Traffic Volume Forecast Using Statistical Regression Learning Forannual Average Daily Traffic (AADT) Estimate

Manju Chouhan[1], Prof. Vinay W. Deulkar[2]
[1, 2]Dept of Civil Engineering
[1, 2]JIT, Borawan, India

Abstract- *With the advent of Intelligent Transportation Systems (ITS), short term traffic prediction is a major feature to develop the smart traffic system for the smart city framework. It allows real time data accumulation, analysis and operations in routing traffic through the cities. From a systems perspective, the traffic volumes in large cities often exhibit spatial and temporal coherence which can be leveraged to predict the short term traffic volume. However, the time series prediction is challenging due to the analysis of extremely large and complex data sets with dependence on a multitude of parameters or features. This paper presents a machine learning based approach based on the principal component analysis (PCA) and the Back Prop algorithm to for sales forecasting. The performance of the proposed system is evaluated in terms of the mean absolute percentage error (MAPE) and the regression. It is shown that the proposed system outperforms the previously existing system working on the benchmark datasets [1].*

Keywords: Annual Average Daily Traffic (AADT) , Intelligent Traffic System (ITS), Short Term Traffic Prediction, Discrete Wavelet Transform, BackProp, Mean Absolute Percentage Error, Regression.

I. INTRODUCTION

Intelligent Traffic Systems (ITS) have seen an unprecedented interest due to the following reasons [1]:

1) Urbanization throughout the world.
2) Mass exodus of populations towards cities.
3) Population explosion.
4) Increased purchasing power capacity of people worldwide.

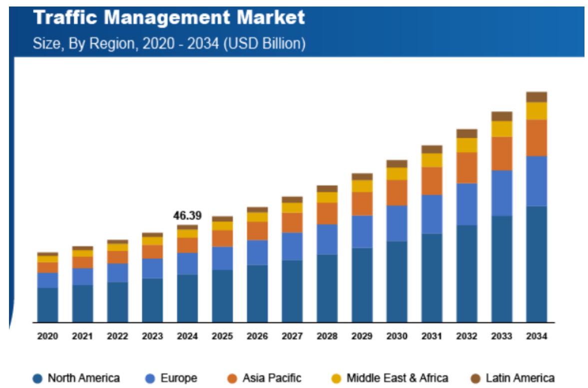

Fig.1 Traffic Management Market Size, Share, Trends, & Industry Analysis
(Source: Polaris Labs)
(https://www.polarismarketresearch.com/industry-analysis/traffic-management-market)

Figure 1 depicts the Intelligent Traffic Systems (ITS) market, for increasing global traffic. The increasing trends have some serious adverse effects [2]:

1) Increased chance of accidents.
2) Large latencies.
3) Increased exhaustion.
4) Lower productivity.
5) Increased pollution levels.
6) Psychological stress.

Intelligent Transportation Systems (ITS) is a combination of leading-edge information and communication technologies used in transportation and traffic management systems to improve the safety, efficiency, and sustainability of transportation networks, to reduce traffic congestion and to enhance drivers' experiences[3]-[4]. This necessitates the estimation of traffic volume for cities with large vehicular populations. Typically, the traffic volume can be modelled as a time series model and can be expressed as [5]:

$$Traffic\ volume = function(time, associated\ variables)$$

The associated variables can be:

1) Speed of vehicles
2) Volume (no. of vehicles)
3) Occupancy

II. MACHINE LEARNING FOR TRAFFIC FORECAST

Due to the enormity of the data to be analysed, machine learning based approaches are being explored off late. The travel time varying throughout the day critically depends on the traffic conditions [6]. A typical variation of the traffic volume is depicted in figure 2.

