World Academics Journal of Engineering Sciences
Research Paper
Vol.7, Issue.2, pp.72-76, June (2020)
E-ISSN: 2348-635X

A Naïve Bayes Classifier Approach to Incorporate Weather to Predict Congestion at Intersections

Saurav Barua

Department of Civil Engineering, Daffodil International University, Dhaka, Bangladesh

Author's Mail Id: saurav.ce@diu.edu.bd Tel.: +88-01715334075

Available online at: www.isroset.org

Received: 26/April/2020, Accepted: 19/May/2020, Online: 30/June/2020

Abstract— The study endeavored to model the influence of weather phenomena on traffic congestion considering various type and time of days at different intersections in Dhaka city. A Naïve Bayes Classifier method was adopted to model this causation relation from the field survey data using Scikit-learn software. The data obtained was divided into training and testing dataset. The proposed Naïve Bayes Classifier model showed 72.25% and 85.03% accuracy in training and testing respectively. The root mean square error (RMSE) and mean absolute error (MAE) of the model were 0.46 and 0.28 respectively. The Naïve Bayes Classifier model showed good prediction potential to assess the influence of weather condition on traffic congestion. The outcomes of the study can be utilized to develop Traffic Management System (ATMS) and Advanced Traveler Information System (ATIS) for Dhaka city. Hence, the motorists can decide which intersections to travel while route choice in advance and congestion will be reduced consequently.

Keywords— Traffic congestion, weather condition, Naïve Bayes Classifier, Scikit-learn, road intersections

I. INTRODUCTION

Traffic congestion incurs loss of time to passengers and motorists, delays, fatigue and stress to road users, unable to predict travel time accurately, wastage of fuels and other huge direct and indirect monetary loses. Traffic fluctuate with respect to day types, i.e. weekday, weekend, holiday and times of days i.e. peak and off peak hour based on travel demand and hence, congestion phenomena changes. Among various other factors, weather condition influences traffic congestion is a widely accepted anecdote; several studies have been conducted to provide a causation link between them [1, 2, 3, 4]. Maritime and aviation authorities use weather information to conduct day-to-day operation smoothly. On contrary, the effects of weather phenomena with respect to day types and time on road traffic condition is yet to be monitored and adopted adequately.

Inclement weather affects travel demand, traffic safety, drivers' behaviour, traffic speed and flow adversely. Heavy rainfall reduces traffic speed [5], snowfall affects free flow speed [6], scheduling of public bus services were hampered by adverse weather [7] and traffic volume varies with snowy and cold weather [8]. Federal Highway Administration (FHWA) under Responsive Weather Management Program (RWMP) developed statistical models to evaluate the effects of weather on traffic flow in various cities of USA [9]. Some of the trending researches include, prediction of traffic congestion occurred by snow and rainfall [10], correlation of the level of traffic congestion with the parameters of weather condition [11], traffic forecast modelling using random forest classifier [12] and logistic regression with fuzzy logic modelling to predict congestion for route choice decision [13]. However, no such study have been conducted relating weather condition and traffic congestion in context of Dhaka city, Bangladesh, where traffic congestion is becoming intolerable day by day.

It is essential to predict when and where traffic congestion will occur and its relation with weather phenomena under different types of days and times. Our study endeavoured to predict relationship between weather and traffic congestion using Naïve Bayes (NB) Classifier, which is a popular Bayesian Network, especially applicable to solve various classification problems [14, 15]. Although several other models have been designed to improve prediction of congestion under harsh weather conditions, Naïve Bayes Classifier have better accuracy in prediction than others [16]. NB classifier models were used to predict traffic flow [17], travel behaviour [18] and traffic incident detection [19] under different weather conditions.

This study predict traffic congestion for various weather conditions under various types and times of days at signalized intersections. The proposed study developed correlation between weather and traffic congestion using Naïve Bayes Classifier. The outcome of the research may assist road users in making better departure time decisions and alternative route choices. It will also enhance the efficiency in traffic management through fusion of weather data into Advanced Traveller Information System (ATIS) and Advanced Traffic Management System (ATMS). The novelty of this study is to model the functional relation between traffic congestion and weather for transportation agencies in Dhaka City to reduce traffic congestion, vis-à-vis, saving travel time and cost.

