TRR
JOURNAL OF THE TRANSPORTATION RESEARCH BOARD

*Research Article*

Transportation Research Record
1–9

DOI: 10.1177/0361198119837236
journals.sagepub.com/home/trr
SAGE

# Estimating Traffic Volume on Minor Roads at Rural Stop-Controlled Intersections using Deep Learning

**Mostafa H. Tawfeek[1] and Karim El-Basyouny[1]**

**Abstract**

Safety Performance Functions (SPFs) are regression models used to predict the expected number of collisions as a function of various traffic and geometric characteristics. One of the integral components in developing SPFs is the availability of accurate exposure factors, that is, annual average daily traffic (AADT). However, AADTs are not often available for minor roads at rural intersections. This study aims to develop a robust AADT estimation model using a deep neural network. A total of 1,350 rural four-legged, stop-controlled intersections from the Province of Alberta, Canada, were used to train the neural network. The results of the deep neural network model were compared with the traditional estimation method, which uses linear regression. The results indicated that the deep neural network model improved the estimation of minor roads' AADT by 35% when compared with the traditional method. Furthermore, SPFs developed using linear regression resulted in models with statistically insignificant AADTs on minor roads. Conversely, the SPF developed using the neural network provided a better fit to the data with both AADTs on minor and major roads being statistically significant variables. The findings indicated that the proposed model could enhance the predictive power of the SPF and therefore improve the decision-making process since SPFs are used in all parts of the safety management process.

The *Highway Safety Manual* (HSM) introduces, in part C, a procedure to use crash prediction models, also known as Safety Performance Functions (SPFs) (*1*). SPFs are equations (i.e., regression models) that mathematically link crash frequencies with traffic volume and other facility features (e.g., lane width, shoulder width, light, traffic control at intersections) based on a group of locations with similar characteristics. These functions are used throughout the safety management process to assess the safety of a transportation facility and to direct improvement funds toward transportation facilities with a high potential for crash reduction. Therefore, the predictive accuracy of such models or functions is crucial for project prioritization and decision-making.

The HSM proposes several SPFs for various highway facility types including roadway segments and intersections in both rural and urban areas (*1*). The HSM also proposes a calibration procedure in addition to the SPFs to accommodate the expected variation between jurisdictions' characteristics, with sufficient accuracy. Therefore, massive efforts were made by different state Departments of Transportation to calibrate HSM SPFs (*2–5*). However, many of these jurisdictions found that the calibration procedure was not accurate enough and inadequate for their conditions. Thus, several jurisdictions developed customized SPFs to fit their crash data better and enhance the prediction accuracy of the SPFs. The major recommendation from these jurisdictions was that jurisdiction-specific SPFs, especially the generalized SPFs, provided better results (*3–11*).

One of the main challenges when calibrating HSM SPFs or developing jurisdiction-specific SPFs, is the availability of sound crash, traffic volume, and supporting datasets. Traffic volume represented as annual average daily traffic (AADT) is an essential component of any SPF. Nevertheless, AADTs in some jurisdictions are seldom available, especially for minor roads at intersections. Consequently, minor road AADT estimation models were proposed to overcome this challenge. The traditional method of developing AADT estimation models used linear regression models (*12–14*). This method usually results in poor goodness-of-fit (GOF) for

[1]Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, Canada

**Corresponding Author:**
Address correspondence to Mostafa H. Tawfeek: mostafa.h.tawfeek@ualberta.ca

AADT estimates on minor roads at intersections, which affects the predictive power of SPFs.

The AADT on minor roads was missing for a significant portion (89%) of the rural unsignalized intersections in the Province of Alberta, Canada. In fact, the number of crashes cannot be predicted accurately without suitable models of AADT estimation for minor roads. Therefore, the aim of this study was to propose a minor roads AADT estimation model using a machine learning technique (i.e., deep neural networks). The AADT estimation models for minor roads were developed based on traffic, road geometry, and network data from the Province of Alberta, Canada. The proposed model, using neural networks, was compared with the traditional estimation method. Two SPFs for four-legged, stop-controlled intersections in rural areas were developed based on the estimated AADT from both methods (i.e., the traditional method and the machine learning method). Three years of crash data from the Province of Alberta were used for the development of these SPFs. The developed SPFs were then compared in relation to function structure and GOF.

## Literature Review

AADT is a crucial component (i.e., covariate or predictor) for the development of intersection-related SPFs as proposed by the HSM (*1*). However, minor road AADT in some jurisdictions is rarely available. Consequently, minor road AADT estimation models were proposed to overcome this challenge. Mohamad et al. (*12*) developed a traffic prediction model for AADT estimation for county roads in Indiana using a multiple linear regression technique. This study explored several variables including relevant demographic variables to be included in the model such as population, number of vehicles registered, income per capita, area type (i.e., rural or urban), mileage of highway, arterial and collector, and accessibility. Among all the explored variables, four were found to be significant variables in the model. Those four variables were county population, county arterial mileage, location, and accessibility. The model was found to provide a good fit for data with an *R-squared* (GOF measure) equal to 0.77. The average, minimum, and maximum differences between the observed and the predicted AADT was 17%, 1.6%, and 34%, respectively.