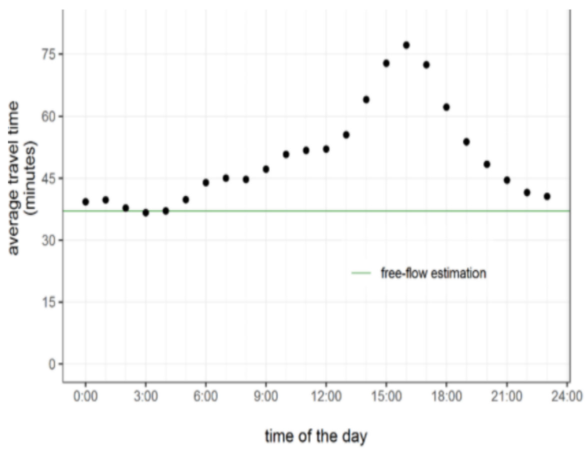

Fig.2 Variation in average travel time as a function of day time.

In recent years, machine learning, deep learning, and probabilistic programming have shown great promise in generating accurate forecasts [7]. Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Other advantages include:

1. **Adaptive learning:** An ability to learn how to do tasks based on the data given for training or initial experience.
2. **Self-Organization:** An ANN can create its own organization or representation of the information it receives during learning time.
3. **Real Time Operation:** ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.

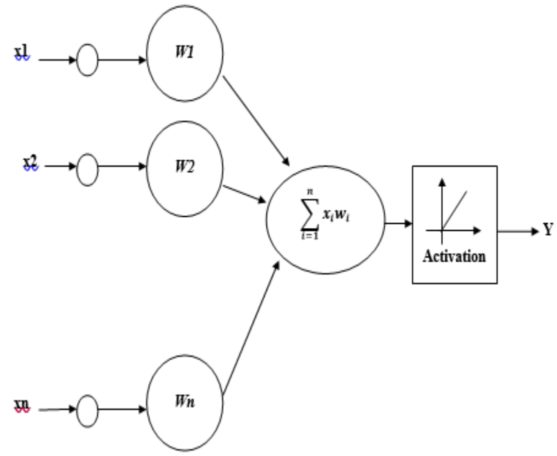

Fig.3 Mathematical Model of Neural Network

The output of the neural network is given by:

$$\sum_{i=1}^{n} XiWi + \Theta \quad (1)$$

Where,
Xi represents the signals arriving through various paths,
Wi represents the weight corresponding to the various paths and
ϴ is the bias.

It can be seen that various signals traverssing different paths have been assigned names X and each path has been assigned a weight W. The signal traverssing a particular path gets multiplied by a corresponding weight W and finally the overall summation of the signals multiplied by the corresponding path weights reaches the neuron which reacts to it according to the bias ϴ. Finally its the bias that decides the activation function that is responsiblefor the decision taken upon by the neuralnetwork. The activation function φ is used to decide upona the final output [8]. The learning capability of the ANN structure is based on the temporal learning capability governed by the relation:

$$\boldsymbol{w(i) = f(i,e)} \quad (2)$$

Here,
w (i) represents the instantaneous weights
i is the iteration
e is the prediction error

The weight changes dynamically and is given by:

$$\boldsymbol{W_k \xrightarrow{e,i} W_{k+1}} \quad (3)$$

Here,

W_k is the weight of the current iteration.

W_{k+1} is the weight of the subsequent iteration.

(i) Regression Learning Model

Regression learning has found several applications in supervized learning algorithms where the regression analysis among dependednt and independent variables is eeded [9]-[10]. Different regression models differ based on the the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used. Regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a relationship between x (input) and y(output). Mathematically,

$$\boldsymbol{y = \theta_1 + \theta_2 x} \qquad \textbf{(4)}$$

Here,
x represenst the state vector of inut variables
y rperesenst the state vector of output variable or variables.
Ө1 and Ө2 are the co-efficents which try to fit the regression learning models output vector to the input vector.

By achieving the best-fit regression line, the model aims to predict y value such that the error difference between predicted value and true value is minimum. So, it is very important to update the θ1 and θ2 values, to reach the best value that minimize the error between predicted y value (pred) and true y value (y). The cost function J is mathematically defined as:

$$\boldsymbol{J = \frac{1}{n}\sum_{i=1}^{n}(pred_1 - y_i)^2} \qquad \textbf{(5)}$$

Here,
n is the number of samples
y is the target
pred is the actual output.

