# Performance Evaluation Of Machine Learning Algorithms In Traffic Flow Prediction

Nazirkar Reshma Ramchandra [1], Dr. C. Rajabhushanam [2]
{reshma174@gmail.com[1], rajabhushanamc.cse@bharathuniv.ac.in[2]}

Research Scholar, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.[1],
Professor, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, Tamilnadu, India.[2]

**Abstract.**The main and foremost reason for traffic congestion is overpopulation and the poor condition of the roads. This mostly happens in the urban cities where all the people in the urban areas go for some work or certain purposes. Due to the current growth of the communication technology various computing techniques are used to predict the outcome based on a given dataset. This research work uses four kinds of machine learning techniques line Deep AutoEncoder (DAN), Deep Belief Network (DBN), Random Forest (RF), and Long Short Term Memory (LSTM) to predict the traffic flow. This proposed system is implemented using Python programming. Lastly, the outcome describes that the proposed model using the LSTM technique produces 94.3% accuracy and less error value.

**Keywords:** Traffic Flow, Machine Learning, Prediction, Accuracy, Recall, Performance.

## 1 Introduction

Traffic congestion on the road is caused because of many and more reasons. The traffic congestion is defined by the overload of vehicles in the road, the insufficient way to move the vehicles further in the road. In road the rate of traffic increases. Some of the most important effects of traffic congestion are because of environment, mechanical and human. The environment affects traffic because some of the bad weather conditions may cause traffic, mainly sometimes due to heavy rain. The mechanical means the vehicle is sometimes defected by mechanical issues the next is mainly by humans. The main cause of traffic because of drunk and drive, humans drive their vehicle after consuming alcohol the vehicle gets out of their control. In this proposed work these machine learning techniques are used to learn generic traffic flow prediction.

The second part of this research article elaborates on the overview of existing techniques used in the application traffic prediction. The third part describes the proposed methods are used in this research work. The fourth section deals with the result part of the current system. Finally, the fifth section concludes the proposed work.

I3CAC 2021, June 07-08, Chennai, India

DOI 10.4108/eai.7-6-2021.2308866

## 2 Literature Review

Traffic prediction is a very critical task in current traffic handling systems. Forecasting accuracy value in traffic system is decreased when uncovered events like accidents, weather changes, and condition of roads. Mining the data from social media like twitter increased the traffic flow prediction because traffic-related data also posted on social media regularly. AniekanEssien et al., 2020 propose a new traffic forecasting system using the concept of deep learning that incorporates traffic data retrieved from twitter and climate data. This model uses LSTM and stacked autoencoder framework for predict traffic flow using tweet data, climate datasets, and traffic. The proposed model has been with the UK city road network. The result of the system describes that it increases the forecasting accuracy level than other classical and machine learning approaches. this improved result direct decreases the irritation of the people, cost benefits for companies, and low impairment to the surroundings[1].

The prediction of traffic flow is a very critical task. Traffic prediction is used to increase the effectiveness of the transport system. Compare to the prediction of traffic flow, traffic jam forecasting researches are very less due to the lack of the best quality data set and recent techniques. Sen Zhang et al., 2019 proposes a common workflow to collect traffic jam data and generate traffic jam datasets depends on the analysis of images. The authors make a new traffic dataset by using traffic jam images from concern state transportation. They also propose a new model DCPN using a deep autoencoder with the concept of decoder and encoder to find about the relation of network transport and forecast traffic jam. The outcome of the proposed model describes the association of congestion intensity for traffic jam forecasting. This proposed work performed better than existing methods in terms of forecasting and computation effectiveness. This research work mainly focuses on forecasting of traffic jam levels with the help of images from the data set [2].

Object identification is one of the important parts of the growth of smart transport for a decreasing number of accidents. A. Dairi et al.., 2015 developed a new technique for recognizing objects in surroundings. This new model is integrated into the greedy method with a reduction of data dimension capability and KNN concept to reliable and accurate identification of the objects. The model was tested by using three data sets available publically. The outcome of the proposed model is compared with the deep belief network technique with the clustering technique. The result represents the proposed model is best to examine urban spots [3].

In recent days smart transport is one of the fast developmental areas by using deep learning methods. Due to the advancement of technology security and safety is increased and maintenance cost is decreased. In this research paper Arya KetabchiHaghighat et al., 2020 provides an outline and broad advantages of deep learning techniques on smart transport and presented the growth of smart transport. Initially, the authors discussed the various concepts of deep learning methods and the importance of the techniques. Lastly elaborates various systems using deep learning techniques are look over and the merits and demerits [4].

