Publications

Glenn Department of Civil Engineering

Real-time Traffic State Estimation with Connected Vehicles

Sakib Mahmud Khan

Kaken Dey

Mashrur Chowdhury

Real-time Traffic State Estimation with Connected Vehicles

Sakib Mahmud Khan, Kakan Dey, and Mashrur Chowdhury, *Senior Member, IEEE*

Abstract — **A novel framework is developed in this research, to increase the real-time roadway traffic condition assessment accuracy, which integrates connected vehicle technology with artificial intelligence paradigm forming a CVT-AI method. Traffic density is a major indicator of traffic conditions. In this paper, the traffic operational condition is assessed based on traffic density. A simulated network of Interstate 26 in South Carolina is developed to investigate the effectiveness of the method. The assumption is that the vehicle on-board units will forward the connected vehicle (CV) generated data to the edge devices (e.g., roadside units) for further processing. CV generated distance headway, number of stops, and speed data are used to estimate traffic density. This study reveals that with 20% and greater CV penetration levels, the accuracy of the density information with the AI-aided CVT is a minimum of 85%. Moreover, this study demonstrates that the integrated CVT-AI method yields a higher accuracy with the increase of CV penetration levels. Level of service (LOS) is the indicator of traffic congestion level on highways, and is described with traffic density in terms of passenger car/mile/lane for a specific free flow speed. LOS estimated using the CVT-AI density estimation method are compared with the density estimation algorithm used by the Caltrans Performance Measurement System (PeMS), which relies on the occupancy and flow data collected by the freeway inductive loop detectors. With a 10% or more CV penetration, higher accuracy is achieved using the CVT-AI algorithm compared to the PeMS density estimation algorithm.**

Index Terms— **Connected vehicles, Density estimation, LOS, Artificial intelligence, PeMS, Loop detector**

I. Introduction

FOR efficient management of intelligent transportation systems (ITS), accurate assessment of real-time traffic operational conditions is critical. ITS technologies and solutions have been developed and deployed over the last three decades as a cost-effective investment to improve capacity of existing transportation system [1]. One key feature of ITS applications is the real-time traffic operational assessment. The goal of real-time condition analyses is to immediately collect the traffic information and estimate the traffic condition for real-time traffic management, congestion monitoring, and road-user (traveler and concerned agencies) information. In recent years, connected vehicle technology (CVT) has been considered as the next big innovation platform for ITS. Over a wireless communication network, connected vehicles (CVs) will reliably share the traffic condition data with surrounding vehicles through vehicle-to-vehicle (V2V) communication and with transportation infrastructures through the vehicle-to-infrastructure (V2I) communication. Once the CV on-board unit compiles traffic data (e.g., vehicle position, number of brakes applied etc.) in an autonomous way at predetermined intervals, they transmit the data to roadside units (RSUs). Later, these data are further processed to derive the additional vehicle kinetics data (e.g., average speed, acceleration etc.) [2].

Reliability of this vehicle-generated data for incident detection, congestion identification, vehicle routing, and increasing energy efficiency has already been explored in previous research [3-8].

Different state departments of transportation (DOTs) have recognized CVT as a viable option for congestion monitoring and management in future [9-11]. The 'Vehicle Data for Traffic Operations' application defined in the Connected Vehicle Reference Implementation Architecture (CVRIA) uses the CV generated data to support traffic operations by implementing localized operational strategies, as shown in Fig. 1 [12]. This application architecture shows the interconnectivity between different physical objects (i.e., entities) to implement real-time operational strategies using CV data. Once the RSUs collect vehicle situation data (including traffic condition monitoring data) from nearby CVs, it sends the data to the Traffic Management Center (TMC). TMC assigns the parameters to control the traffic condition information flow from the RSUs. Finally, estimated traffic condition data by TMC will be forwarded to other centers (i.e., the maintenance and construction center, the transportation information center, and the emergency management center for further actions (e.g., traffic information dissemination to road users, emergency road maintenance etc.).

