

Received November 18, 2020, accepted November 26, 2020, date of publication December 4, 2020, date of current version December 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3042612

Developing Data-Driven Approaches for Traffic Density Estimation Using Connected Vehicle Data

MOHAMMAD A. ALJAMAL^{1,3,4}, (Member, IEEE), MOHAMED FARAG^{1,2}, AND HESHAM A. RAKHA^{1,3}, (Fellow, IEEE)

¹Center for Sustainable Mobility, Virginia Tech Transportation Institute, Virginia Tech, Blacksburg, VA 24061, USA

²College of Computing and Information Technology, Arab Academy for Science, Technology, and Maritime Transport, Alexandria 1029, Egypt

³The Charles E. Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA 24061, USA

⁴Precision Systems, Inc, Washington, DC 20003, USA

Corresponding author: Hesham A. Rakha (hrakha@vt.edu)

This work was supported by the University Mobility and Equity Center (UMEC).

ABSTRACT This paper introduces novel approaches for the estimation of the traffic stream density. First, an artificial neural network (ANN) data-driven approach is developed to estimate the level of market penetration (LMP) of connected vehicles at two fixed locations. Then, the estimated values are used as inputs to a Kalman filter (KF) approach to estimate the vehicle count between these two locations. Second, three data-driven approaches are developed to directly estimate the vehicle count using only connected vehicle data, an ANN, a k-nearest neighbor (k-NN), and a random forest (RF). A congested signalized roadway in downtown Blacksburg, Virginia, is used to test and compare the performance of the estimation approaches. Results demonstrate that the ANN approach produces reasonable errors in estimating the LMPs; however, integrating the ANN with the KF results in larger errors than the errors produced from using the KF with a predefined fixed average value obtained from historical data. The results also demonstrate that the data-driven approaches provide accurate vehicle count estimates, with the ANN being the most accurate of the three approaches. Lastly, the paper compares the three developed data-driven approaches with model-driven approaches (i.e., KF), showing that the ANN outperforms all other approaches. However, taking into consideration that the difference is not large, the computational time needed to train the ANN, the large amount of data needed, and the uncertainty in the performance when new traffic behaviors are observed (e.g., incidents), the use of the KF approach is recommended in the estimation of traffic stream density due to its simplicity and applicability in the field.

INDEX TERMS Connected vehicles, real-time estimation, machine learning, traffic stream density.

I. INTRODUCTION

People wasted around 166 billion hours in traffic congestion in 2017, including around 3.8 billion gallons of fuel [1]. Traffic engineers and researchers are working to provide solutions for the traffic congestion problem. One efficient solution is to deploy Intelligent Transportation System (ITS) applications with the aim of increasing the capacity of the existing traffic infrastructure [2]. One ITS application is the use of connected vehicle (CV) technology, which can allow information exchange between two CVs (V2V communication) and also between any CV and the traffic infrastructure (V2I communication). In the case of traffic congestion, traffic

infrastructures such as the traffic signal controller can send early messages to the surrounding CVs to find alternative routes, leading to a reduction in trip travel times.

Traffic congestion can be represented by the macroscopic traffic stream density (the number of vehicles that traverse a specific roadway segment divided by the length of that segment). Traffic density is considered a spatial rather than a temporal measurement. Consequently, the temporal traffic occupancy measurements, obtained from loop detectors, cannot be used to estimate the traffic density for the entire link unless multiple loop detectors are installed, which results in high costs. A more efficient way to estimate the traffic density is to exploit CV technology with the ability to share real-time information, such as the vehicle's location and speed, anywhere inside the link [3]. Consequently, this paper

The associate editor coordinating the review of this manuscript and approving it for publication was Liang Hu¹.

aims to develop three data-driven approaches for the estimation of the traffic stream density on signalized approaches. The three data-driven approaches include an artificial neural network (ANN), a k-nearest neighbor approach, and a random forest (RF) model. The three estimation approaches are developed using CV data only and compared to real-time model-driven filtering techniques to identify the merits of each approach.

II. RELATED WORK

To estimate the number of vehicles on a road segment, researchers have developed different estimation approaches such as model-driven (filtering techniques) and data-driven (machine learning). In addition, different data sources were used to implement the proposed estimation approaches, such as the data from fixed sensors (e.g., loop detectors), data from two different detection sources (fusion data), and CV data. In this section, we will present and discuss previous studies that developed model-driven and data-driven approaches to estimate the traffic stream density.

A. MODEL-DRIVEN ESTIMATION APPROACHES

For the use of fixed sensors, the input-output approach has been widely used to develop model-driven approaches. One study developed a Kalman filter (KF) approach to estimate the vehicle counts on a signalized link using at least three loop detectors (two at the boundaries of the tested link and the third one in the middle of the link) [4]. Ghosh and Knapp [5] developed a KF approach to estimate the total number of vehicles using data from four loop detectors. Bhourri *et al.* [6] also developed a KF approach, using six loop detectors, to provide accurate estimates for the vehicle counts in an on-ramp segment. In summary, the aforementioned studies require installing multiple fixed sensors to provide accurate estimates. However, the cost of implementing these approaches in the field is high. Moreover, it was found that fixed sensors always produce some noise in their data [7]. Thus, there is an urgent need to use additional data sources to reduce the noise.

Fusion data has been given more attention following the introduction of advanced technologies such as CVs. Recently, researchers have started using fixed sensors together with CV data for seeking better estimation accuracy. A study attempted to provide accurate estimates of traffic density using mobile sensors and loop detector data [8], showing that the estimation accuracy using fusion data outperformed the loop detector data. A recent study utilized CVs and cameras to estimate traffic density in a 500 m highway segment. The model developments were based on the assumption that the average speed of CVs is approximately equal the average speed of traditional vehicles [9]. In that study, a KF model was developed under the consideration of having a linear parameter-varying system with known parameters. The state equation was based on the traffic flow continuity equation, while the measurement equation was based on the average speed of CVs. Wright and Horowitz [10] developed a particle

filter (PF) using fusion loop and CV measurements to estimate the number of vehicles in a freeway section, demonstrating that the use of fusion data resulted in improving the estimation accuracy. Another study [11] developed a KF approach using fused loop and CV data to estimate the number of vehicles in a signalized link.