Xie et al. (*13*) estimated minor road AADTs to calibrate the HSM SPFs for Oregon State highways. Several variables were evaluated, and 12 key variables had the potential to influence the estimation process. Some of these variables were continuous, such as county population, the population of the nearest city, average income per capita, and distance to the nearest freeway. These variables were converted to a logarithmic scale with a base of 10. The remaining eight variables were binary variables in the form of questions, which have the value of 1 if yes, of 0 otherwise. The questions included whether the minor road was a minor arterial, major collector, had a right turn, had a centerline, or had a striped edge line. In addition, these questions included some information about the intersection (i.e., whether the intersection was within the city limit), the adjacent land uses (i.e., the intersection's adjacent land was developed), and the major road (i.e., a right turn was present on it). The developed models had nine variables, excluding the county population, the population of the nearest city, and the average income. The *R-squared* value was around 0.6, and the variables were significant at 90% confidence level.

A recent study in Oregon State used regression models to estimate the entering minor AADT (*15*). The presence of exclusive right- or left-turn lanes on both roads (i.e., major and minor roads), the average number of through lanes on both roads, speed limit, the presence of two-way left-turn lanes, and road width were explored as variables with a potential correlation to the minor road AADT. In this study, the parallel facilities (i.e., parallel minor roads) with similar characteristics were considered; however, the volumes on these facilities were not always available. Consequently, a model with a parallel facility AADT and another model without this AADT were proposed for estimating the minor road entering AADT. The models' variables were major road AADT, parallel AADT (for the first model only), average number of through lanes on major roads, and the functional classes of the major and minor roads. The *R-squared* values for the first model (i.e., with parallel AADT) and the second model (i.e., without the parallel AADT) were 0.63 and 0.60, respectively. In addition, all the parameters were found to be statistically significant at either the 90% or 95% confidence level.

Recently, neural network models and a statistical model were developed to estimate the link-level AADT in urban areas based on land-use characteristics (*16*). Several independent variables were used including land-use characteristics and network characteristics (e.g., number of lanes and median presence) and the dependent variables (i.e., AADT of the links) were obtained from permanent count stations. A total of 285 urban road links was used for the models' development and only 30 were used for the models' validation. The results of the neural network models were compared with the statistical model and to the traditional four-step model outputs. The neural networks showed the lowest error among the compared methods. It is worth mentioning that the primary focus in this paper was on the variables related to land-use characteristics in an urban context only. Furthermore, the models were used mainly to estimate the AADT on links only (*16*).

Certain studies overcame the unavailability of the AADT on minor roads by using different modeling approaches to develop the SPF. For example, one approach uses the number of intersections on a road segment (i.e., intersection density) as an explanatory variable in the SPF. Another approach, which appears in the HSM, is to develop a separate SPF for intersections (*17*, *18*). By comparing the results of both approaches, there was no difference in the quality of the obtained estimates. It is worth mentioning that Mountain et al. (*17*) defined traffic flow as the two-way annual traffic flow of a major road measured in million vehicles per year. In other words, this study avoided the inclusion of the traffic flow on the minor roads in the development of the SPF.

In summary, the lack of minor road AADTs at intersections was identified as a common challenge during the development of SPFs for intersections by several previous studies. To overcome this pitfall, several efforts were made to estimate or predict AADTs for minor roads using linear regression. Some studies used the estimated AADTs to calibrate HSM SPFs or to develop jurisdiction-specific SPFs (*12*, *13*, *15*). These models include several variables such as the class of the minor road, the presence of right turns on minor or major roads, marking of the minor roads (i.e., centerline or striped edge lines), proximity to the city limit, and adjacent land development (*13*). In addition, major road AADT, average number of through lanes on major roads, functional classification of major and minor roads, and AADT of a parallel facility, if available, were used to develop minor roads AADT estimation models (*15*).

Despite these efforts, the estimation models had, on average, an *R-squared* value of approximately 0.6, which indicates a less than ideal yet adequate fit. Therefore, this paper aims at improving the estimation procedure of AADTs on minor roads, to develop SPFs for unsignalized intersections, by adopting a novel machine learning technique. To explore the effect of such estimation procedures on the SPFs, models were developed using the estimated AADT using (i) linear regression and (ii) deep neural network (DNN).

## Methodology

### *Development of Minor Roads AADT Estimation Models*

Several variables were explored to develop estimation models for AADT on minor roads. These variables were obtained from Alberta Transportation for the development of SPFs. The explored variables included road geometry (i.e., number of lanes, the presence of median separation, and the presence of left or right turns), traffic (i.e., AADTs for major roads), and network (i.e., the closeness to urban centers, the type of the closest urban center, and the road functional class). It is worth mentioning that the road functional classification used in the Province of Alberta divides the roadways into four classes. These four classes include national highway system facilities as service class 1, arterials facilities as service class 2, collectors facilities as service class 3, and local facilities as service class 4 (*19*).

A total of 1,350 four-legged, stop-controlled rural intersections with available traffic volume data on both minor and major roads were used in the development of the minor roads AADT estimation models, and in the calculation of the GOF measure for each model. These estimation models calculate the total entering AADT at the minor roads. Only the major road AADT and the distance to the nearest urban area were continuous variables; all the other geometrical variables were binary. The other geometrical binary variables took the form of questions, which had a value of 1 if the answer was yes and 0 otherwise. These questions included whether or not:

- The major road is service class 1.
- The major road is service class 2.
- The major road is service class 3.
- The major road is service class 4.
- The total number of lanes of the major roads is more than two (i.e., multilane road).
- The major road is divided.
- The intersection is lit.
- The major road has right- or left-turn lanes.
- The closest urban area to the intersection is town or city.

