Machine Learning based modelling of Intelligent Traffic System

Abhay Narayan Tripathi[1*], Bharti Sharma[1]
[1]Department of Computer Science,
DIT University, Dehradun, UK, India
e-mail: abhayntripathi74@gmail.com[1*], drbhhartisharaa@gmail.com[1]

ABSTRACT
In the present article, the authors have employed an Artificial Neural Network (ANN) based information processing technique for developing a prediction model of real time traffic volume in an urban area. For developing the ANN-based prediction model, the amount of various vehicles has been collected for 31 working days. This collected data has been used for training and testing the ANN model. Thereafter, to validate the developed ANN model, more 7 days data has been collected and compared with the predicted values of model for the same input attributes. From the results, it has been found that the performance of the developed model is highly adequate for the prediction purpose with the MSE and MAE values for training data and testing data as 0.0840 and 0.0202 and 0.2561 and 0.1511 respectively. Based on the acquired results it is evident that the proposed methodology may be significant in predicting the traffic volume of the anticipated road.

Keywords: Big data, Traffic control, Intelligent Transport Systems
DOI Number: 10.48047/nq.2022.20.19.NQ99110 **NeuroQuantology2022;20(19): 1182-1192**

1. INTRODUCTION

Big data is currently adding value for the users and various organizations with the utilization of various types of use cases. One such example is smart traffic control system. Smart network traffic represents the implementation of the smart technology in traffic systems. In these systems, interconnected data is implemented which helps in communicating different types of information for various services in the city [1]. Smart traffic or transportation in the cities is supported by an Intelligent Transport System (ITS). ITS is currently a boon because it utilizes cloud computing for processing, analyzing, and visualizing a large amount of data. By using these types of electronic technologies intelligent transport systems can easily control and process necessary information. These technologies will enable scientists to build smarter roads, vehicles, and users. These technologies will work by increasing the safety of transportation. It will also enable us to provide the most efficient routes which result in improving the operations of individuals. Thus, the implementation of big data has led to the development of the safety of vehicles with smart planning [2-4].

The intelligent transport system has led to the development of cost-effective data analysis. Deakin et al. [5] proposed a framework to relate big data analytics with a traffic control system. They used several data storage systems and analytical engines for plugging purpose. They also used different types basic set of interfaces and concepts for analysis of the proposed framework.

Amini et al. [6] proposed a flexible and comprehensive architecture for real-time traffic control depending using Big Data analytics. They also identified the heterogeneity and magnitude of the Big Data in the existing approaches in ITS. Jairo et al. [7] analyzed the transport data for Colombia city. Shi and Aty [8] demonstrated the significance of association monitor and safety and progress traffic operation. Their aim was to achieve improvement in the system performance of urban expressways by decreasing crash risk and congestion.

Zhu [9] found the application effect of the realized road traffic situational awareness system. He considered a situational awareness system on the ground of situational awareness technology, combining with convolution neural network database and some other technologies. He also analyzed the road traffic situational awareness system information collection, analysis process, and processing.

In literature survey, it has been found that research work related to big data in the current transport system are very limited. Hou et al. [10] focused on decision-making process for services related to advance travel information systems such as predicting the time of travel, predicting incidents related to traffic management and routes of vehicles, traffic volume etc. Situations that are related to the designing framework are analyzed by using data from videos [11]. The behaviour of drivers is estimated by using the speed of data. Technologies related to modern vehicles such as advanced sensing, increasedsafety analysis of vehicles for self-driving vehicles**.** For this, data mining has been used for detecting various signs of traffic or predicting the crash of traffic. Application of machine learning and huddle methods are majorly used [12]. Therefore, the need for an Intelligent Transport System framework has been implicated by using various types of connected vehicles and travel technologies. In these cases, the volume and speed of data were large.

Keeping all these facts in mind, in the present article, one month data of a busy road has been taken by the authors. Then, based on the acquired traffic data a feed-forward back-propagation ANN has been developed, trained, and tested for predictive modeling of traffic volume. Thereafter, the prediction accuracy of the developed ANN model has been tested for one week of traffic data.