The number of vehicles is increased day by day. Due to the number of vehicles traffic flow is also increased in urban areas. Y. Liu et al., 2017 developed a prediction model by using random forest technique. The important features of the RF are more robustness, more performance and it is suitable to implement real-time problems. Climate type, time, type of road, quality of the road, and days are the important identifiers used to predict the traffic flow. At last, the outcome of the proposed model with the RF technique produces 87.5% accuracy and less error value. The computation process RF approach is simple and it is implemented

easily than other techniques. Due to that reason in data mining, RF is an efficient and fast technique [6].

L. Nie et al., 2017, proposes a new model for traffic forecasting using the DBN concept. This model initially takes on wavelet transform to retrieve the low pass element of traffic. A prediction system is developed by using DBN from the retrieved elements. High pass elements that represent an unequal variation of traffic network, a Gaussian system is used to simulate it. The authors measure the arguments of the Gaussian model with the help of the highest likelihood technique. From the output of the proposed model, the authors say that the newly proposed technique is outperformed than other techniques [7].

## 3 Proposed Methodology

Current computing techniques are playing an important role in real-time applications. Recent techniques are also used to predict the flow of traffic in the real world. This research article mainly concentrates on various machine learning concepts like DAN, DBN, RF, and LSTM. Here the data collected from the online website.

**DAN**

Deep autoencoders are the deep neural networks that are helpful to reproduce the input dataset at the output layers which is the number of neurons in the input layer. These algorithms consist which are encoder, code, and decoder. This is the type of artificial neural network that is used to analyze the data coding in an unsupervised manner. It has two internal layers that described a code that is used to represents the input and it is engaged with two main concepts that are the encoding which map the input into the code and a decoder that maps the code for the rebuilt of the original input dataset. Commonly autoencoder contains an unseen layer h that represents the code to indicates input data. It consists of two major elements: encoder method $h - f(x)$ and decode function$-g(h)$. The framework is represented in figure 1.

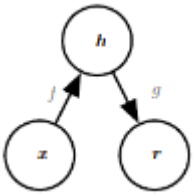



Figure 1 General framework of an autoencoder

The x(input data) mapped to r (output) using the ode value h. It has two elements: encoder f(mapping x to h) and decoder g (mapping h to r)New autoencoders have an idea of encoder and decoder beyond ahead of a deterministic method to stochastic associations $p_{encoder}(h|x)$ and $p_{decoder}(x|h)$. In the olden days, autoencoders are used to reduce the dimension of the data and learning features. Currently, associations between autoencoders and latent identifiers systems are used to construct generative systems. The one drawback that is found in the autoencoder is that when the size that that of the input layer, this concept can potentially learn the function identity. The next algorithm that has been applied to the real-time dataset is DBN, which s commonly called Deep belief Network.

**DBN**

DBN is a generative model based on the probabilistic system. It is constructed by using several layers of RBMs (Restricted Boltzmann Machines), each layer contains an unseen part

and visible part. DBN in the machine learning concepts DBN is considered as the generative graphical model or otherwise a deep neural network that is filled with different layers of latent variables. This algorithm aims to help the classification of the system in various types. It is trained to retrieve the input value with the help of the greedy method. After a trained RBM layer, the preceding unseen layer input is given to the next unseen layer. The pictorial representation of DBN is described in Figure 2[11].

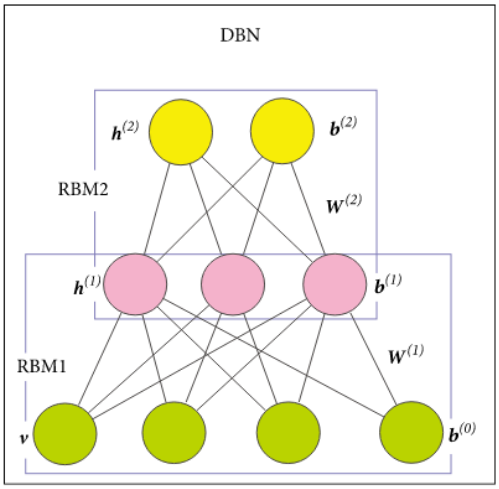



Figure 2 DBN[11]

Figure 2 indicates the two-layer DBN loaded by 2 RBMs. It consists of visible parts and 2 layers of unseen parts. From the figure 2 $h^{(1)}$ and $h^{(2)}$ denotes state vectors of unseen layers $v$ denotees the state vector of the visible part, $W^{(1)}$ and $W^{(2)}$ are symmetrical weights matrices, bias vectors of unseen layers are $b^{(1)}$ and $b^{(2)}$, and bias vector of the noticeable layer is $b^{(0)}$.