(ii) Gradient Descent in Regression Learning

To update θ1 and θ2 values in order to reduce Cost function (minimizing MSE value) and achieving the best fit line the model uses Gradient Descent [11]. The idea is to start with random θ1 and θ2 values and then iteratively updating the values, reaching minimum cost. The main aim is to minimize the cost function J [12].

III. THE STEEPEST DESCENT ALGORITHM AND PRINCIPAL COMPONENT ANALYSIS

The critical aspect about steepest descent is the fact that it repeatedly feeds the errors in every iteration to the network till the errors become constant or the maximum number of allowable iterations are over. This can be mathematically given by:

$if\ PF \neq constant$
$for\ (k = 1, k \leq k_{max} = constant, k = k + 1)$
{
$W_{k+1} = f(X_k, W_k, e_k)$
}
else
{
$W_{k+1} = W_k\ \&\&\ training\ stops$
}

Here,
X_k is the input to the kth iteration
W_k is the weight to the kth iteration
W_{k+1} is the weight to the (k+1)st iteration
e_k is the error to the kth iteration
k is the iteration number
PF is the performance function deciding the end of training
k_{max} is the maximum number of iterations

Thus if the error is within tolerance, which is generally not feasible to find beforehand in time series data, the training is stopped if the performance function (which can be the training error) becomes constant for multiple iterations or the maximum number of iterations are over. Now there are various ways in which the error can be minimized. However, the steepest fall of the error with respect to weights is envisaged. It is depicted in the figure below:It can be seen from figure 1 that although the error in training keeps plummeting in all the three cases of gradient descent, the gradient 3 or g3 attains the maximum negative descent resulting in the quickest training among all the approaches and hence the least time complexity [13]. This would be inferred from the number of iterations which are required to stop training. Thus the number of iterations would be a function of the gradient with which the error falls.

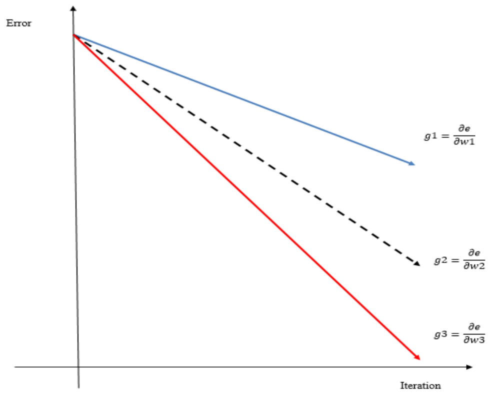

Fig.4 The concept of Steepest Descent

This is mathematically given by:

$$k_n = f(g = \frac{\partial e}{\partial w}) \qquad \textbf{(6)}$$

Here,
k_n is the number of iterations to stop training.
g is the gradient
w is the weight
e is the error
f stands for a function of

The proposed methodology uses two key components one of which is the training algorithm and the other is the training optimization algorithm. Both are explained in this section

a) The BackProp

There are several ways to implement the back propagation technique in the neural networks. One consideration however always remains that of the least time and space complexity so as to reduce the amount of computational cost that is associated with the training algorithm. The essence of the scaled conjugate gradient algorithm is the fact that it has very low space and time complexity making it ideally suited to large data sets to be analyzed in real time applications where the time is a constraint. The training rule for the algorithm is given by[14]:

$$A_0 = -g_0 \qquad \textbf{(7)}$$

A is the initial search vector for steepest gradient search [16]
g is the actual gradient

The BackProp algorithm is thus computed as [17]:

$$w_{k+1} = w_k - \alpha[\frac{\partial^2 e}{\partial w^2}]^{-1}\frac{\partial H}{\partial w} \qquad \textbf{(8)}$$

Here,
w_k & w_{k+1} denote the weights of the present and subsequent iterations.
α denotes the learning rate.
e denotes the error in the present iterations.

b) The Principal Component Analysis (PCA)

The principal component analysis (PCA) is basically a dimensional reduction tool which helps to clear out the redundancies in the training data vector in such a way that the training is optimized for lesser number of variables and mean absolute percentage error hits the least values in the lest number of iterations possible [17].

c) The discrete wavelet transform (DWT)

As the raw data exhibits high, thus the proposed approach employs data cleaning techniques such as the wavelet transform [18].The mathematical formulation for the wavelet transform is given by the scaling and shifting approach of the wavelet function [19].