2. Traffic data analysis before ANN modeling

In the initial stage, traffic data has been analyzed based on the interaction of the factors. The collected data of 31 days has been used for developing the ANN model. The detailed analysis of input data of all factors are mentioned in Table 1 (a-b). These tablesconsist the types of vehicles, their means and standard deviations. From the table 1 (b), it has been observed that the mean and standard deviation of traffic volume is 194.48 and 27.71 respectively.

Table 1. Detailed analysis of input data

	ST/ MT/ Mpeds	BCYCLE	PRICSHAW	TT	BCART/HCART	SCRJPVN	Sbus	Struck	SMBLVAN Tractors
Count	400	400	400	400	400	400	400	400	400
Mean	112.89	14.60	13.67	3.30	0.44	18.98	15.75	17.25	19.97
STD	8.56	4.27	4.55	1.79	0.74	6.07	5.66	7.48	5.329
Min	101	1	3	1	0	14.12	9.56	11.10	13.09
25%	107	12	10	2	0	15.99	12.52	12.53	15.72
50%	110	15	14	3	0	16.78	13.45	13.33	19.21
75%	116	18	18	4	1	17.67	15.93	18.50	21.91
Max	155	25	21	13	2	43.37	31.40	39.33	29.95

	S3_W	SSRMRMpeds	SBCYCLE	SPRICSHAW	STT	SBCARTHCART	Traffic Volume
Count	400	400	400	400	400	400	400
Mean	19.36	19.83	12.54	11.98	14.17	0.25	194.48
STD	5.92	6.24	3.58	5.17	2.80	0.43	27.71
Min	9.96	9.54	9.11	8.57	9	0	134
25%	14.09	16.15	10.23	9.67	12.54	0	177
50%	18.89	17.41	11.23	9.98	13.11	0	189
75%	24	28.10	12.94	15.12	14.40	0.78	207
Max	31.23	34.13	22.29	92.5	20.68	1.42	281

During analysis it has been found that there is no missing in the collected data.It is a well-known fact that the distribution of real-life data is generally skewed. Therefore, in the present study, subplots are plotted to check the distribution and skewness of factors, and shown in Fig. 1. Skewness represents the distribution of factor values around its mean. The skewness of factor's data represents by its tail. If the length of one side tail is longer than another, the distribution of data is assumed to skew which does not show any type of symmetry and is known as asymmetrical distributions. In Fig. 1 it has been observed that ST/MT/Mpeds, BCART/HCART, SCRJPVN, S3_W, SPRICSHAW, STT, TT, SBus, Struck, SSTMTMpeds, SBCYCLE, SBCARTHCART & Traffic volume have left or negatively skewed distribution which shows that the mean of these two factor's data is to the left of the peak value. On the other side, the distribution of PRICSHAW is right or positive-skewed as shown in Fig. 2. In other words, the mean of these factors is to the right side of the peak value.