The first chapter of this manuscript discusses introduction and background prior researches, the next chapter illustrates methodology including NB classifier and study area, the third section contains results and discussion and the last section discusses concluding remarks and future study of the manuscript.

II. METHODOLOGY

Naïve Bayes Classifier

The Naïve Bayes (NB) Classifier assuming that all feature variables are conditionally independence for the value of the target variables. It calculates the probability of a target variable for a particular instances of feature variables and then predicts the class label of the target variable through the highest posterior probability [20]. NB classifier is simple, train vary fast, highly scalable and robust to noise data. Broadly, it is insensitive to irrelevant features, has computational simplicity and scales linearly with the number of feature variables.

Naïve Bayes Classifier based on conditional probability. Assume, n number of feature variables have a vector $x = (x_1, x_2, \ldots\ldots, x_n)$. Feature variables are mutually independent to each other. For each k class labels L_k the instance probabilities is $p(L_k|x_1, \ldots\ldots, x_n)$.

The conditional probability can be rewritten using Bayes' theorem, as:

$$p(L_k|\mathrm{x}) = \frac{p(L_k) \times p(x|L_k)}{p(x)} \qquad (1)$$

In the equation (1), $p(L_k|x)$ is posterior and in numerator, $p(L_k)$ is prior and $p(x|L_k)$ is likelihood. $p(x)$ in denominator is evidence, which is independent of L and for a given x_i, it is constant.

The numerator is equal to the joint probability $p(L_k, x_1, \ldots., x_n)$ and can be written by chain rule as following:

$$\begin{aligned} p(L_k, x_1, \ldots., x_n) &= p(x_1, \ldots., x_n, L_k) \\ &= p(x_1|x_2, \ldots.., x_n, L_k) \times p(x_2, \ldots.., x_n, L_k) \\ &= p(x_1|x_2, \ldots.., x_n, L_k) \times p(x_2|x_3, \ldots.., x_n, L_k) \ldots\ldots \times p(x_{n-1}|x_n, L_k) \times p(x_n|L_k) \times p(L_k) \qquad (2) \end{aligned}$$

Similarly, $p(x_i|x_{i+1}, \ldots., x_n, L_k) = p(x_i|L_k)$ for feature x_i.

The joint probability written as:

$$\begin{aligned} p(L_k|x_1, \ldots, x_n) &\propto p(L_k, x_1, \ldots.., x_n) \\ &= p(L_k) \times p(x_1/L_k) \times p(x_2/L_k) \times p(x_3/L_k) \ldots\ldots \\ &= p(L_k) \times \prod_{i=1}^{n} p(x_i|L_k) \qquad (3) \end{aligned}$$

The conditional distribution of class label L:

$$p(L_k|x_1, \ldots, x_n) = \frac{1}{w} \times \prod_{i=1}^{n} p(x_i|L_k) \qquad (4)$$

Where, $w = p(\mathrm{x}) = \sum_k p(L_k) \times p(x|L_k)$ is evidence. It is a scaling factor dependent on $x_1, x_2, \ldots x_n$.

The Naïve Bayes Classifier combines the independent feature model using maximum posteriori (MAP), which is a decision rule. Bayes classifier function assigns a class label as $\hat{y} = L_k$ for same class k as:

$$\hat{y} = \mathrm{argmax}_{k\epsilon(1,\ldots,k)}\, p(L_k) \times \prod_{i=1}^{n} p(x_i|L_k) \qquad (5)$$

In Gaussian Naïve Bayes Classifier (GNB), for observation value t, the probability distribution of t for a class L_k, $p(x = t|L_k)$ can be calculated by mean μ_k and variance σ_k^2:

$$p(x = t|L_k) = \frac{1}{2\,\pi\sigma_k^2} \times e^{-\frac{(t-\mu_k)^2}{2\sigma_k^2}} \qquad (6)$$

Study area

Traffic conditions in terms of congestions were observed for 15 intersections in Dhaka City during October 2018 to April 2019. The GPS locations of the study intersections are presented in the Table 1 and detail maps of intersections are in Figure 1. A group of surveyors were involved to collect 334 data regarding congestion situations and record corresponding weather conditions, type of days and time of days during field observations at different intersections.