Recently, a few studies have attempted to estimate the number of vehicles on signalized links using CV data only. Several benefits of using CV data have been recognized; for example, the high quality of data compared with existing data sources (e.g., cameras and loop detectors), the data can be collected at any location inside the network, thus offering a clear picture about traffic behavior at any time and location, and the cheap cost of collecting the data given that no additional infrastructure is required. In those studies, the linear KF, linear adaptive KF (AKF), and nonlinear PF model-driven approaches were developed to provide accurate estimates [12]–[15]. Moreover, a comprehensive comparison between the KF, AKF, and PF was performed. It was found that: (1) the AKF and PF are very sensitive to the system's initial condition (e.g., vehicle count), while the simple KF is the least sensitive to the initial condition, and (2) the PF and AKF require more computational time than the KF [15].

B. DATA-DRIVEN APPROACHES

Machine learning techniques require considerable amounts of data to build mathematical models that draw the relationship between the model's inputs and outputs, and thus machine learning is considered a data-driven technique. Data-driven approaches have been employed to estimate traffic state variables such as traffic stream density and speed [14], [16]–[21]. In those previous studies, the proposed estimation approaches have relied on different data sources such as data from fixed sensors and fused data.

ANN and k-NN data-driven approaches were developed to produce reliable estimates of vehicle counts [21]. In that study, authors relied on fixed sensors to obtain traffic speed and flow measurements to build and train the ANN and the k-NN approaches. Fulari *et al.* [16] developed an ANN approach to estimate the number of vehicles using video and Bluetooth data. It was found that the ANN approach performs well if a good quantity of training data is accessible. Fused loop and CV data were used to develop support vector machine and k-NN approaches, with the aim of estimating the level of traffic congestion in a freeway segment [22]. Another study [20] deployed data from fixed sensors and CVs to build different data-driven estimation approaches such as ANN, k-NN, and RF to estimate the hourly traffic volumes. In that study, the ANN was found to outperform the other approaches. Aljamal *et al.* [14] developed an ANN approach to estimate the level of market penetration (LMP) rate of the CVs. In that study, the ANN approach provides the AKF approach with real-time values of the LMPs, resulting in improving the vehicle count estimation accuracy. The LMP represents the percentage of the CVs to the total number of vehicles.

In summary, studies have shown the benefits of using data-driven approaches in addressing different aspects of the traffic state estimation problem. Therefore, the research described in this paper aims to develop data-driven approaches in the application of traffic stream density estimation (vehicle counts). One commonality among the related studies is that they all estimated the vehicle counts using data from fixed sensors or using fused source data (e.g., loop with CV data).

The research described in this study aims to develop different data-driven estimation techniques to estimate the vehicle counts using only CV data. The proposed estimation approaches are applied to test a signalized link in downtown Blacksburg, Virginia. The proposed research extends the state-of-the-art in vehicle count estimation by making three major contributions:

- 1) This study develops three data-driven estimation approaches (ANN, k-NN, and RF) to estimate the vehicle counts in signalized links. The three data-driven approaches are developed using only CV data.
- 2) This research develops a data-driven approach to estimate the LMP for the CVs at the entrance and the exit of the link.
- 3) This study compares the three proposed data-driven approaches with state-of-the-art model-driven estimation approaches KF, AKF, and PF).

The paper is organized as follows: Section III demonstrates the development of the simulation data. Section IV presents the proposed estimation approaches. Section V shows the findings of the estimation approaches. Section VI presents the conclusions of the paper and potential future work.

III. DEVELOPMENT OF SIMULATION DATA

A congested link in downtown Blacksburg, Virginia, was selected to evaluate the proposed estimation approaches. The link falls between two traffic signals, as shown in Figure 1. The link length is 97 meters. The INTEGRATION microscopic traffic assignment and simulation software [23]–[26] was used to simulate the network in Figure 1. The INTEGRATION software tracks vehicle longitudinal motion using the Rakha–Pasumarthy–Adjerid collision-free car-following model, also known as the RPA model [27]. The RPA model captures vehicle steady-state car-following behavior using the Van Aerde model [28], [29]. Movement from one steady state to another is constrained by a vehicle dynamics model described in [30], [31]. Vehicle lateral motion is modeled using lane-changing models described in [25]. The model estimates of vehicle delay were validated in [32], while vehicle stop estimation procedures were described and validated in [33]. The traffic origin-destination (O-D) values for the network were calibrated using real count data. The speed limit of the tested link is 40 km/h, the speed-at-capacity is 32 km/h, the jam density is 160 veh/km/ln, and the saturation flow rate is 1800 veh/h/lane.

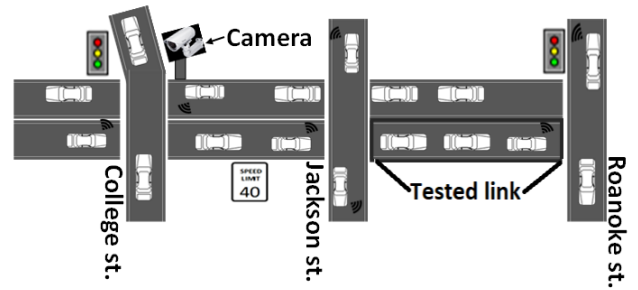


FIGURE 1. Tested link section in downtown Blacksburg, VA.

A. GENERATION OF THE TRAINING DATA SET

Training data are needed to develop machine learning estimation approaches. The INTEGRATION simulation software was used to generate the CV data given that empirical CV data are costly to gather. A total of 1000 scenarios were simulated using different input factors:

- Different scaling factors of the base O-D table,
- Different right-turn traffic volumes that exit Main Street toward Jackson Street, and
- Different random seeds for each LMP scenario.

For right-turn traffic volumes and demand O-Ds, 20 different scaling factors were generated from a uniform distribution, ranging from 0.8 to 1.2; for example, a scenario could have a 0.82 O-D demand scaling factor and a 1.05 right-turn volume demand scaling factor. The INTEGRATION simulation software generates an output time-space file that includes real-time information about the CVs, such as the vehicle's location and speed. In section IV, more details are provided about the inputs and outputs that are considered in the training data set.

IV. METHODOLOGY

In this section, three research approaches are presented: (1) model-driven approaches, (2) integrating data-driven and model-driven approaches, and (3) data-driven approaches. In the first research approach, linear and nonlinear filtering approaches are used to estimate the vehicle counts. The second approach first develops a data-driven approach to estimate the ratio of the number of CVs (N_{cv}) to the total number of vehicles (N_T), and then combines the data-driven approach with the most accurate model-driven approach to finally estimate the vehicle counts. The third approach develops data-driven approaches to directly estimate the vehicle counts.