Table 1 summarizes the variables used to develop the AADT estimation models, and their descriptive statistics. It is noteworthy that while the maximum value of the AADT on major roads is far from the mean value, this maximum AADT is consistent with other rural intersections that have the same characteristics, such as service class, lit, and so forth. Moreover, the maximum AADT is relatively high because the intersection that has this maximum AADT is located on a city boundary.

*Linear Regression Model.* According to the literature, the *Log10* transformation is usually used for the continuous variables in the estimation models to minimize the unbalanced variance (*12*, *13*, *15*). Therefore, when developing the AADT estimation model using linear regression, all the continuous variables including minor and major road AADTs and distance to the nearest urban area were transformed. Linear regression analysis was conducted using the data of 1,350 rural unsignalized intersections to

**Table 1.** Descriptive Statistics of Variables Explored for Models' Development

| | Rural 4-leg stop-controlled intersections | | | |
|---|---|---|---|---|
| Variable name | Min | Max | Mean | St. Dev. |
| AADT major road | 30 | 60,900 | 3,381.16 | 4,724.20 |
| AADT minor road | 0 | 6,330 | 524.35 | 700.25 |
| Service class 1 | 0 | 1 | 0.17 | 0.37 |
| Service class 2 | 0 | 1 | 0.40 | 0.49 |
| Service class 3 | 0 | 1 | 0.35 | 0.48 |
| Service class 4 | 0 | 1 | 0.08 | 0.28 |
| Total no. of lanes is more than 2 | 0 | 1 | 0.18 | 0.38 |
| Major road is divided | 0 | 1 | 0.14 | 0.35 |
| Intersection is lit | 0 | 1 | 0.25 | 0.43 |
| Right turn lane on major road (RTL) | 0 | 1 | 0.53 | 0.50 |
| Left turn lane on major road (LTL) | 0 | 1 | 0.21 | 0.41 |
| Nearest urban area is city or town | 0 | 1 | 0.64 | 0.48 |
| Distance to nearest urban area (km) | 0.321 | 283.65 | 16.64 | 21.63 |

*Note*: AADT = annual average daily traffic; St. Dev. = standard deviation.

develop a model for minor road AADT estimation. The minor road AADT as a dependent variable was assumed to be a function of the variables mentioned in Table 1.

*DNN Model.* Artificial neural networks have been widely used as a modeling and pattern recognition technique in various transportation applications. For instance, neural networks were used to study the route choice behavior (*20*), model car-following behavior (*21*), traffic congestion prediction (*22*), and traffic safety analysis (*23*, *24*). The concept of a neural network modeling approach was initiated to mimic how the human brain works. Similar to the neuron in the human brain, which has a soma and an axon, each node (i.e., hidden unit) in the neural network has an input and an output (*25*). These inputs and outputs connect all the nodes to represent the layered structured neural network, exactly as in the nervous system where the neurons are connected by somas and axons. As shown in Figure 1, a DNN consists of an input layer, hidden layers, and an output layer.

Using the Deep Learning Toolbox in MATLAB (*26*), a multilayer feed-forward neural network with backpropagation was used to estimate the AADTs of minor roads. The Bayesian regularization backpropagation was used as the network training function. The training function updates the weights and the bias of the network according to Levenberg-Marquardt optimization. The main advantage of the Bayesian regularization backpropagation is that it overcomes neural networks overfitting by producing well-generalized networks. This training function estimated the regularization parameters while training the network to improve the generalized network performance (*27*, *28*). This type of neural network was used for transportation applications such as route choice modeling (*29*). For more details about Bayesian regularization backpropagation training function, readers are directed to Mackay (*27*), Hagan et al. (*28*), and Foresee and Hagan (*30*). Also, the maximum number of epochs used to train the networks was set to 300 to reduce the training time of the networks. The performance function for the training of the models was based on the mean square error, and the performance goal was zero.

The input layer of the proposed network consists of 12 hidden units representing 12 variables, which were mentioned in Table 1. The output layer of the network included only one hidden layer that was the estimated AADT of the minor road. For network training purposes, the data available for this study (i.e., 1,350 intersections) were split randomly to 90% as a training set, and 10% as a test set. To reach a reasonable training performance and to have a better generalized neural network, several trials were conducted with different network structures (i.e., various number of layers and number of hidden units). The best performance was achieved by a network with four hidden layers with 20, 19, 19, and 16 hidden units each. This network was then prepared to estimate minor roads' AADT for the development of SPFs.