The scaling, shifting dependence can be defined as:

$$W\varphi(Sc, Sh) = ₩[\mathbf{x}, \mathbf{t}] \qquad \textbf{(9)}$$

Here,
x is the space variable
t is the time variable
₩ is the transform
sc is the scaling factor
sh is the shifting factor

The wavelet transform is an effective tool for removal of local disturbances. Pharmaceutical demands show extremely random behavior and local disturbances. Hence conventional Fourier methods do not render good results for highly fluctuating data sets [20]. The major advantage of the wavelet transform is the fact that it is capable of handling fluctuating natured data and also local disturbances. The DWT can be defined as:

$$\mathbf{W\Phi\ (Jo, k)} = \frac{1}{\sqrt{M}}\sum_n S(n).\Phi(n)_{jo,k} \qquad \textbf{(10)}$$

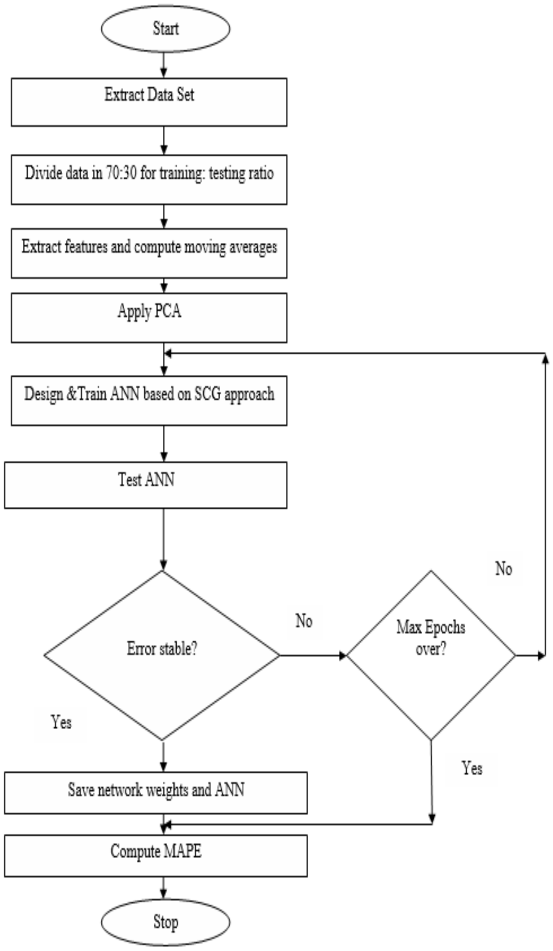

Figure 5. Flowchart of Proposed System

The training is stopped based on the mean square error or mse given by:

$$mse = \frac{\sum_{i=1}^{n} e_i^2}{n} \qquad \textbf{(11)}$$

The final computation of the performance metric is the mean absolute percentage error given by [21]:

$$MAPE = \frac{100}{M}\sum_{i=1}^{N}\frac{E-E_i|}{i} \qquad \textbf{(12)}$$

Here,
n is the number of errors
i is the iteration number
E is the actual value
E_i is the predicted value

IV. RESULTS AND DISCUSSIONS

The system has been designed on MATLAB 2020a. The training algorithm used is the steepest descent based scaled conjugate gradient algorithm. The accuracy of temporal prediction is computed as:

$$Ac = 100 - \frac{100}{M}\sum_{i=1}^{N}\frac{E-E_i|}{i}\ \% \qquad \textbf{(13)}$$

Here,
Ac is the accuracy computed in %

Basically a 10 hidden layer deep neural network is design based on the scaled conjugate gradient approach. The need for a deep neural network is seen as a necessity since the traffic data in terms of traffic volume is substantially complex and exhaustive.