RF

RF approach is one of the ensemble approaches of DTs constructed on an arbitrarily divide dataset. The set of DT classifiers are called as forest. Every DT can be constructed using the identifiers like gain ratio value, Gini Index value, and information gain of every identifier. In classification, every DTs are voted and a famous class only was chosen for final output and regression all trees outputs average values are considered for final output. RF is one of the familiar and powerful classification concepts than other classification techniques. The working procedure of the RF technique is described below.

- From the concern dataset choose the samples randomly
- Make a DT or every sample and find the perdition output from every DT
- Do the vote for predicted output
- Choose the predicted output with the majority votes as the final output of the prediction

When creating RF depends on classification information Gini Index value is used.

$\text{Gini} = 1 - \sum_{i=1}^{C}(p_i)^2$ ---------- (1)

This function uses probability and class value to find the Gini of every node. From equation (1) $p_i$indicates the comparative frequency value of class in the given dataset and c denotes the total amount of classes. The entropy value can decide how data nodes in every branch in DT.

$\text{Entropy} = \sum_{i=1}^{C} -p_i * \log_2(p_i)$ --------- (2)

**LSTM**

The researcher from German introduces LSTM in 1997. It is the kind of RNN, able to learn long term associations. Multiple numbers of gate parts learn to close and open to the regular error value low. O(1) is the computational complexity value of the LSTM technique. The following diagram represents the LSTM model with a single memory part [1].

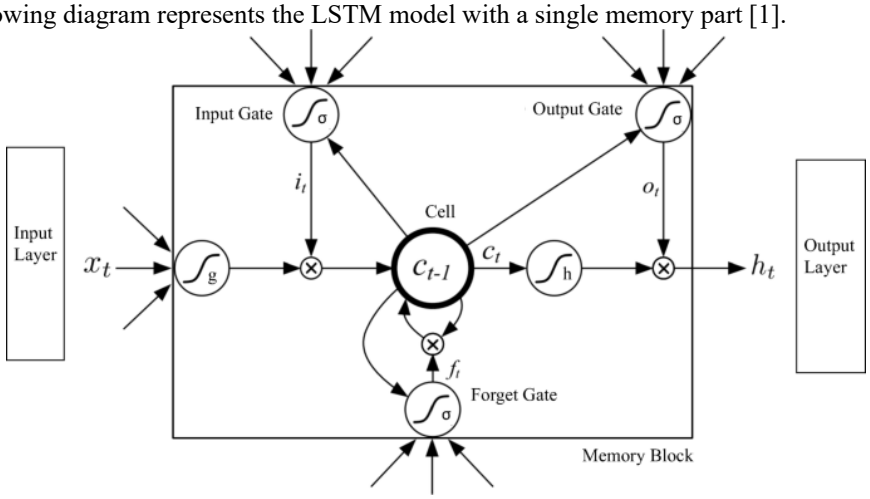



Figure 2 LSTM model with a single memory block [1]

The important parts of the LSTM are:

Forget gate - it deletes data not essential for the end of the task. This part is important to increase the performance of the entire network.

Input gate - this part is accountable for insert data into cells.

Output gate - this part is used to choose the results essential information.

In the model of LSTM, the input series is denoted as $x = x_1 + x_2 + x_3 \ldots . x_t$ and the output series $y = y_1 + y_2 + y_3 \ldots . y_t$ is calculated with the help of historical data. This is attained by using the following equations.

$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i$ ---------- (3)

$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f$ ---------- (4)

$c_t = f_t c_{t-1} + i_t g(W_{xc}x_t + W_{hc}h_{t-1} + b_c$ ---------- (5)

$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o)$----------(6)

$h_t = o_t h(c_t)$----------(7)

From the above equations, W indicates the weight matrixes; the bias vector indicates as b and $\sigma(.)$ represent a sigmoid method. The Sigmoid method is described as follows.

$\sigma(x) = \frac{1}{1+e^{-x}}$ ---------- (8)

$g(x) = \frac{4}{1+e^{-x}} - 2$---------(9)

$h(x) = \frac{2}{1+e^{-x}} - 1$----------(10)

From the above equation $g(.)$ and h (.) denotes the sigmoid methods. The identifiers i is used to indicating the input gate, f denotes the forget fate, o denotes the output gate value, and indicates cell activation vector value. LSTM model can be widely used to predict traffic flow and speed prediction.

**Performance Evaluation**

The main intention of the current research work is to predict the traffic flow by using a given data set using machine learning techniques. The performance of the given techniques can be evaluated by using various parameters like accuracy, precision, MSE, and RMSE.