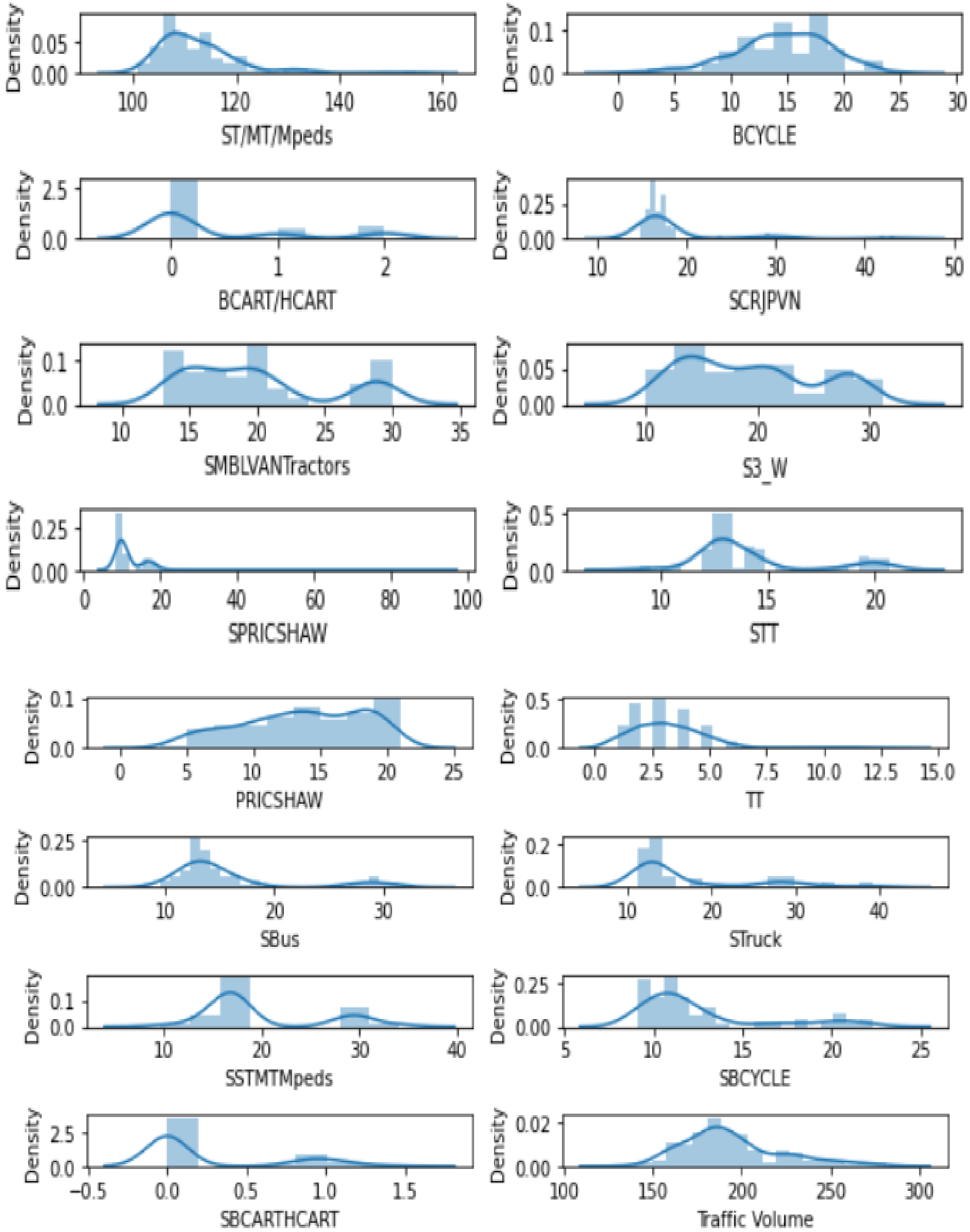

Fig. 2 Interaction matrix of pair plots of traffic volume data

In the present study, a large amount of traffic volume data is used. Therefore for accurate analysis of data, after normalization of the traffic data, a heat map is plotted to represent the numerical data in a color matrix according to the clustering shown in Fig. 3. It represents the behavior or trend of different factors on traffic volume. In Fig. 3, black color represents the negative correlation and white color represents the higher positive correlation. Moreover, Fig. 4 shows correlation plots of traffic volume factors.