Table 1: Designation and Location of study intersections in Dhaka City

Intersections	Intersection Name	GPS Location (Latitude and Altitude)
I-1	Mouchak	23°44′44″N, 90°24′44″E
I-2	Shantinagar	23°44′14″N, 90°25′06″E
I-3	Malibagh	23°44′39″N, 90°24′50″E
I-4	Farmgate	23°45′28″N, 90°23′24″E
I-5	PanthaPath	23°45′03″N, 90°23′13″E
I-6	Sonargaon	23°44′59″N, 90°23′13″E
I-7	GPO	23°43′39″N, 90°24′37″E
I-8	Nilkhet	23°43′57″N, 90°23′12″E
I-9	Science Lab	23°44′19″N, 90°23′00″E
I-10	New Market	23°43′56″N, 90°23′06″E
I-11	Russel Square	23°45′04″N, 90°22′41″E
I-12	Moghbazar	23°44′55″N, 90°24′13″E
I-13	Bonoshree	23°46′03″N, 90°25′22″E
I-14	Mohakhali	23°46′42″N, 90°23′53″E
I-15	Jahangir Gate	23°46′31″N, 90°22′23″E

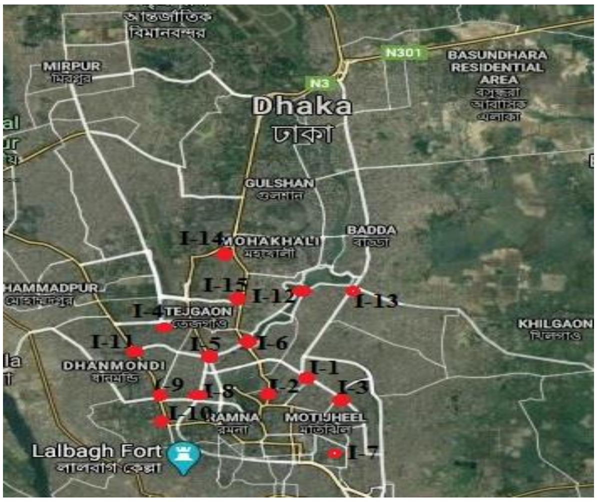

Figure 1: Location showing the study intersections in Dhaka city map

Features and target variables

The target variables of the study is congestion state of intersections. An intersection was marked as 'congested', if the queue of vehicle did not clear within a traffic signal cycle, i.e. there was a queue overflow at the intersection within the recorded time. On contrary, if the queue of vehicle cleared in every green time of a traffic signal, the intersection was marked as 'not congested' for that observation time period. The feature variables of the study are—time of day, type of day and weather condition. The day of field observations were distinguished according to working day types, such as, work days, weekends and holidays. And, the time of field observations were designated as morning rush hour, morning, afternoon, evening peak hour and night. Weather conditions were categorized as clean, light rain and heavy shower. The detail of features and target variables with explanation and their corresponding numerical representations are shown in the Table 2.

Table 2: Numerical representations of categorization of variables

Variables	Types	Explanation	Numerical representation
Time (Observation time)	Morning rush	8AM-10AM	1
	Morning	10AM-12PM	2
	Afternoon	12PM-4PM	3
	Evening rush	4PM-6PM	4
	Nigh	6PM-10PM	5
Day (types of work day)	Working day	Sunday to Thursday	1
	Week end*	Friday, Saturday (*Weekends in Bangladesh)	2
	Holiday	National holiday	3
Weather	Clean	Sunny, cloudy, windy	1
	Light rain	Drizzle with no wind	2
	Heavy rain	Heavy downpour with or without storm	3
Congestion state	Yes	Queue disappears within a signal cycle	1
	No	Long queue after each signal cycle	0

III. DATA ANALYSIS

Relevant details should be given including experimental design and the technique (s) used along with appropriate statistical methods used clearly along with the year of experimentation (field and laboratory).The data analysis used in this study composed of two parts—firstly, preparation of data for analysis and performing descriptive statistics. The data collected from the field surveys were recorded and converted the information into numerical representation as shown in Table 2. For example, the congestion state of traffic corresponding to 'Yes' and 'No' are '1' and '0' respectively. Morning rush, morning, afternoon, evening rush and night time shared 19.2%, 13.5%, 28.1%, 18.6% and 20.7% respectively among the time of day feature. Time of day feature consisted with working day (68.3%), weekend (21.6%) and holiday (10.2%). The weather condition feature comprised with clear, light rain and heavy rain corresponding to 61.7%, 19.5% and 18.9% respectively. The target variable i.e. traffic congestion state divided into Yes (48.8%) and No (51.2%).