## *Development of SPFs*

The Negative Binomial regression framework, as recommended by the HSM (*1*), was used to develop the SPFs for rural four-legged, stop-controlled intersections with two-way, two-lane major roads. To develop these SPFs, the crash data for 12,949 intersections were extracted for three years along with the traffic volumes and other variables mentioned in Table 1. Two functions were developed including SPFs using AADTs estimated from traditional models (i.e., linear regression model) and supervised machine learning (i.e., DNN). Statistical Analysis System (SAS) version 9.4 (*31*) was used to

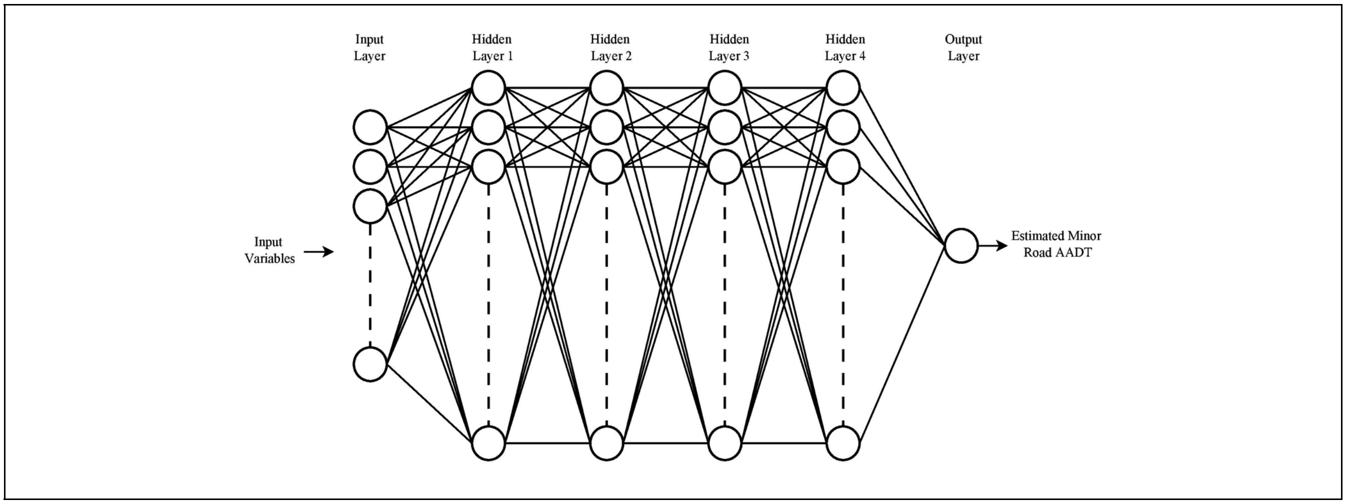


**Figure 1.** The proposed structure of the DNN used for minor road AADT estimation.

develop the Negative Binomial models for this study. All the variables, which were mentioned in Table 1, were considered while developing the SPFs for both cases. The aim was to explore the effect of different methods of minor AADTs estimation on the structure of the developed SPFs and to facilitate the comparison between the developed SPFs.

To assess the appropriateness of the SPFs, parameter significance, as well as the GOF of the developed models, was examined. For parameter significance, the level of significance was 0.01 for a 99% confidence level. For the GOF, the Bayesian information criterion (BIC) and the Akaike information criterion (AIC) are used to assess the developed SPFs (*4*, *10*, *32*). The model, which has smaller BIC and AIC values, is a better model (*4*, *10*, *32*).

## Results and Discussion

### *Minor Roads AADT Models*

*Linear Regression Model.* The first model developed in this study was based on linear regression for rural four-legged stop-controlled intersections. The estimates and the variables significance are shown in Table 2, and the GOF of the model (i.e., *R-squared*) is shown in Figure 2. Figure 2 depicts the relation between the output values (Y) of the model (i.e., predicted AADT of the minor roads) and the target values (T) (i.e., the actual AADT of the minor roads). Moreover, this figure graphically represents the fit of the proposed model to the given data. For a good fit, the data points of the plot will be closer to the 45-degree line where the output and target values are equal, otherwise it is a poor fit. It is worth mentioning that the *R-squared* for this rural model was estimated at 0.66, which is consistent with the *R-squared* values found in previous studies (*13*, *15*). As shown in Table 2, the variables included in the model were all significant at a 99% confidence level. All other variables that were explored during the development process of the models were insignificant and were consequently excluded from the final estimation model. The model mathematical functional form for the rural four-legged, stop-controlled intersections is summarized below.

**Table 2.** Linear Regression Model Parameters for Minor Roads AADT Estimation

| | Rural 4-leg stop-controlled | |
|---|---|---|
| Variable name | Estimates | *p-value* |
| Intercept | 0.867 | <0.001 |
| Service class 1 | –0.338 | <0.001 |
| Service class 2 | –0.151 | <0.001 |
| Light | 0.482 | <0.001 |
| Right-turn lane on major road | 0.123 | <0.001 |
| Major road AADT | 0.509 | <0.001 |

*Note*: AADT = annual average daily traffic.