A. FIRST APPROACH: MODEL-DRIVEN APPROACHES

Linear and nonlinear filtering approaches are presented in this section, namely: 1) KF, 2) AKF, and (3) PF. These filtering techniques are always used to solve state-space models. A state-space model is represented by: (1) a state, and (2) a measurement system. The filtering techniques are mainly

used to provide posterior estimates given some measurements with the aim of minimizing the errors in the a priori estimates.

In this paper, the state-space model presented in [12] is used to estimate the vehicle counts. The state and measurement equations are presented in Equations (1) and (3), respectively.

$$N(t) = N(t - \Delta t) + u(t) \quad (1)$$

$$u(t) = \frac{\Delta t [q^{in}(t) - q^{out}(t)]}{\max(\rho, \rho_{min})} \quad (2)$$

$$TT(t) = H(t) \times N(t) \quad (3)$$

$$H(t) = \frac{2 \times \rho}{q^{in}(t) + q^{out}(t)} \quad (4)$$

where $N(t)$ is the number of vehicles crossing the link at time t , $N(t - \Delta t)$ is the number of vehicles crossing the link in the preceding time interval, $u(t)$ is the system input, ρ is the CVs' LMP, defined as the ratio of the CV counts to the total vehicle counts. In this research approach, the ρ is computed from historical data and assumed to remain constant for the entire simulation. For instance, if a scenario of 10% LMP is evaluated, the ρ value is assumed to be 10%. q^{in} and q^{out} represent the flow of CVs entering and exiting the link, respectively, during Δt . The Δt is updated when 5 CVs traversed the tested link [12]. TT is the average travel time for CVs.

The following subsections present three filtering techniques to solve the described state-space model.

1) THE KF APPROACH

The KF [34] is a linear filtering technique and can be implemented using the following equations:

$$\hat{N}^-(t) = \hat{N}^+(t - \Delta t) + u(t) \quad (5)$$

$$\hat{TT}(t) = H(t) \times \hat{N}^-(t) \quad (6)$$

$$\hat{P}^-(t) = \hat{P}^+(t - \Delta t) \quad (7)$$

$$G(t) = \hat{P}^-(t)H(t)^T [H(t)\hat{P}^-(t)H(t)^T + R]^{-1} \quad (8)$$

$$\hat{N}^+(t) = \hat{N}^-(t) + G(t) [TT(t) - \hat{TT}(t)] \quad (9)$$

$$\hat{P}^+(t) = \hat{P}^-(t) \times [1 - H(t)G(t)] \quad (10)$$

where \hat{N}^- and \hat{N}^+ are the priori and the posterior vehicle count estimates, \hat{TT} is the estimated average travel time, \hat{P}^- and \hat{P}^+ are the priori and posterior covariance estimates for the state system, G is the Kalman gain, and R is the error covariance in the measurement system.

The state error covariance is a tuning parameter that quantifies the uncertainty of the state's estimate represented by the mean. If the state covariance value is low, then the state estimation outcome is accurate and close to the actual value. In real-applications, the user needs to define the initial estimates of the state error covariance $\hat{P}^+(0)$ based on how off the initial vehicle count value is from the actual value. After that, the KF will adjust the initial estimate and provide real-time covariance error estimates using Equations (7) and (10), with the aim of converging the estimation error to zero. For the measurement error covariance value, an offline procedure can

be adopted to provide a reasonable covariance value. The measurement error covariance value is used to represent the accuracy of the system's measurements (in our case the travel time values from the CVs). If we assume that the algorithm is reset each day at midnight, the traffic would be very low and the covariance matrix can be set at a value close to zero. For more details, readers can refer to [12].

2) THE AKF APPROACH

The linear AKF dynamically estimates the noise error values for the state and measurement systems every estimation step. The AKF approach can be solved using the following equations:

$$\hat{N}^-(t) = \hat{N}^+(t - \Delta t) + u(t) + m(t - \Delta t) \quad (11)$$

$$\hat{TT}(t) = H(t) \times \hat{N}^-(t) \quad (12)$$

$$\hat{P}^-(t) = \hat{P}^+(t - \Delta t) + M(t - \Delta t) \quad (13)$$

$$r = \frac{1}{n} \sum_{t=1}^n [TT(t) - \hat{TT}(t)] \quad (14)$$

$$R = \frac{1}{n-1} \sum_{t=1}^n [(r(t) - r).(r(t) - r)^T - (\frac{n-1}{n}) \times H(t)\hat{P}^-(t)H^T(t)] \quad (15)$$

$$G(t) = \hat{P}^-(t)H(t)^T [H(t)\hat{P}^-(t)H(t)^T + R(t)]^{-1} \quad (16)$$

$$\hat{N}^+(t) = \hat{N}^-(t) + G(t) [TT(t) - \hat{TT}(t) - r(t)] \quad (17)$$

$$\hat{P}^+(t) = \hat{P}^-(t) \times [1 - H(t)G(t)] \quad (18)$$

$$m = \frac{1}{n} \sum_{t=1}^n [\hat{N}^+(t) - \hat{N}^+(t - \Delta t) - u(t) + m(t - \Delta t)] \quad (19)$$

$$M = \frac{1}{n-1} \sum_{t=1}^n [(m(t) - m).(m(t) - m)^T - (\frac{n-1}{n}) \times \hat{P}^+(t - \Delta t) - \hat{P}^+(t)] \quad (20)$$

where r and R are the mean and covariance of the measurement noise, n is the number of state noise samples, and m and M are the mean and covariance of the state noise.

3) THE PF APPROACH

The PF [35] is a nonlinear filtering technique. First, the PF generates different particles with unique relative weights. In every estimation step, the system removes the particles with low relative weights and replaces them with new particles (resampling), thus preserving only the important particles. To compute the posterior value, an average value of the remaining important particles are calculated. The PF approach can be implemented using the following steps:

- Initialization: $t = 0$
- $\hat{N}^+(0)$, R , V , and l .
- Generate particles:

$$N^l(0) \sim P(N_0) \quad (21)$$

- For $t = 1 : T$

$$N^l(t) = N^l(t - \Delta t) + u(t) \quad (22)$$

$$TT^l(t) = H(t) \times N^l(t) \tag{23}$$

$$w^l(t) = \frac{1}{\sqrt{2\pi R}} e^{-(TT- TT^l(t))^2/2R} \tag{24}$$

$$\hat{w}^l(t) = w^l(t) / \sum_{l=1}^L w^l(t) \tag{25}$$

After normalizing the weights using Equation (25), the low-weighted particles are replaced with new particles (resampling [35]). After a few iterations in the PF process, the weight will focus on a few particles only and most particles will have insignificant weights, resulting in sample degeneracy [36]. The resampling process is therefore used to tackle the degeneracy problem.

$$\hat{N}^+(t) = \frac{1}{L} \sum_{l=1}^L N^l(t) \tag{26}$$

where V is the variance of the initial vehicle count estimate, N^l is the particles' locations from 1 to L , and TT is the observed measurement from the CVs. More details can be found in [15].