Finally, the features and target variables were modeled using Naïve Bayes Classifier to identify relationship between weather condition and congestion state under different type and time of days. The model was developed in Scikit-learn v. 0.22.2 under python programing language. The pseudo code written in python is presented in the Table 3. The study found that, the motorist were less likely to prefer inclement to go outside. However, the rush hour motorists were more likely to ignore the weather condition, because of their urgency of work. Motorists in weekends and holidays were more rely on weather for the outing decisions, since, people used to travel for recreational and shopping trips require less urgency. Detail codes are presented in Appendix A.

Table 3: Algorithm: Modelling dataset using Naïve Bayes Classifier in Scikit-learn

Input: Weather data (*WD*), time of day (*TiD*), type of Day (*TD*)

Output: Congestion state (*CS*)

1. Import dataset
2. *y* ⟵ Output
x ⟵ Input
3. dataset *D* ⟵ Matrix (x, y)
4. Split D by Training (*Tr*) and Testing (*Te*)
5. *Tr* ⟵ *x_train*, *y_train*
Te ⟵ *x_test*, *y_test*
6. *gnb* ⟵ function *GaussianNB*()
7. *y_predicted* ⟵ *gnb.fit* (*Tr*).*x_predicted*
8. Accuracy_score(*y_test, y_predicted*)

IV. RESULTS AND DISCUSSION

Confusion matrix and error analysis

The NB classifier is a data driven model. The dataset was split by training and testing datasets. Among 334 data, 50% data were used for training and the rest of the data were used for testing. The proposed NB classifier have 72.25% prediction accuracy in training data and 85.03% accuracy in testing data. Mean absolute error (MAE) and Root mean square error (RMSE) were calculated using following formulas:

$$\text{MAE} = \frac{1}{n}\sum_{i-1}^{n}|P_i - E_i| \quad (7)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n}\frac{(P_i - E_i)^2}{n}} \quad (8)$$

Where, P_i = predicted value, E_i = observed value and n = number of observations.

The MAE and RMSE of the proposed NB classifier model were 0.28 and 0.46 respectively. Confusion matrix of testing datasets are presented in the Figure 2. When the observed values and its corresponding predicted values categorize congestion state as 'No', i.e. uncongested condition, the outcomes are true positive (TP). Similarly, when both values categorize congestion state as 'Yes', i.e. congested condition, the outcomes are true negative (TN). The %TP and %TN of the proposed model are 85.71% and 84.21% respectively (Figure 2). False positive (FP) and false negative (FN) are the error estimations of the model. When observed value shows traffic congestion, but the corresponding predicted value shows uncongested condition, the outcomes are FN and vice-versa for the outcomes of FP. Errors of prediction i.e. FP and FN outcomes are 15.79% and 14.29% respectively in the proposed model. % accuracy considers equal weight for FP and FN errors. Precision and recall represent % correctness in each category and higher values depict the observations are categorized i.e. labelled correctly. Values of precision and recall are 86.67% and 54.93% respectively. The f-1 score uses harmonic mean, is an alternative measure to assess the correctness of the model and the value is 67.24%.

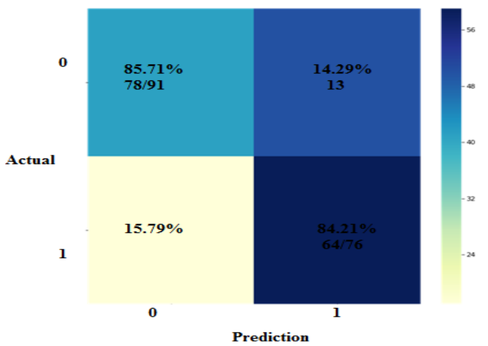

Figure 2: Confusion matrix of testing dataset

Model validations

In order to validate the test results, multinomial logistic regression (MLR) is employed to categorize the datasets. SPSS v. 24.0 has been used to model the categorical relationship among features with congestion state. Accuracy obtained by MLR is 63.80%. Our proposed NB classifier performs better (85.03% accuracy in testing) than MLR for congestion prediction.

