

Research Article

Predicting Freeway Work Zone Delays and Costs with a Hybrid Machine-Learning Model

Bo Du,¹ Steven Chien,^{2,3} Joyoung Lee,³ and Lazar Spasovic³

¹HNTB Corporation, Arlington, VA, USA

²School of Automobile, Chang'an University, Xi'an, Shaanxi, China

³John A. Reif, Jr. Department of Civil and Environmental Engineering, New Jersey Institute of Technology, Newark, NJ, USA

Correspondence should be addressed to Steven Chien; chien@njit.edu

Received 24 March 2017; Accepted 3 July 2017; Published 7 August 2017

Academic Editor: Seungjae Lee

Copyright © 2017 Bo Du et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A hybrid machine-learning model, integrating an artificial neural network (ANN) and a support vector machine (SVM) model, is developed to predict spatiotemporal delays, subject to road geometry, number of lane closures, and work zone duration in different periods of a day and in the days of a week. The model is very user friendly, allowing the least inputs from the users. With that the delays caused by a work zone on any location of a New Jersey freeway can be predicted. To this end, tremendous amounts of data from different sources were collected to establish the relationship between the model inputs and outputs. A comparative analysis was conducted, and results indicate that the proposed model outperforms others in terms of the least root mean square error (RMSE). The proposed hybrid model can be used to calculate contractor penalty in terms of cost overruns as well as incentive reward schedule in case of early work competition. Additionally, it can assist work zone planners in determining the best start and end times of a work zone for developing and evaluating traffic mitigation and management plans.

1. Introduction

Highway maintenance activities usually require lane closures, frequently disrupt traffic operations, and increase delays because of restricted capacity. According to an urban mobility report [1], traffic congestion in 2014 caused urban Americans to travel 6.9 billion hours more and to consume extra 3.1 billion gallons of fuel, which is equivalent to a congestion cost of 160 billion dollars. The delays caused by work zones are nearly 24% of all nonrecurring delay and 10% of overall delay congestion in the United States [2]. Work zones have become a necessary feature on US highways, and they have been the second largest contributor to the nonrecurring delay congestion. The growth in the vehicle-miles travel has far exceeded the increase of lane-miles built on highways. Extending the lifecycle of existing roads via regular maintenance and efficiently utilizing the available capacity to meet the mobility needs are highly desirable actions.

Transportation systems, especially roadway networks, form an integral set of connections for the movement of

passengers and goods, thus aiding in progressive economic development. Restricted capacity due to lane closures and growing demand deteriorate the mobility of interconnected road networks spatially and temporally. Roadway rehabilitation and reconstruction projects usually require traffic lane and/or shoulder closures. These activities result in reduced reliability, increased delays, more frustration for travelers, and more wasted fuel. To improve the quality of life, the environment, and commerce, it is desirable to develop a tool that can correctly predict work zone impacts as well as evaluate the effectiveness of congestion mitigation strategies.

The concept of deterministic queuing theory has been widely used to predict work zone delays based on the relationship between approaching traffic volume and restricted capacity, because of its simplicity of implementation. However, the predicted delays were sometimes inaccurate because the heterogeneity of road geometry (i.e., number of lanes, ramp junctions) and traffic volumes (i.e., fluctuations in numbers and percentages of cars and trucks) upstream of the work zone were simplified.

Machine-learning techniques can be applied to recognize traffic patterns and dynamically self-adjust in order to gain in prominence and maturity. Artificial neural networks (ANNs) have been widely used for predicting work zone delays. With sufficient traffic data, ANNs outperformed other parametric models in predicting work zone impacts because of their flexibility and ability to capture dynamic speed and delay changes over space and time. However, most ANNs developed previously were based on traffic data collected by loop detectors. The loop detector data used to suffer from costly operation and maintenance as well as the accuracy of predicted space-mean speed. The emergence and steady increase in public deployment of probe-vehicle-based data collection systems, providing greater temporal and spatial coverage at relatively inexpensive cost, shall be adapted to predict work zone impacts.

This study aims to develop a hybrid machine-learning model, integrating an artificial neural network (ANN) and a support vector machine (SVM) model, to predict spatiotemporal work zone impact on the New Jersey freeways. It is worth noting that the work zone impacts (i.e., speeds and delays) predicted in this study reflect those conditions that occurred in the roadway segments upstream of the study work zone. The proposed ANN employs the relationship traffic volume and road capacity to predict speeds and delays, while the proposed SVM model is in charge of estimating the restricted road capacity caused by a work zone.

2. Literature Review

Motorists often experience excessive delays caused by restricted work zone capacity, especially during peak hours. Numerous studies have focused on developing models to predict work zone delay. Previous modeling approaches in delay estimation and prediction can be classified into parametric, simulation, and nonparametric approaches.