B. SECOND APPROACH: INTEGRATING DATA-DRIVEN AND MODEL-DRIVEN APPROACHES

In our state-space equations, the ρ variable is found to be the main source of noise in the state-space model [12]. Unlike the first research approach described in IV-A, two ρ variables, instead of one ρ variable, are used in the state-space equations, namely: 1) ρ_{in} , and 2) ρ_{out} . ρ_{in} and ρ_{out} are observed at the entrance and exit of the link, respectively. The ρ_{in} and the ρ_{out} are displayed in Equations (27) and (28), respectively. A_{cv} , A_T , D_{cv} , and D_T are the number of CV arrivals, total number of arrivals, number of CV departures, and total number of departures, respectively. Equations (29) and (30) present the new formulation of the $u(t)$ and $H(t)$ using the two ρ variables.

$$\rho_{in}(t) = A_{cv}/A_T \tag{27}$$

$$\rho_{out}(t) = D_{cv}/D_T \tag{28}$$

$$u(t) = \Delta t \left[\frac{q^{in}(t)}{\rho_{in}(t)} - \frac{q^{out}(t)}{\rho_{out}(t)} \right] \tag{29}$$

$$H(t) = \frac{2}{\frac{q^{in}(t)}{\rho_{in}(t)} + \frac{q^{out}(t)}{\rho_{out}(t)}} \tag{30}$$

It should be noted that the two variables can be measured if two fixed sensors (e.g., cameras) are installed in the entry and exit of the tested link; however, the installation cost is high, thus making this approach undesirable. A more-efficient approach is to employ estimation techniques such as machine learning without the need to add to the existing infrastructure. Hence, in this research approach, an ANN is developed to estimate the ρ_{in} and ρ_{out} variables.

1) ANN APPROACH

The ANN data-driven model is a combination of simple units (nodes) that are connected by links. The ANN aims

to recognize relationships between enormous amounts of data by adding certain number of neurons in the assigned hidden layers. The ANN contains three layers: the input layer, the hidden layer, and the output layer [37]. The mechanism behind the ANN is that every node receives/sends signals from incoming/outgoing links by performing computations. The links that connect the nodes in the network have certain weight values, and these weights determine the strength of connection between the nodes.

a: ANN INPUTS AND OUTPUTS

In this section, the aim was to use the nearest existing fixed sensor with the CV data to build the ANN model. As seen in Figure 1, an existing camera is located upstream of the tested link (at the intersection of College Street). The camera in the field measures the total traffic counts at the intersection. Consequently, the total traffic count variable is used as an input for the ANN model. In addition, CVs are used to generate the inputs of the ANN model as their ability to provide measurements at any location inside the network.

Seven inputs are used to build the ANN approach, as follows:

- 1) The total traffic counts obtained from the camera (C_T),
- 2) The number of CVs on the tested link (N_{cv}),
- 3) The number of CVs at the entrance of the link (A_{cv}),
- 4) The space-mean speed of CVs (u_s),
- 5) The average speed for CVs at link entrance ($S1$),
- 6) The average speed for CVs at link exit ($S2$), and
- 7) The estimation interval time (Δt).

Figure 2 displays the ANN inputs and outputs. To build a strong ANN approach, the inputs must relate to the outputs, which allows the ANN to define the relationship between the inputs and the outputs. For instance, a high traffic volume (C_T , N_{cv} , and A_{cv}) means that we have more vehicles in the link, which results in having large values in the denominator in Equations (27) and (28). The speed factor (u_s , $S1$, and $S2$) is also an important indicator of the level of congestion. A congested link can also result in having large values in the denominator in the two equations. It should be noted that the Δt variable strongly relates to the output variables. Remember that Δt is not a constant value and is updated when new 5 CVs are observed at the end of the link. A high Δt value means that the number of CVs is low, which results in low output values. The ANN output variables are ρ_{in} and ρ_{out} . In reality, the ρ output values vary between 0 and 1; 0 means that no CVs are observed, while the value of 1 means that the number of CVs is equal to the total number of vehicles.

The ANN approach was trained offline using the training data set. The data set was divided into three portions: 70% for training, 15% for validation, and 15% for testing. The reported results in section V used external data sets to evaluate the proposed approaches at different LMPs.

The developed ANN consists of single hidden layer with 10 neurons, with the use of a transfer function of hyperbolic

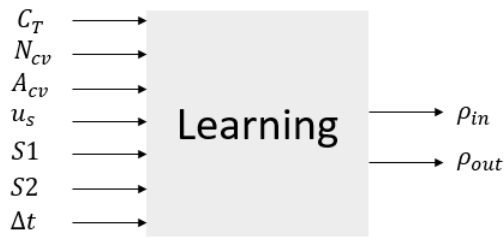


FIGURE 2. Estimating the ρ_{in} and ρ_{out} variables.

tangent sigmoid and the Levenberg-Marquardt (LM) optimization method.

After developing and training the ANN approach, the estimated values for ρ_{in} and ρ_{out} are used in our most accurate model-driven approach to estimate the main research goal, which is the vehicle count.

C. THIRD APPROACH: DATA-DRIVEN APPROACHES

The third research approach aims to directly estimate the vehicle counts by developing different data-driven estimation approaches, namely: (1) ANN, (2) k-NN, and (3) RF. The data-driven approaches were developed using CV data only without the need to use data from the camera. Six inputs were considered to train and build the data-driven approaches, as shown in Figure 3.

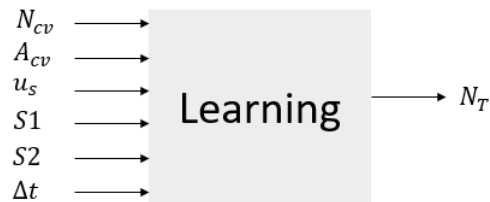


FIGURE 3. Estimating the vehicle counts (N_T) on the tested link.

1) ANN APPROACH

To estimate the vehicle counts, the ANN approach was developed using the following parameters:

- The structure of the ANN consists of single hidden layer with 10 neurons.
- A transfer function of hyperbolic tangent sigmoid and the LM optimization method were used.
- Number of Epoch was 332.
- Training time was around 12 minutes. The ANN was trained in MATLAB R2019a on a Dell PC with 8.0 GB RAM.
- Learning rate was 0.01.
- MU control parameter was 1×10^{-5} .
- R value was almost 0.86 for the training, validation, and testing data set. The R value measures the correlation between model outputs and desired outputs. A value close to 1.0 means that the model outputs are very close to desired outputs.