Table 4: Classification table of MLR model

Classification			
	Predicted		
Observed	0	1	Percent Correct
0	109	62	63.70%
1	59	104	63.80%
Overall Percentage	50.30%	49.70%	63.80%

Feature importance

Weather condition, types of day and time of day are the features in this study. Importance of the features are calculated using permutation of features. Within three set of features, one feature is dropped and computed the MAE of the model. The value of MAE increases as a feature drop and the features are ranked based on ascending order of the MAE value, i.e. the model without a feature, which has higher MAE have more importance priority. Relative importance of the features are computed with respect to the most important feature.

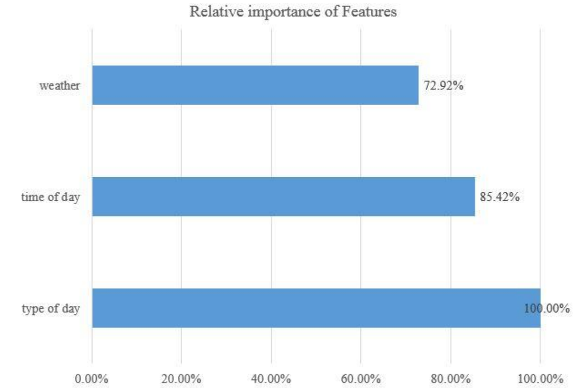

Figure 3: Relative importance of features in the proposed NB classifier model

Among the three features, types of days, i.e. working day, weekend and holidays have most influence on congestion condition at the intersections. Time of days, i.e. morning rush, morning, afternoon, evening rush, night have 85.42% relative importance. On contrary, weather conditions, such as, clean, light rain and heavy rain have lowest importance 72.92% on congestion condition at intersections (Figure 3). It means that, motorist decision to travel much influenced by type of day and time of day rather than weather phenomena.

V. CONCLUSION

The proposed NB classifier model can predict influence of weather condition on traffic congestion state under various scenarios with better accuracy. The competitive multinomial logistic regression is less suitable to model such phenomena. NB classifier is faster, robust and computational efficient technique. The suitability of the model increases with the numbers of training data and other features which are highly influence the traffic congestion.
Traffic congestion influences traffic flow, travel time, vehicular speed and traffic density in a roadway. Our study endeavoured to incorporate the real-world information, such as, type of time of day, such as, working day, weekend, morning hour, afternoon, evening rush hour and so on, for the prediction of congestion state and related those to the corresponding weather condition. The study opens the door to use incident, delay and level of service (LOS) to determine their relation with weather condition.

Since congestion losses valuable working hour, information of congestion state in advance can save the time of road users. Incorporating weather information into Advanced Traveller Information System (ATIS) can help motorists to decide while travelling, which intersections to use and when the intersection is free from congestion precisely. Besides, concern authorities can use weather data incorporating Advanced Traffic Management System (ATMS) to make appropriate decision on congestion reduction dynamically. An extensive study regarding the influence of weather on traffic congestion at intersections in Dhaka city needs to be perform before wide scale practical implementation.

ACKNOWLEDGMENT

The research is supported by the Department of Civil Engineering, Daffodil International University (DIU), Bangladesh. The author expresses his gratitude to the undergraduate students of Daffodil International University (DIU), who were involved in the conduction of field survey.

REFERENCES

[1] A. Luchetta, S. Manetti, and F. Francini, "Forecast: a neural system for diagnosis and control of highway surfaces," *IEEE Intelligent Systems*, Vol. **13**, No. **3**, pp. **20–26**, **1988**.

[2] M. Aron, M. Ellenberg, P. Fabre, and P. Veyre, "Weather related traffic management," *In Towards an Intelligent Transport System Proceedings of the First World Congress on Applications of Transport Telematics and Intelligent Vehicle Highway Systems*, Vol. **3**, pp. **1089–1096**, **1995**.

[3] R. Kulmala, "Recent developments in weather related traffic management," *In Proceedings volume from the 8th IFAC/IFIP/IFORS Symposium on Transportation Systems*. Vol. **2**, pp. **711–714**, **1997**.

[4] D.J. Dailey and Washington State Transportation Commission, "The use of weather data to predict non-recurring traffic congestion (No. TNW2006-11)," *Transportation Northwest (Organization)*, **2006**.