Parametric models commonly employed the deterministic queuing theory for predicting work zone delay [3–6]. Chien and Schonfeld [4] used deterministic queuing theory to predict user queuing delay caused by a work zone using a single lane closure on a four-lane highway (i.e., two lanes per direction). Du and Chien [6] formulated the work zone delay considering time-varying traffic pattern, work zone capacity adjustment factors (e.g., light condition, heavy-vehicle percentage, and lane width), and shoulder usage. Deterministic queuing models are suitable for predicting delay for planning purposes but sometimes they have a limited ability to provide accurate prediction (i.e., delays), especially under significantly fluctuated traffic condition over time [7, 8]. Further, the applied work zone capacity was either given or based on some simplified empirical equations, which also degrade the accuracy of prediction.

Another well-known parametric approach for delay prediction is shockwave theory, first introduced by Lighthill and Whitham [9] and Richards [10]. The shockwave theory assumes that traffic flow is analogous to fluid flow and employs a flow-speed-density relationship to analyze the transition of traffic flow over space and time. Since the shockwave theory has been focused mainly on noncongested traffic

conditions, it is therefore not very reliable for delay prediction under congested traffic conditions [11].

Recently, researchers applied microscopic traffic simulation to quantify work zone delay [12–14]. Well-calibrated simulation models are capable of generating high fidelity traffic measures under various work zone configurations. CORSIM [12] and VISSIM [15] are among the most widely used models. However, the simulation approach for delay prediction suffers from high computational time and the results only represent the traffic measures for a specific work zone on a specific segment of a highway.

To overcome the limitations of parametric and simulation approaches, nonparametric models were developed. Many studies have successfully applied ANNs to predict various traffic measures, such as traffic flow [16, 17], freeway work zone capacity [18, 19], and work zone delay [7, 8, 20]. Ghosh-Dastidar and Adeli [20] presented a multilayer feedforward neural network model (i.e., Levenberg-Marquardt neural network model) for delay and queue length prediction at freeway work zones. However, the ANN model was trained using simulated data instead of real data collected from the fields.

With technological advancements, a wide variety of massive traffic data from infrastructure sensors and probe vehicles has become increasingly available. This new and rich data has made way for big data analytics as an emerging method for predicting freeway spatiotemporal work zone delay. Du et al. [7] developed a multilayer feedforward ANN model to predict work zone delay using the probe-vehicle data (i.e., speeds under normal and work zone conditions) subject to the condition when traffic volume and capacity information are missing. Based on the prediction results of three examples, we found that the ANN model outperformed analytical models in terms of the accuracy in predicting delays caused by reconstruction projects. However, the accuracy of that ANN model can be improved if the relationship of approaching traffic volume and work zone capacity can be captured. Hence, this paper intends to enhance the proposed ANN by integrating the proposed SVM model, responsible for predicting the work zone capacity subject to various work zone configurations.

3. Data Collection

To develop a model for accurately predicting speed/delay caused by an expected work zone with lane closures on New Jersey freeways, it requires a robust database to host and aggregate the data from multiple data sources. Based on the use of available work zone data for year 2014, the collected data, along with the corresponding databases applied for developing the proposed model, are listed below:

- (i) OpenReach: OpenReach [21] is an incident reporting system, in which work zone information is also included. The work zone information needed for model development includes work zone location, starting/ending time, number of closed lanes, and duration and length, to name a few.
- (ii) New Jersey Straight Line Diagram (NJS LD): NJS LD [22] is a road geometry database maintained by the

New Jersey Department of Transportation (NJDOT), which provides the roadway inventory and geometry data. The work zone information needed from NJSLD includes road name, functional classification, direction, and total number of lanes.

- (iii) New Jersey Congestion Management System (NJCMS): NJCMS [23] is a data management and data analysis system used for predicting congestion impacts on New Jersey highways. It has the weekday traffic data that is necessary for predicting work zone impacts and user delay cost.
- (iv) Probe-vehicle data: traffic speeds for freeway segments: the main traffic speed data used in this study was the historical speed data reported by INRIX [24], which provides space-mean speeds based on data collected from probe vehicles. There are more than 1,200 Traffic Message Channels (TMCs) in New Jersey covering interstate and express freeways. The INRIX raw speed data, which included more than 1.5 billion records, was collected for 24 hours a day over a one-year period, from January to December 2014.
- (v) Plan4Safety: crash records provide crash location, date, and time of the crash for New Jersey highways [25], which are then used to screen out work zones where accidents occurred.

In order to identify work zones with full and accurate information needed for model development, Figure 1 illustrates a data cleaning procedure applied to identify work zone data suitable for developing the proposed model.

Step 1. Identify historical work zone events from the OpenReach incident database. Remove work zones with uncompleted information (e.g., missing work zone milepost, starting/ending date, and duration).

Step 2. Add the standard route identifier (SRI), work zone direction, and number of lanes-closed information to each work zone based on the NJSLD database.

Step 3. Neglect accident-related historical work zones by crosschecking accidents recorded in the Plan4Safety database.