Accuracy means the exact prediction percentage value for the given test data. The following equation (11) is used to compute the accuracy value.

$$\text{accuracy} = \frac{\text{correct predictions}}{\text{all predictions}} \text{ ------- (11)}$$

The precision metric can be evaluated by using the following equation (12)

$$\text{precision} = \frac{\text{true positives}}{\text{true positives+false positives}} \text{ ------- (12)}$$

The recall value is calculated by using the equation (13)

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \text{ ---------- (14)}$$

Mean squared error (MSE) is measured by using the following equation.

$$MSE(y_{true}, y_{pred}) = \frac{1}{n_{samples}} \sum (y_{true} - y_{pred})^2 \text{ ---------- (15)}$$

Equation (16) is used to calculate Root Mean Square (RMSE) value.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} |e_i^2|} \text{ ---------- (16)}$$

Here $e_i, i = 1,2, \ldots . . n$ indicates n number of sample modal error values, $x_i$ denotes the input value and $y_i$ represents the output value.

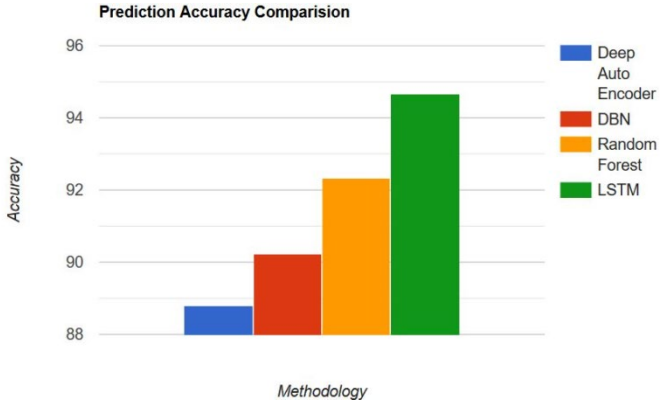
## 4 Result And Discussion

The performances of machine learning concepts are analyzed by using various metrics. Here the concepts are applied to the online traffic dataset. This proposed system is implemented by using python programming languages. The performance of the given system is evaluated with the help of accuracy, precision, recall, MSE, and RMSE metrics. The attributes like day, weather condition, events are mainly used to predict the traffic flow on a particular place. The following table shows the values of the metrics.

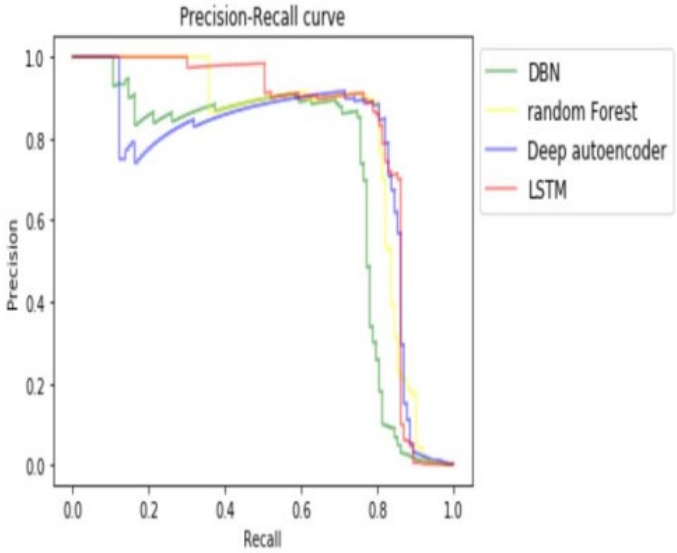
Table 1 Values of Metrics

| Metrics/Method | ADN | DBN | RF | LSTM |
|---|---|---|---|---|
| Accuracy | 88.6% | 90.43 | 92.6 | 94.3 |
| MSE | 3340.25 | 1709.71 | 346.35 | 250.93 |
| RMSE | 57.794895968 | 41.3486396 | 18.6104809 | 15.84077 |

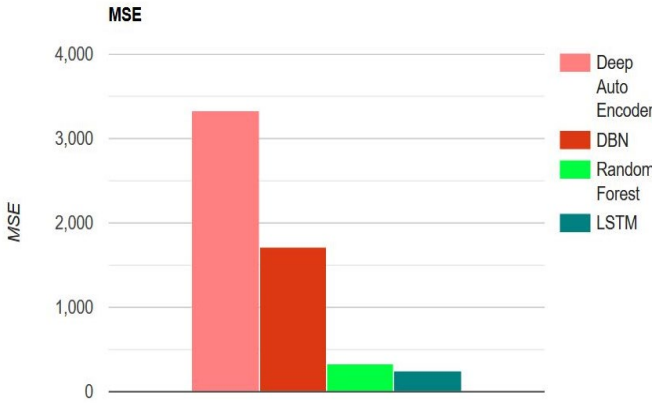
Figure 3 demonstrates the performance comparison of various metrics like accuracy, recall, RMSE, and MSE.