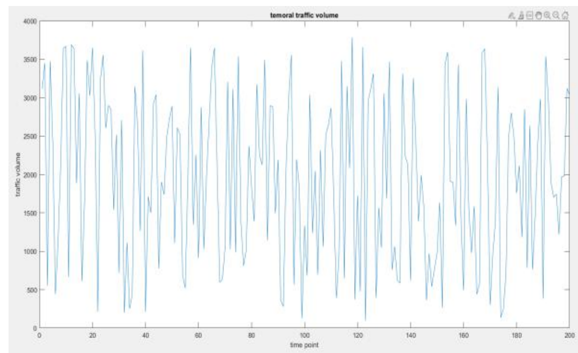

Fig.6. Raw Traffic Volume

Figure above depicts the raw data.

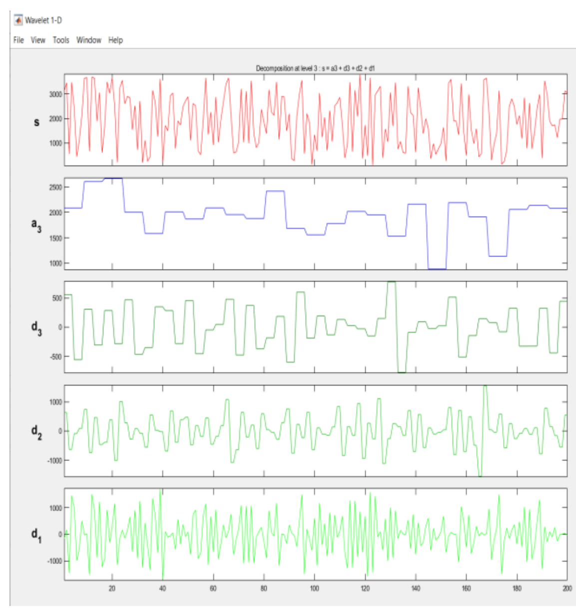

Fig.7 Wavelet Decomposition at level 3.

Figure above depicts the wavelet decomposition of raw data.

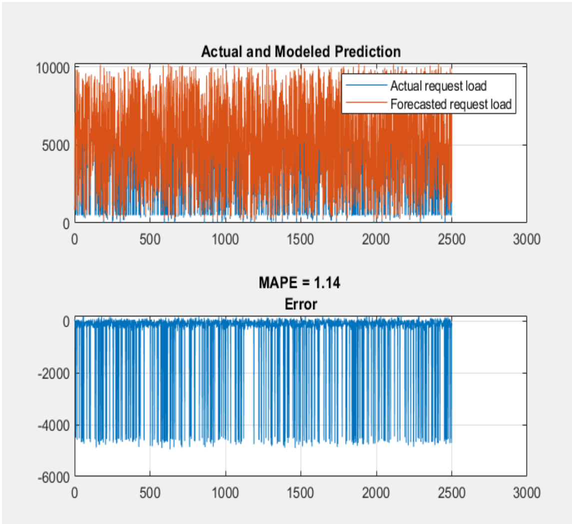

Fig.8 Actual and Modelled values

The figure above depicts the actual and the modelled prediction values in which the red curve is the values corresponding to forecasted values and the blue curve corresponds to the actual values. A difference between them yields the error.

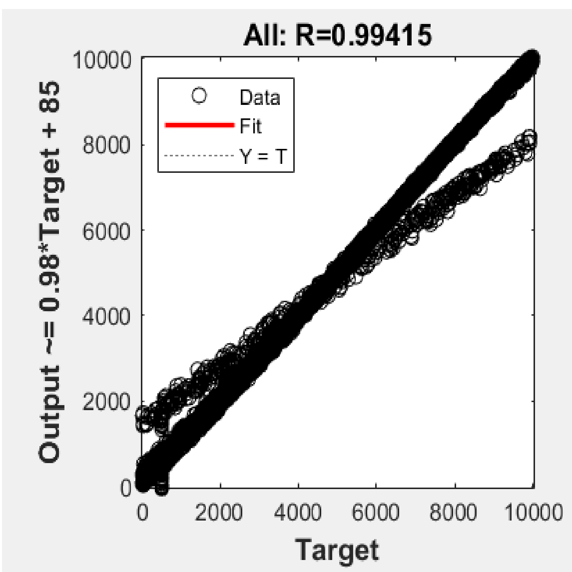

Fig.9 Regression

The figure above depicts the regression obtained in the proposed approach which is a sort of similarity among two random variables. The maximum allowable regression is unity depicting complete similarity.