Step 4. Map the aggregated 15-minute speed data from INRIX for each TMC located upstream of each work zone identified in Step 3.

After conducting the data cleaning procedure, there are 181 work zones qualified for developing the proposed model.

4. Methodology

Based on the limitations of previous models discussed in the literature review, a hybrid machine-learning model is proposed for predicting the spatiotemporal delays caused by a prescheduled freeway work zone. The model integrates an SVM model and a multilayer feedforward ANN model. As discussed earlier, the SVM model is in charge of predicting

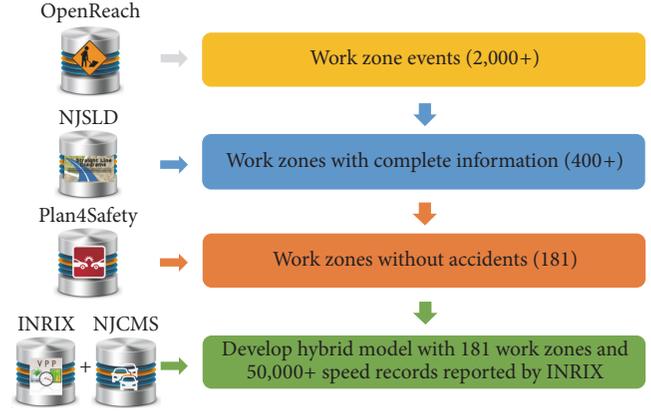


FIGURE 1: Data processing procedure.

the restricted capacity caused by a work zone, and the ANN model is for predicting delays.

As discussed in the literature review, SVM is very friendly to use and has the particular strength of overcoming overfitting problem and local minima. In this study, the proposed SVM employed six training vectors obtained from the 181 work zones identified in previous section, including number of lanes, number of open lanes, work zone length, upstream traffic volume, truck percentage, and average upstream speed, suggested by the Pearson and Spearman test. These work zones were randomly divided into three subsets for training, validating, and testing the proposed model. Then, the SVM model mapped these training vectors into a higher dimensional space using a radial basis kernel function. Then the optimal hyperplanes were determined by maximizing the margins of the training vectors. Finally, a trained SVM model was created to predict the work zone capacity (C_w).

To be more specific, given a nonlinear training data set of l instance-label pairs (x_p, C_{wp}) , $p = 1, \dots, l$, where l is the total number of training samples (i.e., work zones), $x_p \in R^6$ consists of six training vectors identified earlier, and $C_{wp} \in R$ is the work zone capacity of the corresponding sample p . The nonlinear relationship between x_p and C_{wp} can be linearized as

$$C_w = f(x) = \omega^T \phi(x) + \beta, \quad (1)$$

where ω is a vector perpendicular to the hyperplane, T is the transposition of the matrix, β is a constant, and ϕ is a nonlinear transformation from R^6 to a higher dimensional space. To find ω and β , SVM requires the solution by solving the following optimization problem:

$$\begin{aligned} \min_{\omega, \beta, \xi_p} \quad & \frac{1}{2} \omega^T \omega + \lambda \sum_{p=1}^l \xi_p \\ \text{subject to} \quad & C_{wp} (\omega^T x_p + \beta) \geq 1 - \xi_p \\ & \xi_p \geq 0, \end{aligned} \quad (2)$$

where ξ_p is a slack variable and λ is a regularization parameter. With the implementation of a sequential minimal

optimization algorithm [26] and performing an exhaustive search [27, 28], the SVM model is developed.

To develop the ANN model, a Pearson and Spearman test was conducted. Results show that the factors affecting the speeds of upstream work zone include average speed of upstream segment i at time j under normal condition (s_{ij}); traffic volume approaching work zone at time j (Q_j); work zone capacity (C_w); and distance from segment i to the work zone (d_i). Therefore, the average speed of upstream segment i at time j under work zone condition (y_{ij}) can be represented by

$$y_{ij} = f(s_{ij}, Q_j, C_w, d_i). \quad (3)$$

As Q_j increases, the resulting travel time and delay increase, especially when it is close to the restricted capacity caused by a work zone. To represent the relationship among s_{ij} , Q_j , and C_w , the concept of the BPR function [29] was adapted, assuming that the weighted speed of segment i at time j denoted as v_{ij} is the historic speed under normal condition multiplied by a reduction factor that is a function of approaching volume to work zone capacity ratio. Thus,

$$v_{ij} = \frac{s_{ij}}{1 + a(Q_j/C_w)^b}, \quad (4)$$

where s_{ij} is the average speed of segment i at time j under normal condition (km/h); Q_j is the traffic volume approaching the work zone at time j (vph); C_w is the work zone capacity (vph) predicted from SVM model [30]; a and b are the arrays of freeway model coefficients; i is the i th freeway segment upstream of work zone ($1 \leq i \leq m$); j is the j th time interval after work zone started ($1 \leq j \leq n$); m is the number of segments upstream of work zone; and n is the number of time intervals (e.g., 15 minutes per interval) between work zone starting time and ending time.