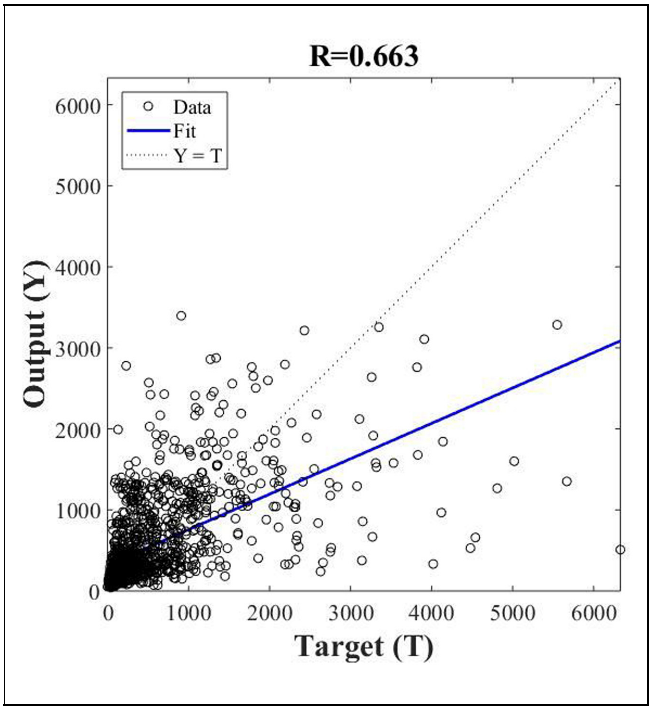


**Figure 2.** Linear regression model GOF results.

$$\log AADT_{Min} = 0.867 - 0.338 \times S.C.1 - 0.151 \times S.C.2 + 0.482 \times Lit + 0.123 \times RTL + 0.509 \times \log AADT_{Maj} \quad (1)$$

where

$AADT_{Maj}$ = average annual daily traffic on the major road (vehicles per day (vpd)).

$AADT_{Min}$ = average annual daily traffic on the minor road (vpd).

$S.C.1$ = major road is service class 1; 1 if true, 0 otherwise.

$S.C.2$ = the major road is service class 2; 1 if true, 0 otherwise.

$Lit$ = major road at the intersection is lit; 1 if true, 0 otherwise.

$RTL$ = one direction or both directions of the major road at the intersection has a right turning lane(s); 1 if true, 0 otherwise.

*DNN Model.* Different neural network sizes were tested to obtain the best performance for AADT on minor roads estimation. A four hidden layers neural network with 20, 19, 19, and 16 hidden units each was trained for the estimation of the AADT. Figure 3 showed the results of the model GOF represented as *R-squared* values. These values indicate the GOF for the training set, test set, and the entire dataset (i.e., including both the training and test sets). In addition, Figure 3 exhibited the relation between the target (i.e., actual) and the output (i.e., estimated by the model) values of minor roads' AADT using the neural networks for the different datasets. As shown in the figure, the model estimates match the actual minor roads' AADT since a substantial portion of the target values equal or were close to the output values. This was reflected by the *R-squared* values, which were recorded as 0.94, 0.67, and 0.89 for the training set, the test set, and the entire dataset, respectively. These results indicated that the minor roads AADT estimation using the neural network model outperformed the linear regression model. The improvement in the AADT estimation was around 35% when comparing the overall *R-squared* value of the neural network with the *R-squared* value of the linear regression. It is worth mentioning that the GOF of both models was compared based on the entire dataset because the traditional statistical techniques (e.g., linear regression) do not often split the data into training and test sets, especially when the linear regression is used for AADT estimation as discussed earlier (*12–14*).

## *Safety Performance Functions*

The HSM defines intersection crashes as those that occur in the physical area and immediate vicinity of an intersection, and intersection-related crashes as the ones occurring just before or after the intersection based on the characteristics of the crash (*1*). An SPF for an intersection predicts the average crash frequency that would occur per year given specific intersection characteristics. According to the HSM, traffic volumes at an intersection for both major and minor roads should be included in the SPF (*1*). Despite the tangible improvement that the DNN model had when compared with linear regression, both models were also verified by developing SPFs based on the AADT of the minor roads, which were estimated from each method (i.e., DNN and the linear regression). The primary goal of developing these SPFs is to investigate to what extent minor roads AADT estimation model results could affect the structure of the proposed SPFs. Therefore, several variables were considered during the development of the SPF, including traffic volumes, turn lanes on the major roads, major road service class, and other variables that were mentioned in Table 1. However, only the statistically significant variables were included in the final SPFs presented in Tables 3 and 4.

Two SPFs, shown in Tables 3 and 4, were developed based on the used minor roads AADT estimation models (i.e., the linear regression model and the neural network model). The results of the SPF developed using the volumes estimation of the linear regression model are shown in Table 3. From this table, six variables were significant at a 99% confidence level. These six variables included the AADT on the major road, whether the