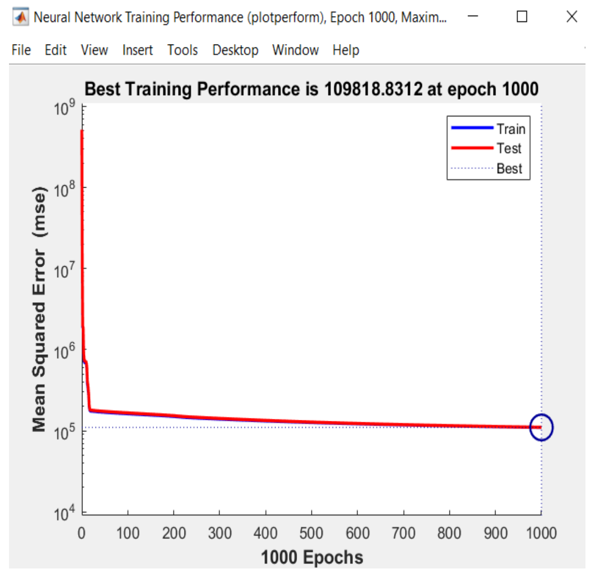

Fig.10 Performance Function

The performance function that decides the culmination of training is the mean squared error as a function of iterations.

Table1. Summary of Results

S.No	PARAMETER	VALUE
1.	Model	DNN
2.	Architecture	Back Prop
3.	Iterations	1000
4.	**MAPE (Proposed Work)**	**1.14%**
5.	**MAPE (Previous Work)**	**15.8% (SARIMA) 11.7% (ML) 12.8% (LSTM) 2.5% (Power-function-based LSTM Model)**

Table 1 presents the summary of results.

Theperformance of the proposed approach is found better compared to previously existing technique [11] which attains a best case MAPE of 2.5%. The limitations of the proposed work was the absence of data pre-processing and optimization tools such as the discrete wavelet transform and PCA.

V. CONCLUSION

This paper presents a mechanism for short term traffic forecast. The neural network architecture is used to implement machine learning. The architecture of the approach is the use of the back propagation based approaches to utilize the knowledge about the errors in each iteration to affect the weights of each subsequent iteration. The Back Prop based steepest descent training rule is utilized in this approach to reduce the number of iterations and also the mean absolute percentage error. The principal component analysis (PCA) is used as a data optimization tool while the (DWT) has been used a data smoothening filter. It is shown that the proposed work performs better in terms of mean absolute percentage error (MAPE) compared to the previously existing technique for the same standard dataset. It can be attributed to the back propagation approach as well as the use of the principal component analysis to fine tune and optimize the training process. The proposed approach attains an MAPE of 1.14%. It is substantially improved compared to previous work [11].

REFERENCES

[1] A Pompigna, F Rupi, “Comparing practice-ready forecast models for weekly and monthly fluctuations of average daily traffic and enhancing accuracy by weighting methods”, Journal of Traffic and Transportation Engineering , Elsevier 2018, vol. 5, Issue 4, pp. 239-253.

[2] I. M. Wagner-Muns, I. G. Guardiola, V. A. Samaranayke and W. I. Kayani, "A Functional Data Analysis Approach to Traffic Volume Forecasting," in IEEE Transactions on Intelligent Transportation Systems, 2018, vol. 19, no. 3, pp. 878-888.

[3] AS Soliño, ALL Galera,, “Measuring uncertainty of traffic volume on motorway concessions: a time-series analysis”, Transportation Research Procedia, Elsevier 2017, Volume 27, pp.3-10.