2) k-NN APPROACH

The k-NN [38] is used for classification and regression applications. The k-NN does not build a model but requires storing the entire data set. To estimate a new value using the k-NN, the following information is required: (1) having access to the training records, (2) defining the distance metric to compute the distance between the records, and (3) identifying the value of the number of nearest neighbors (k). The results section will test different k values to find the optimal k value for the k-NN approach. The new estimated value is computed by taking the average value of the nearest neighbors.

3) RF APPROACH

The RF [39] is a supervised learning technique and can be used in classification and regression. The RF is a set of decision trees. Each decision tree is constructed using a subset of inputs. The desired estimation values are given based on the majority votes from all trees. The advantage of using the RF is the ability to handle a large data set without the need to create dummy variables. For the purpose of this study, 100 trees were used to develop the RF.

V. RESULTS AND DISCUSSION

This section tests the accuracy of the three research approaches on a signalized link in downtown Blacksburg, Virginia. The relative root mean square error (RRMSE) and the root mean square error (RMSE) are used to evaluate and compare the proposed estimation approaches. The RRMSE and RMSE can be computed using Equations (31) and (32), respectively.

$$\text{RRMSE (\%)} = 100 \sqrt{\frac{\sum_{s=1}^S [\hat{N}^+(s) - N(s)]^2}{\sum_{s=1}^S N(s)}} \quad (31)$$

$$\text{RMSE (veh)} = \sqrt{\frac{\sum_{s=1}^S [\hat{N}^+(s) - N(s)]^2}{S}} \quad (32)$$

where $N(s)$ is the actual vehicle count, $\hat{N}^+(s)$ is the estimated vehicle count value, and S is the total number of estimations.

A. FIRST RESEARCH APPROACH

This section evaluates the three estimation model-driven approaches: 1) KF, 2) AKF, and 3) PF, using data from CVs only. The three approaches are used to estimate the number of vehicles crossing the tested link. Table 1 presents the RRMSE and RMSE values at different LMPs; 1, 3, 5, 8, 10, 15, 20, 30, 40, 50, 60, 70, 80, and 90%. For most of the LMP scenarios, the KF approach produces the lowest error values, while the PF approach outperforms the KF and the AKF for a few scenarios (1, 70, 80, and 90%). However, the nonlinear PF and the AKF require more computational time and they are also very sensitive to the initial conditions [15]. Consequently, the use of the linear KF approach is highly recommended due to its simplicity and high-performance accuracy.

TABLE 1. RRMSE and RMSE values of KF, AKF, and PF approaches for different LMPs.

| LMPs % | RRMSE (%), RMSE (veh) | | |
|--------|-----------------------|---------|---------|
| | KF | AKF | PF |
| 1 | 44, 3.2 | 58, 4.3 | 37, 2.8 |
| 3 | 39, 3.0 | 48, 3.6 | 39, 3.0 |
| 5 | 37, 2.8 | 44, 3.2 | 38, 2.9 |
| 8 | 36, 2.8 | 40, 3.0 | 37, 2.9 |
| 10 | 36, 2.8 | 38, 2.9 | 37, 2.9 |
| 15 | 36, 2.8 | 40, 3.0 | 39, 3.0 |
| 20 | 37, 2.8 | 39, 3.0 | 39, 3.0 |
| 30 | 38, 2.9 | 38, 2.9 | 42, 3.2 |
| 40 | 37, 2.9 | 38, 2.9 | 39, 3.0 |
| 50 | 37, 2.8 | 38, 2.9 | 37, 2.8 |
| 60 | 31, 2.4 | 38, 2.9 | 31, 2.4 |
| 70 | 26, 2.0 | 32, 2.5 | 25, 1.9 |
| 80 | 23, 1.8 | 31, 2.4 | 20, 1.5 |
| 90 | 20, 1.5 | 30, 2.2 | 14, 1.1 |

B. SECOND RESEARCH APPROACH

First, the ANN approach is developed to estimate the percentage of the CVs to the total number of vehicles at the entry and the exit of the tested link, ρ_{in} and ρ_{out} , respectively. Table 2 presents the RRMSE values for estimating the two variables. The results demonstrate that the ANN produces reasonable error values; the errors for estimating ρ_{in} vary between 14 and 25%, while the error values are between 10 and 23% for ρ_{out} .

TABLE 2. RRMSE of ρ_{in} and ρ_{out} at different LMPs.

| LMPs % | RRMSE (%) | |
|--------|-------------|--------------|
| | ρ_{in} | ρ_{out} |
| 10 | 14 | 19 |
| 20 | 18 | 23 |
| 30 | 22 | 22 |
| 40 | 25 | 21 |
| 50 | 25 | 19 |
| 60 | 24 | 16 |
| 70 | 24 | 14 |
| 80 | 25 | 12 |
| 90 | 24 | 10 |

After that, the estimated ρ values are used as inputs to the KF to estimate the vehicle counts on the tested link. A new approach, named KFNN, was developed based on integrating the KF and the ANN approaches. Remember that the KF approach uses an average one value from the actual ρ values in its equations, while the KFNN approach uses real-time values for ρ_{in} and ρ_{out} in the KF equations at every estimation step. Table 3 shows the RRMSE values for estimating the vehicle counts using the KF and KFNN approaches. The table demonstrates that the KF approach outperforms the KFNN approach. Investigations were undertaken to find the reason. The investigations found that the ANN may over-estimate ρ_{in}

TABLE 3. RRMSE of KF and KFNN approaches at different LMPs.

| LMPs % | RRMSE (%) | |
|--------|-----------|------|
| | KF | KFNN |
| 10 | 36 | 64 |
| 20 | 37 | 50 |
| 30 | 38 | 52 |
| 40 | 37 | 54 |
| 50 | 37 | 58 |
| 60 | 31 | 59 |
| 70 | 26 | 57 |
| 80 | 23 | 52 |
| 90 | 20 | 35 |

TABLE 4. RRMSE of k-NN approach using different k values.

| k | RRMSE (%) |
|-----------|--------------|
| 1 | 24.63 |
| 2 | 21.63 |
| 3 | 20.47 |
| 4 | 19.91 |
| 5 | 19.49 |
| 6 | 19.27 |
| 7 | 19.07 |
| 8 | 18.88 |
| 9 | 18.81 |
| 10 | 18.68 |
| 11 | 18.54 |
| 12 | 18.58 |
| 13 | 18.48 |
| 14 | 18.47 |
| 15 | 18.47 |

and under-estimate ρ_{out} or vice versa for the same estimation step, resulting in large errors in the state equation compared to the errors from using the average ρ . Such large errors make the error corrections from the KF difficult. In conclusion, the use of one single ρ value in the state-space equations is sufficient to produce accurate estimates.