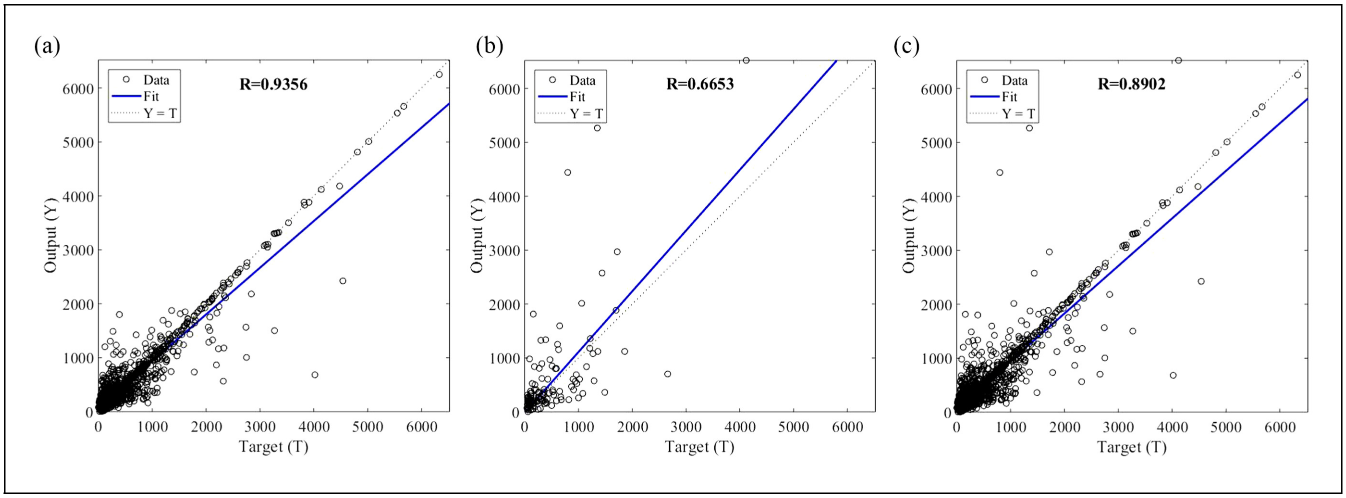


**Figure 3.** Neural network model GOF results for (*a*) the training set, (*b*) the test set, and (*c*) the entire data (i.e., including the training and test sets).

**Table 3.** The SPF Developed Based on the Linear Regression Estimations of Minor Roads' AADT

| Parameter | Estimate | *p*-value |
|---|---|---|
| Intercept | –8.783 | $<0.001$ |
| $AADT_{maj}$ | 0.871 | $<0.001$ |
| Service class 3 | 0.159 | $<0.001$ |
| RTL | 1.037 | $<0.001$ |
| LTL | 0.902 | $<0.001$ |
| CTC | 0.142 | 0.0028 |
| Udist | –0.128 | $<0.001$ |
| Dispersion | 1.218 | |
| AIC (smaller is better) | 14,521.717 | |
| BIC (smaller is better) | 14,581.467 | |

*Note*: SPF = Safety Performance Functions; AADT = annual average daily traffic; RTL = right turn lane on major road; LTL = left turn lane on major road; CTC = closest urban area to the intersection is either a town or a city; Udist = distance to the closest urban area; AIC = Akaike information criterion; BIC = Bayesian information criterion.

**Table 4.** The SPF Developed Based on the Neural Network Estimations of Minor Roads' AADT

| Parameter | Estimate | *p-value* |
|---|---|---|
| Intercept | –9.008 | $<0.001$ |
| $AADT_{maj}$ | 0.702 | $<0.001$ |
| $AADT_{min}$ | 0.207 | $<0.001$ |
| Service class 1 | 0.335 | 0.0065 |
| Service class 2 | 0.351 | $<0.001$ |
| Service class 3 | 0.363 | $<0.001$ |
| Lit | 0.436 | $<0.001$ |
| RTL | 0.909 | $<0.001$ |
| LTL | 0.622 | $<0.001$ |
| CTC | 0.180 | $<0.001$ |
| Udist | –0.140 | $<0.001$ |
| Dispersion | 1.075 | |
| AIC (smaller is better) | 14,417.400 | |
| BIC (smaller is better) | 14,507.025 | |

*Note*: SPF = Safety Performance Functions; AADT = annual average daily traffic; RTL = right turn lane on major road; LTL = left turn lane on major road; CTC = closest urban area to the intersection is either a town or a city; Udist = distance to the closest urban area; AIC = Akaike information criterion; BIC = Bayesian information criterion.

major road is service class 3, whether the major road has a right (RTL) or left turn lane (LTL), whether the closest urban area to the intersection is either a town or a city (CTC), and the distance to this closest urban area (Udist). As indicated in Table 3, the traffic volume on the minor road (i.e., AADT) was not included since it was statistically insignificant. It is worth mentioning that other variables were also eliminated because they were not significant, such as whether the major road is a service class 1 or 2, and whether the intersection was lit or not. Table 3 also included the GOF measures used to assess the developed SPF such as AIC and BIC.

Table 4 shows the results of the SPF developed using the minor roads AADT, estimated using the neural network model. This SPF, as shown in the table, included 10 variables that significantly affect crash occurrence at a 99% confidence level. Among these 10 variables, and in contrast to the first SPF, the traffic volume variables on both the major and minor roads, represented as AADT, were found to be statistically significant. This reflects that the SPF developed based on more accurate AADT estimation had a more reasonable structure by including the volumes of both roads as exposure measures in the SPF. In addition to the traffic volumes, and as shown in Table 4, the SPF contained other variables such as major road service class, whether the intersection was lit or not, presense of an RTL or LTL on the major road, whether the closest urban area to the intersection is either a town or a city, and the distance to this closest urban area.

In relation to GOF measures, the SPF shown in Table 4 had smaller values of the AIC and BIC when compared with the SPF in Table 3. These results indicated that the SPF, which was developed based on the neural network estimates of minor roads' AADT, had a better fit to the data when compared with the SPF developed based on the linear regression estimates of minor roads' AADT.