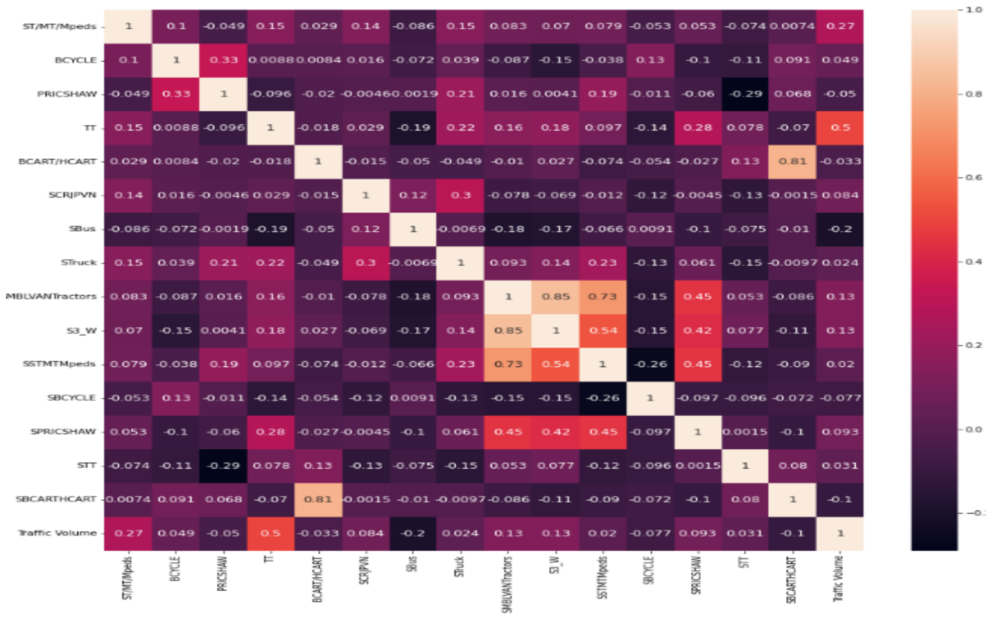

Fig. 3 Heat map of traffic volume data

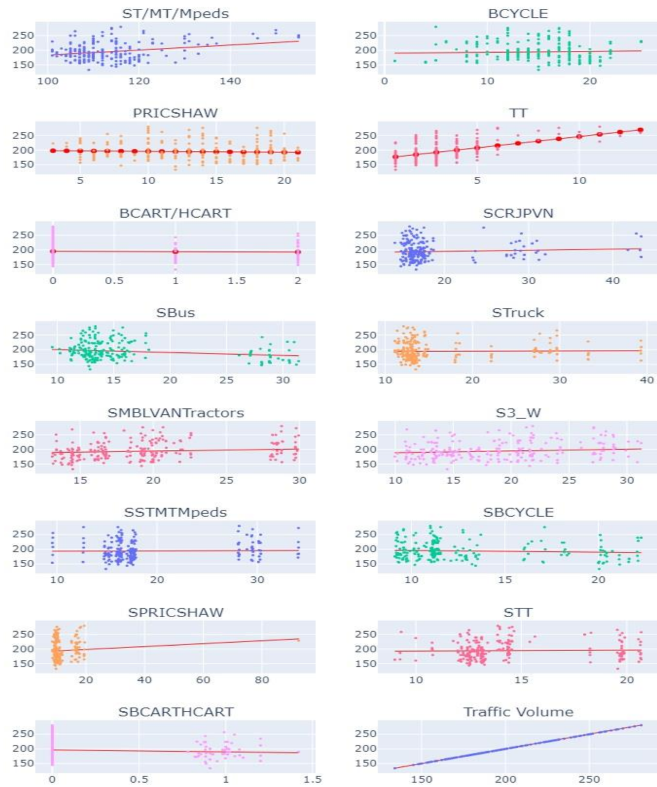

Fig. 4 Correlation matrix of correlated data with Traffic volume data

3. RESULTS & DISCUSSION

After the analyzing of collected data of traffic, an ANN model has been developed for traffic volume prediction. ANN is a type of computing system which simulates the pattern of knowledge transfer of the human brain through artificial neurons. A basic model of an artificial neuron is shown in Fig. 5. In this model, inputs, outputs, non-linearity function, and produced signal are represented by X_i, O_i, f_i, and R respectively. These networks are not only used in the area of engineering but also in social science, management, and humanities. They have a wide range of applications in the different sectors of engineering such as robotics, machine learning, manufacturing, process optimization, control systems, power systems, and pattern recognition, etc. [12] In a neural network, interconnected artificial neurons are built like a human brain. They are also known as the processing units which consist of input and output units. Like a human brain, ANN also follows some learning rules called algorithms to give an outcome. These algorithms are used to continuously modify the connected weights of artificial neurons. [13]