In next section, data-driven approaches are developed to directly estimate the vehicle counts without the need for model-driven approaches.

C. THIRD RESEARCH APPROACH

This section utilizes the three data-driven approaches to estimate the number of vehicles traversing the tested link. The data source used to train and build the three approaches was only CV data without the need of the camera data. For practical considerations, the only information that is needed in practice is as follows: (1) the number of connected vehicles (CVs) that enter the subject link, (2) the number of CVs on the subject link, (3) the space-mean speed of CVs, (4) the average speed for CVs at the entrance and the exit of the link, and (5) the estimation time interval duration.

TABLE 5. RRMSE and RMSE values using data-driven approaches.

| LMPs % | RRMSE (%), RMSE (veh) | | |
|--------|-----------------------|---------|---------|
| | ANN | k-NN | RF |
| 1 | 27, 2.1 | 40, 3.1 | 42, 3.2 |
| 3 | 29, 2.2 | 37, 2.8 | 42, 3.2 |
| 5 | 29, 2.2 | 37, 2.8 | 38, 2.8 |
| 8 | 28, 2.1 | 36, 2.7 | 38, 2.8 |
| 10 | 27, 2.1 | 35, 2.8 | 38, 2.8 |
| 15 | 25, 1.9 | 33, 2.4 | 36, 2.7 |
| 20 | 24, 1.8 | 33, 2.4 | 36, 2.7 |
| 30 | 22, 1.7 | 30, 2.3 | 34, 2.6 |
| 40 | 20, 1.5 | 27, 2.1 | 30, 2.3 |
| 50 | 17, 1.3 | 24, 1.8 | 26, 2.0 |
| 60 | 14, 1.1 | 22, 1.6 | 23, 1.7 |
| 70 | 11, 0.9 | 20, 1.5 | 21, 1.6 |
| 80 | 9, 0.7 | 18, 1.4 | 19, 1.5 |
| 90 | 8, 0.6 | 17, 1.3 | 17, 1.3 |

TABLE 6. RRMSE and RMSE values using model- and data-driven approaches.

| LMPs % | Model-Driven Approaches | | | Data-Driven Approaches | | |
|--------|-------------------------|---------|---------|------------------------|---------|---------|
| | KF | AKF | PF | ANN | k-NN | RF |
| 1 | 44, 3.2 | 58, 4.3 | 37, 2.8 | 27, 2.1 | 40, 3.1 | 42, 3.2 |
| 3 | 39, 3.0 | 48, 3.6 | 39, 3.0 | 29, 2.2 | 37, 2.8 | 42, 3.2 |
| 5 | 37, 2.8 | 44, 3.2 | 38, 2.9 | 29, 2.2 | 37, 2.8 | 38, 2.8 |
| 8 | 36, 2.8 | 40, 3.0 | 37, 2.9 | 28, 2.1 | 36, 2.7 | 38, 2.8 |
| 10 | 36, 2.8 | 38, 2.9 | 37, 2.9 | 27, 2.1 | 35, 2.8 | 38, 2.8 |
| 15 | 36, 2.8 | 40, 3.0 | 39, 3.0 | 25, 1.9 | 33, 2.4 | 36, 2.7 |
| 20 | 37, 2.8 | 39, 3.0 | 39, 3.0 | 24, 1.8 | 33, 2.4 | 36, 2.7 |
| 30 | 38, 2.9 | 38, 2.9 | 42, 3.2 | 22, 1.7 | 30, 2.3 | 34, 2.6 |
| 40 | 37, 2.9 | 38, 2.9 | 39, 3.0 | 20, 1.5 | 27, 2.1 | 30, 2.3 |
| 50 | 37, 2.8 | 38, 2.9 | 37, 2.8 | 17, 1.3 | 24, 1.8 | 26, 2.0 |
| 60 | 31, 2.4 | 38, 2.9 | 31, 2.4 | 14, 1.1 | 22, 1.6 | 23, 1.7 |
| 70 | 26, 2.0 | 32, 2.5 | 25, 1.9 | 11, 0.9 | 20, 1.5 | 21, 1.6 |
| 80 | 23, 1.8 | 31, 2.4 | 20, 1.5 | 9, 0.7 | 18, 1.4 | 19, 1.5 |
| 90 | 20, 1.5 | 30, 2.2 | 14, 1.1 | 8, 0.6 | 17, 1.3 | 17, 1.3 |

Vehicle-to-Infrastructure (V2I) communication can provide this information to the traffic signal controller.

First, different neighbors (k) were tested to calibrate and train the k-NN approach, as shown in Table 4. The optimal k was found to be 14, with an RRMSE of 18.47%.

After calibrating the data-driven estimation approaches, external data were used to test and evaluate the performance of the estimation approaches. Table 5 presents the RRMSE and RMSE values using the three data-driven estimation approaches: ANN, k-NN, and RF. The results demonstrate that the ANN outperforms the k-NN and the RF for all LMP scenarios.

Next, the paper compares the performance of the model-driven approaches (KF, AKF, and PF) and the data-driven approaches (ANN, k-NN, and RF) for

the application to traffic stream density. Table 6 summarizes the RRMSE and RMSE values using the six estimation approaches. The table demonstrates that the ANN approach produces the most accurate estimates compared with the other approaches. However, it is worth mentioning the difficulties of applying this approach in the field due to the huge amount of data needed to train and build the ANN approach, especially for a large network (e.g., Los Angeles, CA). Moreover, sudden changes in traffic behaviors (e.g., incidents) would not always ensure accurate estimates and thus might lead to worsen the performance of the traffic signal controller. Consequently, we recommend using the KF approach for traffic density estimation due to its simplicity and applicability in the field.

