent Research and Future Research Needs , Procedia Engineering 187 (2017) 59–66, ScienceDirect, 2017

[21] Yangzexi Liu et al., Characteristic Analysis of Mixed Traffic Flow of Regular and Autonomous Vehicles Using Cellular Automata , Hindawi Journal of Advanced Transportation., (2017).

[22] Zhenke Luo, Yue Liu, Chen Guo,., Operational characteristics of mixed traffic flow under bi-directional environment using cellular automaton, Journal of Traffic and Transportation Engineering (English Edition), 1(6) (2014) 383-392, ISSN 2095-7564, https://doi.org/10.1016/S2095-7564(15)30288-9.

[23] Zhongqi Yuan et al., Characteristic Analysis and Prediction of Traffic Accidents in the Multiethnic Plateau Mountain Area, J. Transp. Eng., Part A: Systems, 146(8) (2020) ASCE.