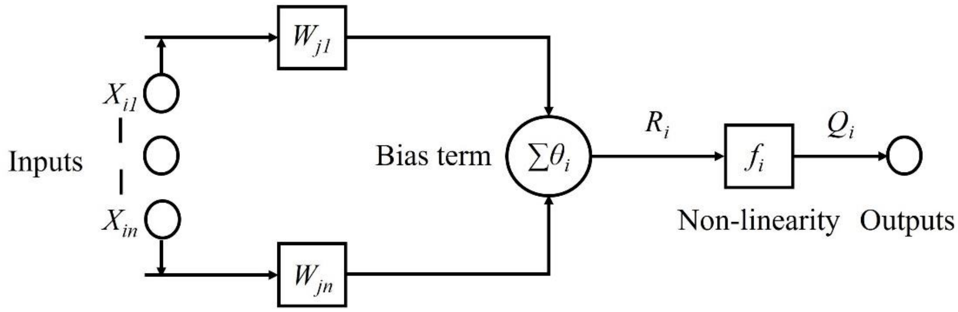

Fig. 5 Model of artificial neuron

In the present research work, a feed forward back propagation ANN has been developed to predict production time (PT) on the shop floor of an automobile manufacturing company. For developing the ANN-based prediction model, the production time involved in different processes has been collected for 31 working days. This collected data has been used for training and testing the developed ANN model. After the analysis of shop floor data, normalization of testing and training data has been performed due to a comparison between the multiple data series of various dimensions. The schematic of the developed ANN model is shown in Fig. 6.

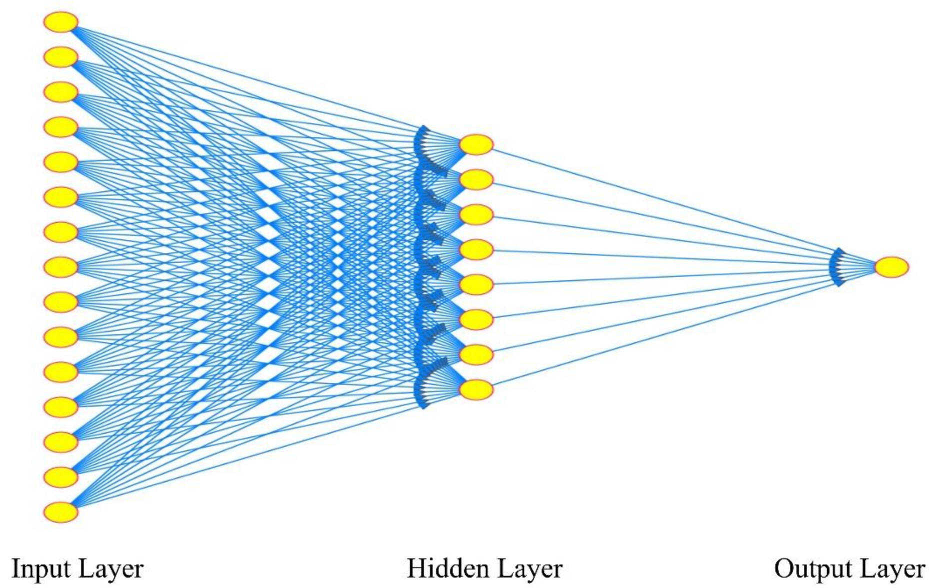

Fig. 6 Schematic of artificial neural network model

The type of layers in the developed network was dense with two hidden layers. The accuracy of the developed prediction model has been determined on the basis of mean absolute error (MAE) and mean square error shown in Fig. 7 and 8. MAE represents the mean of the absolute values of each prediction error i.e. the difference between the actual value and model-predicted value on all instances of the train data-sets. Notwithstanding, it quantifies the absolute average distance between the actual data and the anticipated data, but it fails to manage with large errors in prediction. On the other side, MSE measures the squared average distance between the actual data and the anticipated data. It effectively manages with bigger errors better than MAE. The values of MSE and MAE on train data of the developed model have been found as 0.0840 and 0.0202 and 0.2561 and 0.1511 respectively. These values show that the developed model has a higher degree of fitness and adequacy.