With the weighted speed (v_{ij}) from (4) and the distance from segment i to the work zone (d_i), the work zone speed (y_{ij}) can be represented by

$$y_{ij} = f(v_{ij}, d_i). \quad (5)$$

The Neural Network Toolbox in MATLAB [31] was used for developing the ANN model. As discussed earlier, there were 181 qualified work zones, which were divided into three groups (i.e., 70%, 20%, and 10% of total work zones, resp.) for training, validation, and testing purposes. The root mean square error (RMSE) formulated as (6) was used as an index to determine the optimal parameters of a and b in (4), the suitable training algorithm, and optimal numbers of hidden layers and neurons. The lower the RMSE value, the better the model performance.

$$\text{RMSE} = \sqrt{\frac{1}{mn} \sum_{i,j} (\hat{y}_{ij} - y_{ij})^2}, \quad (6)$$

where \hat{y}_{ij} is the predicted speed of segment i at time j under work zone condition (km/h); and y_{ij} is the INRIX reported

speed of segment i at time j under work zone condition (km/h).

By performing an exhaustive searching, the optimized a and b were 0.1 and 2.7, respectively, and the proposed ANN model consists of eight neurons and one hidden layer. The Levenberg-Marquardt algorithm was applied to train the ANN, which yielded the lowest RMSE.

The proposed hybrid machine-learning model is shown in Figure 2, which is an ANN model integrating an SVM model for approximating work zone capacity. The SVM model employed six training vectors defined previously to predict work zone capacity (C_w). The ANN model consists of an input layer with two neurons representing the weighted speed (v_{ij}) and distance from upstream segment i (d_i), one optimized hidden layer with eight neurons, and an output layer with one neuron representing predicted work zone speed (\hat{y}_{ij}). In the input layer, the predicted work zone capacity (C_w) from SVM model along with normal speed (s_{ij}) and approaching traffic volumes (Q_j) was used for calculating the weighted speed (v_{ij}). It is worth noting that the proposed model can predict speeds since the beginning of a freeway work zone until 2 hours after the work zone has been removed. The model predicts speeds up to 16 km upstream of the work zone.

As mentioned earlier, the accurate prediction of traffic delay is of utmost importance in supporting the efficient planning of work zones for transportation agencies (e.g., traffic management centers, metropolitan planning organizations, and state DOTs). The predicted spatiotemporal speeds under work zone condition with the proposed hybrid model can be used for assessing work zone impacts (e.g., delay and delay cost). The work zone delay (D) can be defined as the additional delay produced by the reduced speed caused by the work zone (\hat{y}_{ij}) over the normal speed (s_{ij}), which can be calculated by

$$D = \sum_{i,j} \max \left\{ l_i \left[\frac{1}{\hat{y}_{ij}} - \frac{1}{s_{ij}} \right] V_{ij}, 0 \right\}, \quad (7)$$

where D is the total delay caused by the work zone (veh-hr); l_i is the length of freeway segment i (km); and V_{ij} is the traffic counts of segment i at time j (veh).

In addition to delay, the proposed hybrid model determines the delay cost to road users caused by work zones. Considering the values of travel time delay for passenger cars and trucks, the delay cost is equal to the sum of delays consumed by passenger cars and trucks multiplied by the corresponding values of time. Thus,

$$C_d = D(P_c \mu_c + P_t \mu_t), \quad (8)$$

where C_d is the delay cost to road users caused by the work zone (\$/zone); P_c is the percent of passenger cars; P_t is the percent of trucks; μ_c is the value of travel time delay for passenger cars (\$/veh-hr); and μ_t is the value of travel time delay for trucks (\$/veh-hr).

The percent of passenger cars and trucks can be obtained from NJCMS and the monetary values of travel time delay for passenger cars and trucks were 18.15 \$/veh-hr and 30.25 \$/veh-hr, respectively [32].