VI. SUMMARY AND CONCLUSION

The paper presents three approaches to estimate the number of vehicles along signalized links. The first approach includes three model-driven estimation techniques (KF, AKF, and PF) using solely CV data. The first approach uses a single average ρ value, obtained from the actual historical LMPs, in the state-space equations. The second research approach develops an ANN to estimate two ρ variables, ρ_{in} and ρ_{out} , to be used in the state-space equations. Fused CV and camera data are utilized to build the ANN. After that, the second approach integrates the ANN with the KF (KFNN approach) to estimate the number of vehicles on signalized links. The third approach includes three data-driven techniques (ANN, k-NN, and RF) to directly estimate the number of vehicles using only CV data. The three research approaches were applied on a signalized link in downtown Blacksburg, Virginia. The main findings and conclusions of the paper are summarized as follows:

- The use of CV data is sufficient to provide accurate vehicle count estimates.
- The use of two estimated variable values in the state-space equations is not recommended as it may produce undesired large errors in the state equation. It was found that the ANN approach may over-estimate the first variable and under-estimate the second variable or vice versa for the same estimation step. Consequently, the second research approach is not recommended.
- The ANN is the most accurate estimation approach. However, taking into consideration the large amount of data needed to train and build the ANN, the long computational time needed to build the ANN, and the constraints on keeping the traffic behavior the same as the behavior in the training data set, the use of the KF approach is highly recommended for the application of traffic density due to its simplicity and applicability in the field.

Proposed Future work entails testing the performance of the traffic signal controller using the outcomes of the KF approach as inputs and developing online learning techniques to estimate the number of vehicles to adapt for local traffic conditions.

REFERENCES

- [1] Trident Academy of Technology Institute. (2019). *Commuters Waste About 54 Hours a Year in Traffic, and it's Not Just During Rush Hour*. [Online]. Available: https://www.nwintimes.com/news/national/commuters-waste-about-hours-a-year-in-traffic-and-it/collection_eac664b9-5300-55bf-bffe-3dbc68a8483.html
- [2] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 630–638, Sep. 2010.
- [3] M. A. Aljamal, "Real-time estimation of traffic stream density using connected vehicle data," Ph.D. dissertation, Charles E. Via, Jr. Dept. Civil Environ. Eng., Virginia Tech, Blacksburg, VA, USA, 2020.
- [4] G. Vigos, M. Papageorgiou, and Y. Wang, "Real-time estimation of vehicle-count within signalized links," *Transp. Res. C, Emerg. Technol.*, vol. 16, no. 1, pp. 18–35, Feb. 2008.
- [5] D. Ghosh and C. H. Knapp, "Estimation of traffic variables using a linear model of traffic flow," *Transp. Res.*, vol. 12, no. 6, pp. 395–402, Dec. 1978.
- [6] N. Bhouiri, H. H. Salem, M. Papageorgiou, and J. M. Blossville, "Estimation of traffic density on motorways," in *Proc. IFAC/IFIP/IFORS Int. Symp. (AIPAC)*, 1989, pp. 579–583.
- [7] L. E. Y. Mimbela and L. A. Klein, "Summary of vehicle detection and surveillance technologies used in intelligent transportation systems," Southwest Technol. Develop. Inst., New Mexico State Univ., Las Cruces, NM, USA, Tech. Rep., 2000. [Online]. Available: <https://www.fhwa.dot.gov/ohim/tvtv/vdstits.pdf>
- [8] J. C. Herrera and A. M. Bayen, "Traffic flow reconstruction using mobile sensors and loop detector data," presented at the 85th Transp. Res. Board Annu. Meeting, Washington, DC, USA, 2008, Paper 08-1868.
- [9] N. Bekiaris-Liberis, C. Roncoli, and M. Papageorgiou, "Highway traffic state estimation with mixed connected and conventional vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3484–3497, Dec. 2016.
- [10] M. Wright and R. Horowitz, "Fusing loop and GPS probe measurements to estimate freeway density," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3577–3590, Dec. 2016.
- [11] X. Di, H. X. Liu, and G. A. Davis, "Hybrid extended Kalman filtering approach for traffic density estimation along signalized arterials: Use of global positioning system data," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2188, no. 1, pp. 165–173, Jan. 2010.
- [12] M. A. Aljamal, H. M. Abdelghaffar, and H. A. Rakha, "Real-time estimation of vehicle counts on signalized intersection approaches using probe vehicle data," *IEEE Trans. Intell. Transp. Syst.*, early access, Feb. 21, 2020, doi: 10.1109/TITS.2020.2973954.
- [13] M. A. Aljamal, H. M. Abdelghaffar, and H. A. Rakha, "Kalman filter-based vehicle count estimation approach using probe data: A multi-lane road case study," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 4374–4379.
- [14] M. A. Aljamal, H. M. Abdelghaffar, and H. A. Rakha, "Developing a neural-Kalman filtering approach for estimating traffic stream density using probe vehicle data," *Sensors*, vol. 19, no. 19, p. 4325, 2019.
- [15] M. A. Aljamal, H. M. Abdelghaffar, and H. A. Rakha, "Estimation of traffic stream density using connected vehicle data: Linear and nonlinear filtering approaches," *Sensors*, vol. 20, no. 15, p. 4066, 2020.
- [16] S. Fulari, L. Vanajakshi, and S. C. Subramanian, "Artificial neural network-based traffic state estimation using erroneous automated sensor data," *J. Transp. Eng., A, Syst.*, vol. 143, no. 8, 2017, Art. no. 05017003.
- [17] C. Antoniou and H. N. Koutsopoulos, "Estimation of traffic dynamics models with machine-learning methods," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1965, no. 1, pp. 103–111, Jan. 2006.
- [18] T. Wassantachat, Z. Li, J. Chen, Y. Wang, and E. Tan, "Traffic density estimation with on-line SVM classifier," in *Proc. 6th IEEE Int. Conf. Adv. Video Signal Based Surveill.*, Sep. 2009, pp. 13–18.
- [19] A. Jahangiri, H. A. Rakha, and T. A. Dingus, "Adopting machine learning methods to predict red-light running violations," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 650–655.
- [20] P. Sekuła, N. Marković, Z. V. Laan, and K. F. Sadabadi, "Estimating historical hourly traffic volumes via machine learning and vehicle probe data: A Maryland case study," *Transp. Res. C, Emerg. Technol.*, vol. 97, pp. 147–158, Dec. 2018.
- [21] J. Raj, H. Bahuleyan, and L. D. Vanajakshi, "Application of data mining techniques for traffic density estimation and prediction," *Transp. Res. Procedia*, vol. 17, pp. 321–330, Dec. 2016.
- [22] S. M. Khan, K. C. Dey, and M. Chowdhury, "Real-time traffic state estimation with connected vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 7, pp. 1687–1699, Jul. 2017.
- [23] M. V. Aerde and H. A. Rakha, "Integration release 2.40 for windows: User's guide-volume I: Fundamental model features," Center Sustain. Mobility, VTTI, Virginia Tech, Blacksburg, VA, USA, Tech. Rep., 2020, doi: 10.13140/RG.2.2.16142.15682/1.
- [24] M. V. Aerde and H. A. Rakha, "Integration release 2.40 for windows: User's guide-volume II: Advanced model features," Center Sustain. Mobility, Virginia Tech Transp. Inst., Blacksburg, VA, USA, Tech. Rep., 2020, doi: 10.13140/RG.2.2.24434.32969.
- [25] H. A. Rakha and Y. Zhang, "The integration 2.30 framework for modeling lane-changing behavior in weaving sections," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1883, no. 1, pp. 140–149, 2004.
- [26] M. A. Aljamal, H. A. Rakha, J. Du, and I. El-Shawarby, "Comparison of microscopic and mesoscopic traffic modeling tools for evacuation analysis," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2321–2326.
- [27] H. A. Rakha, P. Pasumarthy, and S. Adjerid, "A simplified behavioral vehicle longitudinal motion model," *Transp. Lett., Int. J. Transp. Res.*, vol. 1, no. 2, pp. 95–110, 2009.
- [28] H. A. Rakha, "Validation of van aerde's simplified steady-state car-following and traffic stream model," *Transp. Lett., Int. J. Transp. Res.*, vol. 1, no. 13, pp. 227–244, 2009.
- [29] N. Wu and H. Rakha, "Derivation of van aerde traffic stream model from tandem-queueing theory," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2124, no. 1, pp. 18–27, Jan. 2009.
- [30] H. Rakha, I. Lucic, S. H. Demarchi, J. R. Setti, and M. V. Aerde, "Vehicle dynamics model for predicting maximum truck acceleration levels," *J. Transp. Eng.*, vol. 127, no. 5, pp. 418–425, Oct. 2001.
- [31] H. A. Rakha, M. Snare, and F. Dion, "Vehicle dynamics model for estimating maximum light duty vehicle acceleration levels," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1883, no. 1, pp. 40–49, 2004.
- [32] F. Dion, H. Rakha, and Y.-S. Kang, "Comparison of delay estimates at under-saturated and over-saturated pre-timed signalized intersections," *Transp. Res. B, Methodol.*, vol. 38, no. 2, pp. 99–122, Feb. 2004.
- [33] H. Rakha, Y.-S. Kang, and F. Dion, "Estimating vehicle stops at undersaturated and oversaturated fixed-time signalized intersections," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1776, no. 1, pp. 128–137, Jan. 2001.
- [34] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, Mar. 1960.
- [35] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *J. Amer. Stat. Assoc.*, vol. 93, no. 443, pp. 1032–1044, Aug. 1998.
- [36] T. Li, T. P. Sattar, and S. Sun, "Deterministic resampling: Unbiased sampling to avoid sample impoverishment in particle filters," *Signal Process.*, vol. 92, no. 7, pp. 1637–1645, Jul. 2012.
- [37] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1994.
- [38] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
- [39] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.