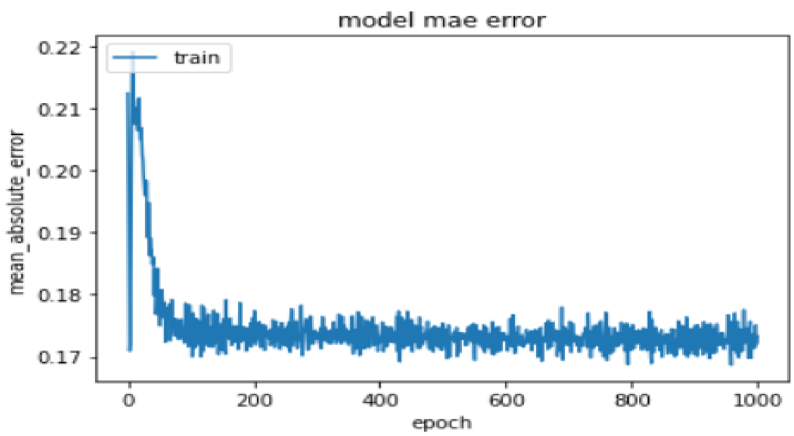

Fig. 7 Mean absolute error for training data sets

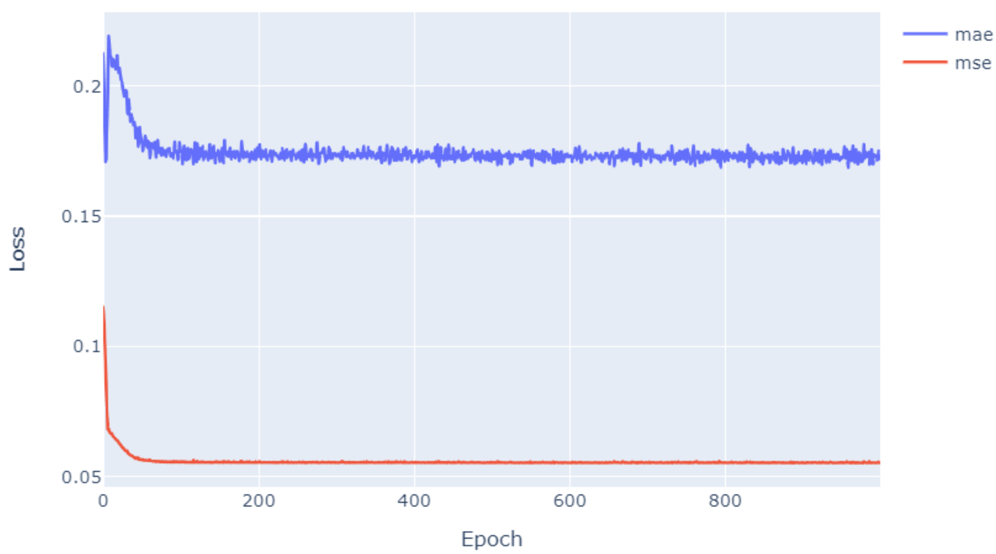

Fig. 8 Variation of loss for MAE and MSE of the developed model

After developing the machine learning-based prediction model for 31 days of data of traffic volume, authors have been used 7 days more data to validate the accuracy of the developed model. During validation, the predicted model provides MSE and MAE values for training data and testing data as 0.0840 and 0.0202 and 0.2561 and 0.1511 respectively. The values of these model performance measures show that the developed model is highly adequate for prediction purposes. The variation between actual and predicted values is shown in Fig. 9. The values of 10 optimal sets with actual and predicted values are listed in Table 2.