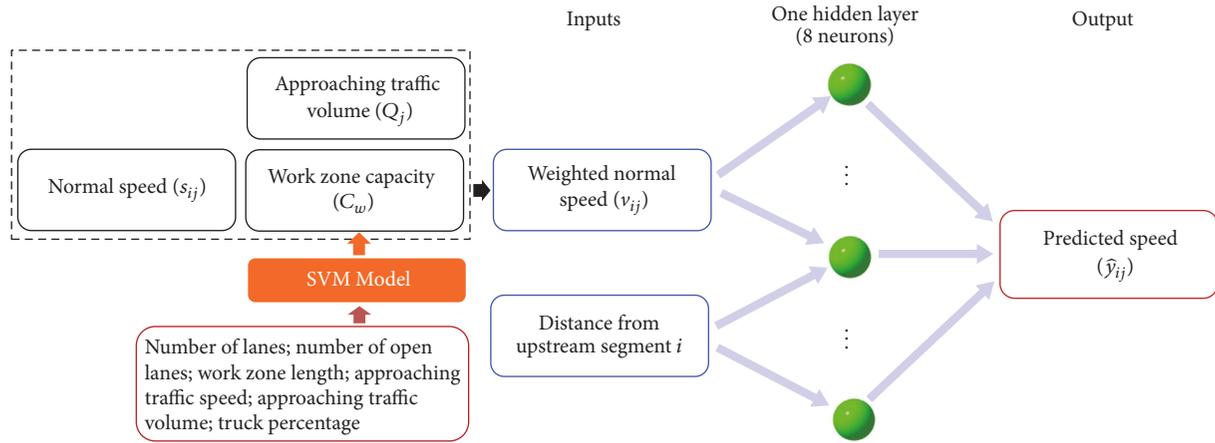


FIGURE 2: Configuration of the proposed hybrid model.

5. Model Evaluation

In this section, scenario-based analyses were conducted to evaluate the effectiveness of the proposed hybrid model. The first analysis was to test the model with historic data, using randomly selected 10% of 181 work zones. The RMSE was an indicator to measure the difference between the predicted and INRIX reported speeds under the work zone condition. In the second analysis, the proposed model was evaluated with new work zones in 2015, in which congestion delay and delay cost were applied to assess the model performance. Results from the proposed hybrid model (Model 3) were compared with the prediction results using Models 1 and 2.

Model 1. It is a previous ANN model developed by Du et al. [7] in which the relationship between approaching traffic volume and work zone capacity was not applied to predict work zone speeds.

Model 2. The proposed ANN model with the work zone capacity suggested by the Highway Capacity Manual (HCM) [33] was applied. Note that the work zone capacity suggested by the HCM can be formulated as in (9).

Model 3. The proposed ANN model with the work zone capacity predicted by SVM was applied.

$$C_w = (1600 + I) f_{HV} N_o - R, \tag{9}$$

where C_w is the work zone capacity (vph); I is the adjustment factor for type and intensity of work activity (vphpl); f_{HV} is the heavy-vehicle adjustment factor indicated in the HCM [33]; N_o is the number of open lanes; and R is the ramp volume (vph).

Firstly, we assessed the RMSEs with all models based on 10 historic freeway work zones occurring in 2014 with different lane configurations. As shown in Table 1, after introducing traffic volume and work zone capacity (i.e., Models 2 and 3), the RMSEs in general are significantly reduced, comparing to those with Model 1 for all testing work zones on freeways

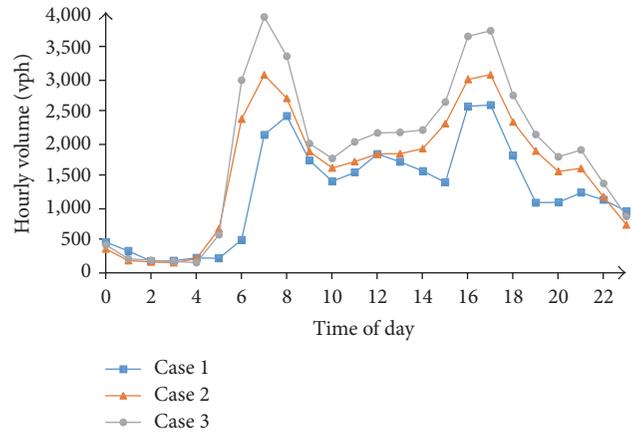


FIGURE 3: Hourly traffic distribution.

with different numbers of lanes per direction. Moreover, Model 3 with SVM has demonstrated itself as a better tool in predicting spatiotemporal work zone impacts. Note that a low RMSE indicates that the predicted work zone speed is accurate and reliable and lies within a narrow band offsetting the INRIX reported speed.

In the second analysis, the comparison was conducted by using three short-term work zones performed in 2015, and the characteristics of them, which consist of different attributes, such as time of a day, road geometry, and traffic pattern, are shown in Table 2. Case 1 was a 3.6 km long work zone with two-lane closure on a three-lane I-78 westbound section on October 14, 2015, between 11 PM and 6 AM. Case 2 was a 0.5 km zone on NJ-21 southbound. Case 3 was a 0.3 km zone on I-280 eastbound. In addition, the work zone capacities suggested by the SVM model as well as the HCM [33] are summarized in Table 2. The hourly traffic distributions for all three cases are shown in Figure 3, which can be used for calculating work zone delay and cost.

For each case, the delays were calculated using all the three models and compared to the “ground truth” information based on INRIX reported speeds. In addition to delay

TABLE 1: RMSE of the three models (km/h).

Number of lanes per direction	Number of tested work zones	Model 1	Model 2	Model 3
2	5	6.2	6.1	3.9
3	9	8.2	5.8	3.6
4 and more	4	6.9	6.7	4.4
Overall		7.3	6.1	3.9

TABLE 2: Work zone characteristics of case studies.