## Summary and Conclusions

This paper applied a machine learning technique (i.e., DNN) to estimate the AADT on minor roads (i.e., output variable) at rural, stop-controlled intersections. The training of the DNN considered several road geometry, traffic, and network characteristics as input variables. These input variables included the number of lanes, the presence of median separation, the presence of left or right turns as road geometric variables, AADTs of major roads as a traffic variable, the closeness to urban centers, the type of the closest urban center, and the road functional class as network variables.

The DNN used 1,350 rural stop-controlled intersections in the Province of Alberta, Canada, as training examples. Based on the estimation accuracy, a DNN of four hidden layers with 20, 19, 19, and 16 hidden units for each layer, respectively, was chosen to estimate the AADT. The *R-squared* values of the neural network were 0.94, 0.67, and 0.89 for the training set, the test set, and the entire dataset, respectively. These values indicated a strong fit between the estimated AADT values from the DNN model and the AADT values collected from the field.

To compare the performance of the proposed DNN model with the traditional estimation model, a linear regression model was developed using the same dataset to estimate the AADT on minor roads. The *R-squared* value of this model was 0.66. This result indicated that the DNN model outperformed the traditional method when it comes to estimating AADT on minor roads. The improvement was estimated to be around 35% when comparing the overall *R-squared* values of the DNN and the linear regression model.

Moreover, an SPF was developed using the estimated AADT on minor roads from each model (i.e., DNN and linear regression models). AADT on minor roads was estimated for 12,949 rural four-legged, stop-controlled intersections using the two models. Two SPFs were then developed using the variables mentioned in Table 1, which included the estimated AADT on minor roads. The SPF, which was developed using the linear regression model estimations, did not include AADT on minor roads as a significant variable in the SPF. On the other hand, the AADT on minor roads had a significant effect on crash occurrence according to the SPF that was developed using the DNN model. Furthermore, the GOF measures indicated that the SPF, which was developed using the DNN model, had a better fit to the crash data. Therefore, the performance of estimating AADT on minor roads using the DNN model for developing SPFs surpassed the traditional method of AADT estimation.

This paper addresses a common challenge in the development of intersection-related SPFs, which is the lack of available AADTs on minor roads at rural intersections. Unlike previous studies, which used neural networks to estimate the AADT on low-volume roads based on temporary traffic counts, such as 48-hour traffic counts (*33*, *34*), this study estimated the AADT on intersections' minor roads where no traffic counts were conducted or recorded previously. Moreover, the proposed model was trained using many training examples (1,350 intersections) and a wide range of input variables covering traffic, network, and road geometry characteristics. The availability of these datasets contributed to the high estimation accuracy of the DNN. However, the amount of data used in this study might be a challenge for some jurisdictions. Future research should investigate the minimum number of intersections and minimum number of input variables that would produce a sufficient estimation accuracy. Furthermore, the influence of the input variables on the predictions of the DNN should be explored because the DNN does not provide a direct relation between the input variables and the estimation accuracy.

### Author Contributions

The authors confirm contributing to the paper as follows: study conception and design: MHT and KE; methodology and data collection and analysis: MHT; interpretation of results: MHT and KE; draft manuscript preparation: MHT and KE. Both authors reviewed the results and approved the final version of the manuscript.

### References

1. American Association of State Highway Transportation Officials. *Highway Safety Manual with 2014 Supplement*, 1st ed. AASHTO, Washington, D.C., 2010.
2. Troyer, D., K. Bradbury, and C. Juliano. Strength of the Variable. *Transportation Research Record: Journal of the Transportation Research Board*, 2015. 2515: 86–93.
3. Brimley, B. K., M. Saito, and G. G. Schultz. Calibration of *Highway Safety Manual* Safety Performance Function. *Transportation Research Record: Journal of the Transportation Research Board*, 2012. 2279: 82–89.
4. Mehta, G., and Y. Lou. Calibration and Development of Safety Performance Functions for Alabama. *Transportation Research Record: Journal of the Transportation Research Board*, 2013. 2398: 75–82.