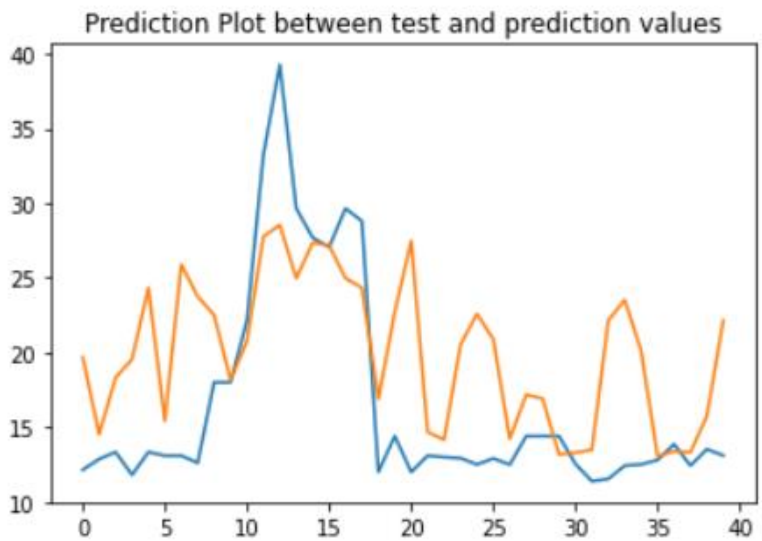

Fig. 9 Prediction plot between test and prediction values of traffic volume

Table 2. Optimal sets of traffic volume parameters

No.	Actual value	Predicted value
1	12.14	19.68
2	12.86	14.52
3	13.33	18.31
4	11.80	19.54
5	13.33	24.33
6	13.09	15.41
7	13.09	25.88
8	12.60	23.75
9	18.00	22.49
10	18.00	18.16

5. Conclusions

In the present research article, the authors have performed prediction modeling for traffic volume data predication by using machine learning. The following conclusions are drawn from the present study:

1. The traffic volume and vehicle types of acity road has been considered to perform the present case study. Bus,

Truck, Ricshaw, carts etc. have been considered as variable parameters to determine traffic volume.

2. Machine learning-based ANN model has been successfully developed by using 31 days of data of the calculated traffic volume. The values of MSE and MAE on the train data of the developed model have been found as 0.0840 and 0.0202 respectively.
3. The accuracy of the developed machine learning-based prediction model has been checked for the next 7 days of data of the traffic. It has been observed that the model performance is highly adequate for the prediction purpose with the MSE and MAE values for training data and testing data as 0.0840 and 0.0202 and 0.2561 and 0.1511 respectively.
4. The authors of the present study strongly believe that the results of the present study could be beneficial for traffic volume prediction and make city planning according to it.

REFERENCES

1. Amin, M. A., Hadouej, S., & Darwish, T. S. (2019). Big data role in improving intelligent transportation systems safety: A survey. In International Conference on Emerging Internetworking, Data & Web Technologies (pp. 187-199). Springer, Cham.
2. Darwish, T. S., Bakar, K. A., Kaiwartya, O., & Lloret, J. (2020). Trading: Traffic-aware data off loading for the big data-enabled intelligent transportation system. IEEE Transactions on Vehicular Technology, 69(7), 6869-6879.
3. Lu, K., Liu, J., Zhou, X., & Han, B. (2021). A Review of Big Data Applications in Urban Transit Systems. IEEE Transactions on Intelligent Transportation Systems, 22(5), 2535-2552.
4. Qureshi, K. N., & Abdullah, A. H. (2013). A survey on intelligent transportation systems. Middle-East Journal of Scientific Research, 15(5), 629-642.
5. Deakin, E. (2004). Mainstreaming Intelligent Transportation Systems: Survey of California Leaders. Transportation Research Record: Journal of the Transportation Research Board, 1885(1), 65-70.
6. Amini, S., Gerostathopoulos, I., & Prehofer, C. (2017). Big data analytics architecture for real-time traffic control. In 2017 5th IEEE international conference on models and technologies for intelligent transportation systems (MT-ITS) (pp. 710-715).
7. Shi, Q., & Abdel-Aty, M. (2015). Big data applications in real-time traffic

operation and safety monitoring and improvement on urban expressways. Transportation Research Part C: Emerging Technologies, 58, 380-394.

8. Zhu, Q. (2019). Research on road traffic situation awareness system based on image big data. IEEE Intelligent Systems, 35(1), 18-26.

9. Hou, Z., Zhou, Y., & Du, R. (2016). Special issue on intelligent transportation systems, big data, and intelligent technology. Transportation Planning and Technology, 39(8), 747-750.

10. Awaysheh, F., Cabaleiro, J. C., Pena, T. F., & Alazab, M. (2019). Big data security frameworks meet the intelligent transportation systems trust challenges. In 2019 18th IEEE International Conference on Trust, Security and Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE) (pp. 807-813).

11. Haykin, S. "Neural Networks: A Comprehensive Foundation, MacMillan College Publishing Co," New York, 1994.

12. Yegnanarayana, B. *Artificial neural networks*: PHI Learning Pvt. Ltd., 2009.