[5] T. Maze, M. Agarwai, G. Burchett, "Whether matters to traffic demand, traffic safety, and traffic operations and flow," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. **1948**, pp. **170-176**, **2016**.

[6] A. Ibrahim, F. Hall, "Effect of adverse weather conditions on speed-flow-occupancy Relationships," *Transportation Research Record: Journal of the Transportation Research Board*, Vol. **1457**, pp. **184-191**, **1994**.

[7] S. Datla, S. Sharma, "Impact of cold and snow on temporal and spatial variations of highway traffic," *Journal of Transport Geography*, Vol. **16**, No. **5**, pp. **358-372**, **2008**.

[8] M. Hofmann and M. O'Mahony, "The impact of adverse weather conditions on urban bus performance measures," *In Proceedings. 2005 IEEE Intelligent Transportation Systems*, pp. **84-89**, September **2005**.

[9] R. Hranac, E. Sterzin, D. Krechmer, H. Rakha, M. Farzaneh, "Empirical studies on traffic flow in inclement weather," *FHWA-HOP-07-073, Federal Highway Administration, USA*. **2006**.

[10] H. Ikeuchi, K. Hatoyama, R. Kusakabe and I. Kariya. "Development of a Statistical Model to Predict Traffic Congestion in Winter Seasons in Nagaoka," *Japan Using Publicly Available Data. In Intelligent Transport Systems for Everyone's Mobility*, *Springer*, Singapore, pp. **265-278**, **2019**.

[11] B. Deb, S.R. Khan, K.T. Hasan, A.H. Khan and M.A. Alam. "Travel time prediction using machine learning and weather impact on traffic conditions," *In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT) IEEE*, pp. **1-8**, March **2019**.

[12] Y. Liu and H. Wu. "Prediction of road traffic congestion based on random forest," In 2017 *10th International Symposium on Computational Intelligence and Design (ISCID) IEEE*, Vol. **2**, pp. **361-364**, December **2017**.

[13] A.S. Tomar, M. Singh, G. Sharma and K.V. Arya. "Traffic management using logistic regression with fuzzy logic," *Procedia computer science*, Vol. **132**, pp. **451-460**, **2018**.

[14] I. Rish. "An empirical study of the naive Bayes classifier," *In IJCAI 2001 workshop on empirical methods in artificial intelligence*, Vol. **3**, No. **22**, pp. **41-46**, August **2001**.

[15] N. Friedman, D. Geiger, M. Goldszmidt, "Bayesian Network Classifiers," *Machine Learning*, Vol. **29**, No. **131–163**, **1997**. DOI:10.1023/A:1007465528199

[16] G. Wang and J. Kim, "The prediction of traffic congestion and incident on urban road networks using naive bayes classifier," *In Australasian Transport Research Forum (ATRF), 38th*. January **2016**.

[17] F.I. Rahman, A. Hasnat, A.A. Lisa, "Traffic flow prediction by incorporating weather information in Naïve Bayes Classifier," *Journal of Advanced Civil Engineering Practice and Research*, Vol. **8**, pp. **10-16**, **2019**. Available at: http://ababilpub.com/download/jacepr8-3/

[18] S. Saneinejad, M.J. Roorda, and C. Kennedy, "Modelling the impact of weather conditions on active transportation travel behavior", *Transportation research part D: transport and environment*, Vol. **17**, No. **2**, pp. **129-137**, **2012**.

[19] Q. Liu, J. Lu, S. Chen, and K. Zhao, "Multiple Naïve bayes classifiers ensemble for traffic incident detection," *Mathematical Problems in Engineering*, **2014**.

[20] M.L. Zhang, J.M. Peña, and V. Robles. "Feature selection for multi-label naive Bayes classification," *Information Sciences*, Vol. **179**, No. **19**, pp. **3218-3229**, **2009**.

AUTHORS PROFILE

Mr. Saurav Barua completed B.Sc., Bangladesh University of Engineering and Technology (2008), M.S., Rutgers, the State University of New Jersey, USA (2012). He is currently working as Assistant Professor in Department of Civil Engineering, Daffodil International University, Dhaka, Bangladesh since 2019. He has published more than 20 research manuscripts in reputed international journals and conferences including Elsevier, Springer and Taylor & Francis, which are available online. His main research work focuses on Transportation Engineering and Machine learning. He has 6 years of teaching experience in several universities and 8 years of engineering and research experiences.