5. Qin, X., C. Zhi, and K. Vachal. Calibration of Highway Safety Manual Predictive Methods for Rural Local Roads. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014, pp. 1–16.
6. Srinivasan, R., and D. Carter. *Development of Safety Performance Functions for North Carolina*. Publication FHWA/NC/2010-09. FHWA, U.S. Department of Transportation, 2011.
7. Srinivasan, S., P. Haas, N. S. Dhakar, R. Hormel, D. Torbic, and H. Douglas. *Development and Calibration of Highway Safety Manual Equations for Florida Conditions*. Final Report (FDOT BDK77 977-06 12). Vol. 6, 2011.
8. Tegge, R. A., J.-H. Jo, and Y. Ouyang. *Development and Application of Safety Performance Functions for Illinois*. Publication FHWA-ICT-10-066. FHWA, U.S. Department of Transportation, 2010.
9. Aziz, S. R., and S. Dissanayake. *Calibration of Highway Safety Manual Methodology for Rural Multilane Divided Highways in Kansas*. Publication K-TRAN: KSU-14-3. Kansas Department of Transportation. 2015.
10. Young, J., and P. Y. Park. Benefits of Small Municipalities using Jurisdiction-specific Safety Performance Functions rather than the Highway Safety Manual's Calibrated or Uncalibrated Safety Performance Functions. *Canadian Journal of Civil Engineering*, Vol. 40, No. 6, 2013, pp. 517–527.
11. Tawfeek, M. H., and K. El-Basyouny. *Safety Prediction Modelling and Deployment of Safety Analyst Software*. Alberta Transportation, 2016.
12. Mohamad, D., K. Sinha, T. Kuczek, and C. Scholer. Annual Average Daily Traffic Prediction Model for County Roads. *Transportation Research Record: Journal of the Transportation Research Board*, 1998. 1617: 69–77.
13. Xie, F., K. Gladhill, K. K. Dixon, and C. M. Monsere. Calibration of Highway Safety Manual Predictive Models for Oregon State Highways. *Transportation Research Record: Journal of the Transportation Research Board*, 2011. 2241: 19–28.
14. Dixon, K., C. Monsere, F. Xie, and K. Gladhill. *Calibrating the Highway Safety Maual Predictive Methods for Oregon Highways*. Publication FHWA-OR-RD-12-07. Kansas Department of Transportation. 2012.
15. Karen, D., C. Monsere, R. Avelar, J. Barnett, P. Escobar, and Y. W. Sirisha Kothuri. *Improved Safety Performance Functions for Signalized Intersections Report*. Publication FHWA-OR-RD-16-03. Portland State University, Civil and Environmental Engineering Faculty publications and presentations. Portland, OR, 2015.
16. Duddu, V. R., and S. S. Pulugurtha. Principle of Demographic Gravitation to Estimate Annual Average Daily Traffic: Comparison of Statistical and Neural Network Models. *Journal of Transportation Engineering*, Vol. 139, No. 6, 2013, pp. 585–595.
17. Mountain, L., B. Fawaz, and D. Jarrett. Accident Prediction Models for Roads with Minor Junctions. *Accident Analysis and Prevention*, Vol. 28, No. 6, 1996, pp. 695–707.
18. Ambros, J., J. Sedoník, and Z. Křivánková. How to Simplify Road Network Safety Screening? *Advances in Transportation Studies*, Vol. 44, 2018, pp. 151–158.
19. Stantec Consulting. *Provincial Highway Service Classification*. Alberta Transportation. 2007.
20. Yang, H., R. Kitamura, P. R. Jovanis, K. M. Vaughn, and M. A. Abdel-Aty. Exploration of Route Choice Behavior with Advanced Traveler Information using Neural Network Concepts. *Transportation*, Vol. 20, 1993, pp. 199–223.
21. Jia, H., Z. Juan, and A. Ni. Develop a Car-following Model Using Data Collected by "Five-wheel System." *Proceedings of IEEE Conference on Intelligent Transportation Systems, ITSC*, Vol. 1, 2003, pp. 346–351.
22. Dougherty, M. S., H. R. Kirby, and R. D. Boyle. The Use of Neural Networks to Recognise and Predict Traffic Congestion. *Traffic Engineering & Control*, Vol. 34, No. 6, 1993, pp. 311–314.
23. Chang, L. Y. Analysis of Freeway Accident Frequencies: Negative Binomial Regression versus Artificial Neural Network. *Safety Science*, Vol. 43, No. 8, 2005, pp. 541–557.
24. Abdelwahab, H., and M. Abdel-Aty. Artificial Neural Networks and Logit Models for Traffic Safety Analysis of Toll Plazas. *Transportation Research Record: Journal of the Transportation Research Board*, 2002. 1784: 115–125.
25. McCulloch, W. S., and W. Pitts. A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, Vol. 5, No. 4, 1943, pp. 115–133.
26. The Mathworks Inc. *MATLAB - MathWorks*. http://www.mathworks.com/products/matlab/.
27. MacKay, D. J. C. Bayesian Interpolation. *Neural Computation*, Vol. 4, No. 3, 1992, pp. 415–447.
28. Hagan, M. T., H. B. Demuth, and M. H. Beale. *Neural Network Design*. Boston, Pws Pub, 1995.
29. Sun, B., and B. B. Park. Route Choice Modeling with Support Vector Machine. *Transportation Research Procedia*, Vol. 25, 2017, pp. 1806–1814.
30. Dan Foresee, F., and M. T. Hagan. Gauss-Newton Approximation to Bayesian Learning. *IEEE International Conference on Neural Networks - Conference Proceedings*, Vol. 3, 1997, pp. 1930–1935.
31. SAS Institute. *Welcome to SAS 9.4 Help and Documentation*. SAS Institute, Cary, NC, 2014.
32. Kaaf, K. A., and M. Abdel-Aty Transferability and Calibration of Highway Safety Manual Performance Functions and Development of New Models for Urban Four-Lane Divided Roads in Riyadh. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015, pp. 1–18.
33. Sharma, S., P. Lingras, F. Xu, and P. Kilburn. Application of Neural Networks to Estimate AADT on Low-volume Roads. *Journal of Transportation Engineering-ASCE*, Vol. 127, No. 5, 2001, pp. 426–432.
34. Jin, L., C. Xu, and J. D. Fricker. Comparison of Annual Average Daily Traffic Estimates: Traditional Factor, Statistical, Artificial Neural Network, and Fuzzy Basis Neural Network Approach. *Network*, No. 765, 2008, pp. 1–19.


*The Standing Committee on Artificial Intelligence and Advanced Computing Applications (ABJ70) peer-reviewed this paper (19-02772).*