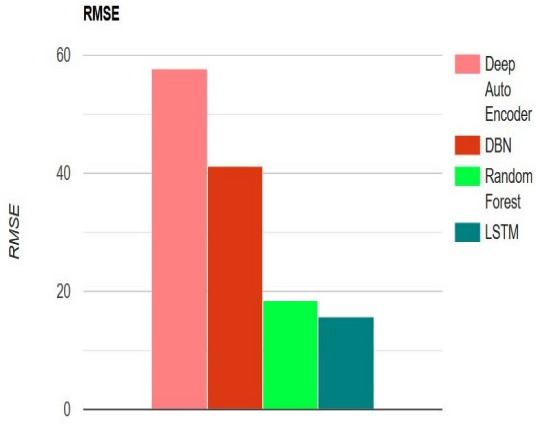

**2 a) Accuracy Comparison**




**2 b) Precision-Recall Curve**




**2 c) MSE Metric Comparison**




**2 d) RSME metric Comparison**

## Conclusion

Proper traffic prediction is one of the challenging areas in the current situation. Due to the overpopulation number of vehicles is also increased in urban places. Advanced techniques like machine learning techniques are used to predict the traffic flow properly. Weather condition, vehicle speed, day, events, the green and red cycle time of the signals are the important parameters that are used to predict the traffic flow. The data collected from online websites and implemented by using python programming. In this research work uses four types of machine learning algorithms to predict the traffic flow. From the outcome of the four techniques, LSTM produces better accuracy of 94.3%.

## References

[1] AniekanEssien&IliasPetrounia& Pedro Sampaio1 & Sandra Sampaio, "A deep-learning model for urban traffic flow prediction with traffic events mined from twitter", Springer World Wide Web.

[2] Sen Zhang, Yong Yao , Jie Hu , Yong Zhao , Shaobo Li &Jianjun Hu (2019), "Deep Autoencoder Neural Networks for Short-Term Traffic Congestion Prediction of Transportation Networks" , Sensors, Vol. 19, 2229, pp. 1-19.

[3] A. Dairi, F. Harrou, Y. Sun & M. Senouci (2018), "Obstacle Detection for Intelligent Transportation Systems Using Deep Stacked Autoencoder and k -Nearest Neighbor Scheme," IEEE Sensors Journal, vol. 18, no. 12, pp. 5122-5132.

[4] Arya KetabchiHaghighat, VarshaRavichandra-Mouli, Pranamesh Chakraborty, YasamanEsfandiari, Saeed Arabi& Anuj Sharma (2020), Applications of Deep Learning in Intelligent Transportation Systems, Journal of Big Data Analytics in Transportation.

[5] S. Uma Devi & S. Nirmala SugirthaRajini (2019), " Detection of Traffic Violation Crime Using Data Mining Algorithms", Jour of Adv Research in Dynamical & Control Systems, Vol. 11, No.9, pp. 982-987

[6] Y. Liu & H. Wu (2017), "Prediction of Road Traffic Congestion Based on Random Forest," 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, pp. 361-364.

[7] L. Nie, D. Jiang, S. Yu & H. Song(2017), "Network Traffic Prediction Based on Deep Belief Network in Wireless Mesh Backbone Networks," IEEE Wireless Communications and Networking Conference (WCNC), pp. 1-5.

[8] L. Han & and Y. Huang, "Short-term traffic flow prediction of road network based on deep learning," in IET Intelligent Transport Systems, vol. 14, no. 6, pp. 495-503.
[9] L. Nie, D. Jiang, L. Guo, S. Yu & H. Song(2016), "Traffic Matrix Prediction and Estimation Based on Deep Learning for Data Center Networks," IEEE Globecom Workshops (GC Wkshps), pp. 1-6.
[10] YuhanJia, Jianping Wu &Yiman Du(2016), "Traffic speed prediction using deep learning method," IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), pp. 1217-1222.
[11] Jiangeng Li, Xingyang Shao &Rihui Sun(2019), "A DBN-Based Deep Neural Network Model with Multitask Learning for Online Air Quality Prediction", ",Hindawi Journal of Control Science and Engineering, , pp.1-9.