	Case 1	Case 2	Case 3
Route	I-78 WB	NJ-21 SB	I-280 EB
Number of lanes	3	3	2
Milepost	47.3–49.3	4.2–4.5	14.1–14.3
Work zone length (km)	3.6	0.5	0.3
Starting time	11 PM, 10/2015	10 AM, 11/2015	10 AM, 12/2015
Ending time	6 AM, 10/2015	3 PM, 11/2015	3 PM, 12/2015
Duration (hours)	7	5	5
Number of closed lanes	2	1	0 ^a
C_w with SVM (vph)	1,524	3,222	3,798
C_w with HCM (vph)	1,395	2,976	3,910

Note. ^aShoulder closure.

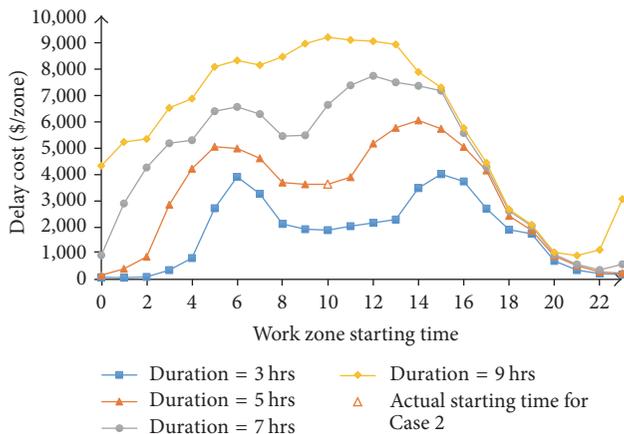


FIGURE 4: Predicted delay costs versus starting time for various work zone durations (Case 2).

and delay cost in Table 3, the numbers in the parentheses represent the error percentage occurring between the predicted and ground truth delay, which indicates that the model performance, in terms of prediction accuracy, is consistent with what we found in Table 2. The delay predicted by Model 3 is closer to the ground truth, which outperforms Models 1 and 2. Note that the delay cost was computed using (8). It is also worth noting that for Case 1, the error percentage differences of three models seem minor because of low traffic volumes during nighttime. When work zones are placed in daytime with higher traffic volumes (i.e., Cases 2 and 3), Model 3 becomes very effective and outperformed other two models.

Figure 4 shows the variation of delay cost versus start time under Case 2 with various work zone durations. Considering a 5-hour work zone, we found that the best starting time that

yields the least cost would be 12 AM. However, if this work zone must be performed during the daytime (i.e., between 6 AM and 6 PM), the best starting time would be 10 AM. We also found that when the 5-hour work zone ends close to or at peak hours, the residual queue must wait for extra time to be cleared, which results in more delay and cost. As the duration is greater than 7 hours, the delay cost reaches the minimum at 10 PM because of light traffic volumes between 10 PM and 5 AM. Figure 4 also indicates that a work zone performed in the daytime with longer duration would raise the delay cost, especially if the work zone schedule crosses peak hours. In general, low delay cost may be expected as the work zone is performed during the nighttime, albeit the labor cost is expected to be high. This also explains the work zone practices often seen in daily commutes.

In addition, the delay costs for various starting times and durations, as shown in Figure 4, can be used as a guideline to form the basis for awarding or deducting payments to contractors for early and late project completions, respectively. For example, in Case 2, assuming that the contractor delays two hours to open the closed lane for traffic (i.e., takes seven hours instead of five hours to complete the work). The transportation agency could charge \$3,005 in penalties to the contractor, for late completion because of the excess delay. Note that this charge may vary depending on the traffic volume distribution, work zone starting time, and the duration of the work zone.

Figure 5 illustrates and exposes the delay cost versus the start time for various demand levels, from 80% to 150% of the original volume in Case 2. We found that the costs are close when the start time ranges from 11 PM to 3 AM (next day) for various demand level because of light traffic. However, the delay cost significantly increases as approaching traffic volume increases, especially if the work zone schedule crosses

TABLE 3: Results comparison.

	Models	Case 1	Case 2	Case 3
Delay ^a (error %)	Model 1	60 (12%)	179 (17%)	152 (14%)
	Model 2	78 (15%)	170 (11%)	158 (11%)
	Model 3	75 (10%)	165 (8%)	169 (5%)
	Ground truth ^c	68	153	177
Delay cost ^b	Model 1	1,306	3,894	3,297
	Model 2	1,698	3,700	3,427
	Model 3	1,633	3,586	3,666
	Ground truth ^c	1,480	3,330	3,840

Note. ^aUnit: veh-hr; ^bUnit: \$; ^cINRIX reported speeds.