MOHAMMAD A. ALJAMAL (Member, IEEE) received the B.Sc. degree in civil engineering from the Jordan University of Science and Technology, Irbid, Jordan, in 2014, and the M.Sc. and Ph.D. degrees in civil engineering from Virginia Tech University, Blacksburg, VA, USA, in 2017 and 2020, respectively. He is currently a Transportation Engineer with Precision Systems, Inc., Washington, DC, USA. His research interests include traffic flow theory, intelligent transportation systems, machine learning, connected automated vehicles, and traffic modeling and simulation. He is also a member of ITE and Jordanian Engineers Association (JEA). He received the US DOT Award for Outstanding Student of the Year in 2018, the First place in the IEEE ITSC 2020 UAS4T Competition, and Thomas N. Hunnicutt Fellowship from the Department of Civil and Environmental Engineering, Virginia Tech, in 2017. In addition, he is a Reviewer for several journals and conferences, including the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (ITS), IEEE ACCESS, the IEEE ITS Conference, and the TRB.



MOHAMED FARAG received the B.Sc. degree (Hons.) in computer engineering from Alexandria University, Alexandria, Egypt, in 2006, the M.Sc. degree in computer science from Arab Academy for science, Technology, and Maritime Transport, Alexandria, Egypt, in 2010, and the Ph.D. degree in computer science from Virginia Tech University, Blacksburg, VA, USA, in 2016. He is currently an Assistant Professor of computer science with the Department of Computer Science, Arab

Academy for Science, Technology, and Maritime Transport, and a Research Scholar with the Center of Sustainable Mobility, Virginia Tech Transportation Institute. His research interests include information retrieval, machine learning, large-scale data analysis, and big data.



HESHAM A. RAKHA (Fellow, IEEE) received the Ph.D. degree from Queen's University, Kingston, ON, USA, in 1993. He is currently the Samuel Reynolds Pritchard Professor of Engineering with the Charles E. Via, Jr. Department of Civil and Environmental Engineering and the Bradley Department of Electrical and Computer Engineering (Courtesy), Virginia Tech, and the Director of the Center for Sustainable Mobility, Virginia Tech Transportation Institute. His research inter-

ests include large-scale transportation system optimization, modeling, and assessment. He also works on optimizing transportation system operations, including vehicle routing, developing various network and traffic signal control algorithms, developing freeway control strategies (speed harmonization and ramp metering), and optimizing vehicle motion [lateral and longitudinal control of connected automated vehicles (CAVs)] to enhance their efficiency and reduce their energy consumption while ensuring their safety. He was the author or coauthor of six conference best papers awards, namely: 19th ITS World Congress in 2012, 20th ITS World Congress in 2013, VEHITS in 2016, VEHITS in 2018, and TRB in 2020; received the most cited paper award from the *International Journal of Transportation Science and Technology (IJTST)* in 2018; and received First place in the IEEE ITSC 2020 UAS4T Competition. In addition, he received the Virginia Tech's Dean's Award for Outstanding New Professor in 2002, the College of Engineering Faculty Fellow from 2004 to 2006, and the Dean's Award for Excellence in Research in 2007. He is an Editor of *Sensors* (the Intelligent Sensors Section), an Academic Editor of the *Journal of Advanced Transportation*, an Associate Editor of the IEEE Transactions on Intelligent Transportation Systems and the *Journal of Intelligent Transportation Systems: Technology, Planning and Operations*. Furthermore, he is on the Editorial Board of the *Transportation Letters: The International Journal of Transportation Research* and the *International Journal of Transportation Science and Technology*.

• • •