[4] D. Wang, Q. Zhang, S. Wu, X. Li and R. Wang, "Traffic flow forecast with urban transport network," 2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE), 2016, pp. 139-143.

[5] K. Xie and R. Li, "A combined forecasting method for traffic volume," 2016 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), 2016, pp. 95-100

[6] M Meng, J Zhang, YD Wong, PH Au, “Effect of weather conditions and weather forecast on cycling travel behavior in Singapore”, International Journal of Sustainable Transportation, Taylor and Francis, 2016, vol.10, issue.0, pp. 773-780.

[7] M Moniruzzaman, H Maoh, W Anderson, “Short-term prediction of border crossing time and traffic volume for commercial trucks: A case study for the Ambassador Bridge”. Transportation Research Part C: Emerging Technologies, Elsevier 2016, Vol. 63, pp. 182-194.

[8] J Tang, G Zhang, Y Wang, H Wang, F Liu, “A hybrid approach to integrate fuzzy C-means based imputation method with genetic algorithm for missing traffic volume data estimation”, Transportation Research Part C: Emerging Technologies, Elsevier 2015, Volume 51, pp.29-40.

[9] K. C. Dey, A. Mishra and M. Chowdhury, "Potential of Intelligent Transportation Systems in Mitigating Adverse Weather Impacts on Road Mobility: A Review," in IEEE Transactions on Intelligent Transportation Systems, 2015, vol. 16, no. 3, pp. 1107-111.

[10] H. Ibrahim and B. H. Far, "Data-oriented intelligent transportation systems," Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), 2014, pp. 322-329.

[11] D.C. Han, “Prediction of Traffic Volume Based on Deep Learning Model for AADT Correction”, Applied Sicences, MDPI, 2024, vol.14, no.20, pp.1-19.

[12] S Park, M Kim, J Kim, “Hourly Long-Term Traffic Volume Prediction with Meteorological Information Using Graph Convolutional Networks”, Applied Sicences, MDPI, 2024, vol.16, no.6, pp.1-12.

[13] Y. Chen, J. Guo, H. Xu, J. Huang and L. Su, "Improved Long Short-Term Memory-Based Periodic Traffic Volume Prediction Method," in IEEE Access, vol. 11, pp. 103502-103510, 2023.

[14] T Ma, C Antoniou, T Toledo, “Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast”, Journal of Transportation Research Part, Emerging Technologies, Elsevier 2022, Volume 111, pp. 352-372.

[15] Haghighat, A.K., Ravichandra-Mouli, V., Chakraborty, P. et al. Applications of Deep Learning in Intelligent Transportation Systems. J. Big Data Anal. Transp, Springer 2021, vol. 2, 115–145.

[16] A. Haydari and Y. Yılmaz, "Deep Reinforcement Learning for Intelligent Transportation Systems: A Survey," in IEEE Transactions on Intelligent Transportation Systems, 2020, vol. 23, no. 1, pp. 11-32.

[17] Muthuramalingam S., Bharathi A., Rakesh kumar S., Gayathri N., Sathiyaraj R., Balamurugan B. (2019) IoT Based Intelligent Transportation System (IoT-ITS) for Global Perspective: A Case Study, ntelligent Systems Reference Library, Springer 2020, vol 154, pp.279-300.

[18] CF Daganzo, A variational formulation of kinematic waves: basic theory and complex boundary conditions, Transportation Research Part B: Methodological, Elsevier 2005, vol.39, Issue 2, pp.187-196

[19] T Ma, Z Zhou, C Antoniou, “Dynamic factor model for network traffic state forecast”, Transportation Research Part B: Methodological, Elsevier, 2018, vol.118, pp.281-317

[20] Z. Diao et al., "A Hybrid Model for Short-Term Traffic Volume Prediction in Massive Transportation Systems," in IEEE Transactions on Intelligent Transportation Systems, 2018, vol. 20, no. 3, pp. 935-946
[21] S Satunin, E Babkin, “A multi-agent approach to intelligent transportation systems modeling with combinatorial auctions”, Expert Systems with Applications, Elsevier 2014, vol. 41, no. 15, pp.6622-6633.