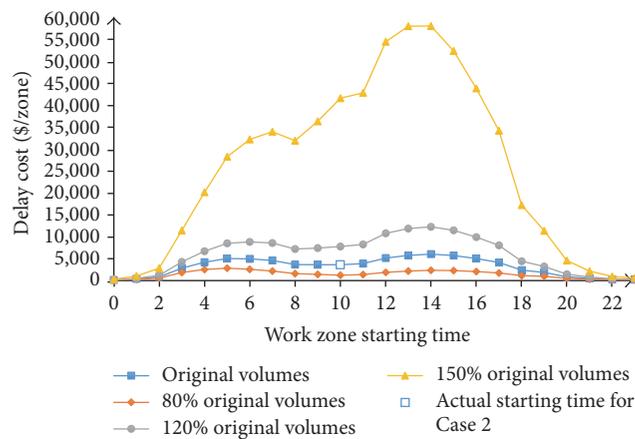


FIGURE 5: Predicted delay costs versus starting time for various traffic volumes (Case 2).

peak hours. The results would give transportation agencies a competitive edge by examining the delay costs versus work zone starting time subject to different traffic volumes.

6. Conclusions

In this paper, a hybrid machine-learning model was developed using work zone, road geometry, traffic volume, and probe-vehicle data to predict users' delays on the freeway segments upstream of a work zone in New Jersey. The restricted capacity caused by lane closure was approximated using SVM, which serves as an input of the proposed ANN model to predict the spatiotemporal speeds under work zone condition.

A total number of 181 historical work zones occurring in 2014 on New Jersey freeways were obtained, which are used to train the proposed hybrid model and to test its performance. The root mean square error (RMSE) is employed for performance evaluation. Comparing with a previous ANN model [7], we found that the developed model here with the work zone capacities suggested by HCM [33] and SVM has demonstrated itself as a sound model which improved prediction accuracy in terms of reduced RMSE. The results also suggest that the proposed hybrid machine-learning model with SVM outperforms the others for all three real-world study cases

with greater prediction accuracy, especially when work zones are placed in daytime facing high traffic volumes.

The key advantage of the proposed model is that it does not require users to set various adjustment factors based on practical experience. It is very convenient for practitioners to assess the congestion impact and determine the optimal work zone schedule yielding the least delay (see Figures 4 and 5). Based on the predicted spatiotemporal speeds, a proper traffic mitigation and management plan may be prepared accordingly. In addition, it can be used to calculate the contractor penalty in terms of cost overruns, as well as an incentive reward schedule in case of early work competition.

The findings of this study point to areas of high potential for future research. First, although results in favor of the proposed hybrid machine-learning model were reported in this paper, calibrations on additional real-world work zone data, especially during peak periods, are needed to improve the prediction accuracy. Additional research is needed to investigate the impact of the relationship between approaching traffic volume and work zone capacity on the prediction accuracy. Finally, the model developed in this paper is only for predicting work zone delay on New Jersey freeways. Future studies will focus on enhancing the proposed model to deal with work zones on arterials with signalized intersections.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper is based on a research project sponsored by the New Jersey Department of Transportation (NJDOT). This support is gratefully acknowledged but implies no endorsement of the conclusions by NJDOT.

References

- [1] D. Schrank, B. Eisele, T. Lomax, and J. Bak, *2015 Urban Mobility Scorecard*, Texas Transportation Institute, Texas & University System, College Station, 2015.
- [2] Federal Highway Administration, "Fact and Statistics—Work Zone Delay," 2016, http://www.ops.fhwa.dot.gov/wz/resources/facts_stats/delay.htm.

- [3] C. L. Dudek and S. H. Richards, "Traffic Capacity through Urban Freeway Work Zones in Texas," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 869, pp. 14–18, Transportation Research Board of the National Academies, Washington, D.C., Wash, USA, 1982.
- [4] S. Chien and P. Schonfeld, "Optimal work zone lengths for four-lane highways," *Journal of Transportation Engineering*, vol. 127, no. 2, pp. 124–131, 2001.
- [5] J. Weng and Q. Meng, "Estimating capacity and traffic delay in work zones: an overview," *Transportation Research—Part C: Emerging Technologies*, vol. 35, pp. 34–45, 2013.
- [6] B. Du and S. I.-J. Chien, "Feasibility of shoulder use for highway work zone optimization," *Journal of Traffic and Transportation Engineering*, vol. 1, no. 4, pp. 235–246, 2014.
- [7] B. Du, S. Chien, J. Lee, L. Spasovic, and K. Mouskos, "Artificial neural network model for estimating temporal and spatial freeway work zone delay using probe-vehicle data," in *Transportation Research Record: Journal of the Transportation Research Board*, no. 2573, pp. 164–171, Transportation Research Board of the National Academies, Washington, D.C., Wash, USA, 2016.
- [8] B. Du, *An Artificial Neural Network Model for Predicting Freeway Work Zone Delays with Big Data [Ph.D. Thesis]*, New Jersey Institute of Technology, Newark, NJ, USA, 2017.
- [9] M. J. Lighthill and G. B. Whitham, "On kinematic waves. II. A theory of traffic flow on long crowded roads," *Proceedings of the Royal Society: London A: Mathematical, Physical and Engineering Sciences*, vol. 229, pp. 317–345, 1955.
- [10] P. I. Richards, "Shock waves on the highway," *Operations Research*, vol. 4, no. 1, pp. 42–51, 1956.
- [11] F. G. Habtemichael, M. Cetin, and K. A. Anuar, "Incident-induced delays on freeways: quantification method by grouping similar traffic patterns," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 2484, pp. 60–69, Transportation Research Board of the National Academies, Washington, D.C., Wash, USA, 2015.
- [12] S. I.-J. Chien, D. G. Goulias, S. Yahalom, and S. M. Chowdhury, "Simulation-based estimates of delays at freeway work zones," *Journal of Advanced Transportation*, vol. 36, no. 2, pp. 131–156, 2002.
- [13] Q. Meng and J. Weng, "Cellular automata model for work zone traffic," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 2188, pp. 131–139, Transportation Research Board of the National Academies, Washington, D.C., Wash, USA, 2010.
- [14] Y. Chung, H. Kim, and M. Park, "Quantifying non-recurrent traffic congestion caused by freeway work zones using archived work zone and ITS traffic data," *Transportmetrica*, vol. 8, no. 4, pp. 307–320, 2012.
- [15] P. Edara, C. Sun, and Z. Zhu, *Calibration of Work Zone Impact Analysis Software for Missouri*, Missouri Department of Transportation, 2013.
- [16] X. Jiang and H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *Journal of Transportation Engineering*, vol. 131, no. 10, pp. 771–779, 2005.
- [17] K. Kumar, M. Parida, and V. K. Katiyar, "Short term traffic flow prediction in heterogeneous condition using artificial neural network," *Transport*, vol. 30, no. 4, pp. 397–405, 2015.
- [18] L. Neubert, L. Santen, A. Schadschneider, and M. Schreckenberg, "Statistical Analysis of Freeway Traffic," in *Traffic and Granular Flow*, pp. 307–314, Springer, New York, NY, USA, 2000.
- [19] A. Karim and H. Adeli, "Radial basis function neural network for work zone capacity and queue estimation," *Journal of Transportation Engineering*, vol. 129, no. 5, pp. 494–503, 2003.
- [20] S. Ghosh-Dastidar and H. Adeli, "Neural network-wavelet microsimulation model for delay and queue length estimation at freeway work zones," *Journal of Transportation Engineering*, vol. 132, no. 4, pp. 331–341, 2006.
- [21] CoVal Systems, "What is OpenReach?" 2016, <http://www.covalsystems.com/latest/openreach/openreach.html>.
- [22] New Jersey Department of Transportation, "Straight Line Diagrams," 2015, <http://www.state.nj.us/transportation/refdata/sldiag>.
- [23] New Jersey Department of Transportation, "Congestion Management System," 2012, <http://www.state.nj.us/transportation/capital/stip1221>.
- [24] INRIX, "INRIX Home Page," 2016, <http://www.inrix.com>.
- [25] A. Maher, M. Jafari, E. Bossett, M. OConnell, and J. Buisson, "Plan4Safety," 2016, <http://cait.rutgers.edu/tsrc/plan4safety>.
- [26] J. Xiao and Y. Liu, "Traffic incident detection using multiple-kernel support vector machine," in *Transportation Research Record: Journal of the Transportation Research Board*, No. 2324, pp. 44–52, Transportation Research Board of the National Academies, Washington, D.C., Wash, USA, 2012.
- [27] R. Hajdin and H.-P. Lindenmann, "Algorithm for the planning of optimum highway work zones," *Journal of Infrastructure Systems*, vol. 13, no. 3, pp. 202–214, 2007.
- [28] J. Weng and Q. Meng, "Effects of environment, vehicle and driver characteristics on risky driving behavior at work zones," *Safety Science*, vol. 50, no. 4, pp. 1034–1042, 2012.
- [29] Bureau of Public Roads, *Traffic Assignment Manual*, U.S. Department of Commerce, Urban Planning Division, Washington, D.C., Wash, USA, 1964.
- [30] B. Du, J. Lee, S. Chien, B. Dimitrijevic, and K. Kim, "Short-term freeway work zone capacity estimation using support vector machine incorporated with probe-vehicle data," in *Presented at 94th Transportation Research Board Annual Meeting*, Washington, D.C., Wash, USA, 2015.
- [31] MathWorks, "Neural Network Toolbox in MATLAB," 2016, <https://www.mathworks.com/products/neural-network>.
- [32] New Jersey Department of Transportation, "Road User Cost Manual," 2015, <http://www.state.nj.us/transportation/eng/documents/RUCM>.
- [33] Highway Capacity Manual, "Transportation Research Board, Washington, D.C," 2010.



Hindawi

Submit your manuscripts at
<https://www.hindawi.com>