Estimation of traffic density is the most important task in real-time roadway network condition assessment [13, 14]. Several other traffic measures like speed or traffic volume can act as surrogate indicators for a congested situation, but density has been identified as the most important parameter to identify traffic congestion [15]. Traditionally real-time traffic operational analysis by density estimation has been conducted using embedded inductive loop detectors, surveillance cameras, and hybrid methods (e.g., loop detector-probe vehicles, loop detector-chase-car method, etc.) [16, 17]. The freeway Performance Measurement System (PeMS), managed by Caltrans utilizes single inductive loop detectors to collect occupancy (i.e., the amount of the time vehicles occupy the detector) and flow data (i.e., the number of vehicles that cross the detector for a specific time period) to assess freeway traffic density [18]. To estimate density, at first PeMS employs an adaptive g-factor (i.e., the effective length of the vehicle) approach to derive speed. Here the basic assumption is that an initial g-factor can be assessed when traffic is under free-flow conditions [19]. Although the estimated density is used in planning and other air quality studies [20], the density estimation algorithm used in PeMS has limitations. For example, loop detectors can be damaged due to vehicle weight and/or poor maintenance. If no data are received due to loop detector malfunction or the collected values are not reliable

S.M. Khan and M. Chowdhury are with the Glenn Department of Civil Engineering, Clemson University, Clemson, SC 29634 USA (e-mail: sakibm@g.clemson.edu, mac@clemson.edu), K Dey is with the Department of Civil and Environmental Engineering, West Virginia University, Morgantown, WV 26505 USA (e-mail: kakan.dey@mail.wvu.edu)
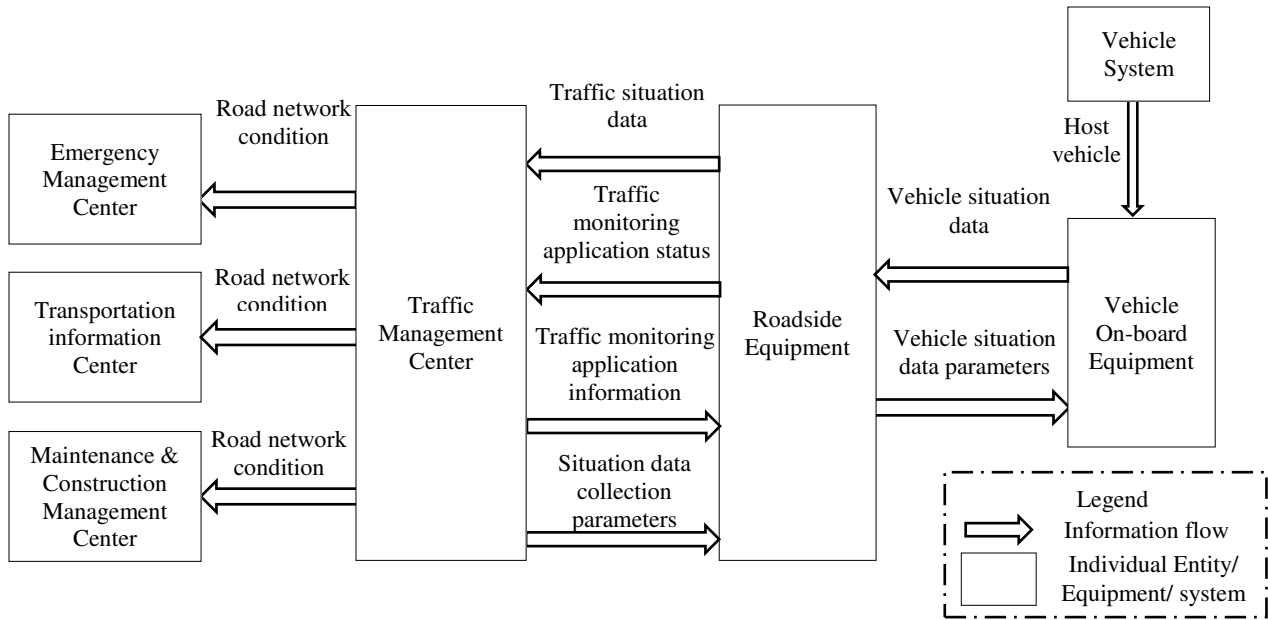
Fig. 1. Architecture of a real-time traffic operational condition assessment application in a CV environment (adapted from [12])

(i.e., flawed), those data are replaced with estimated or imputed data, which may lead to erroneous density estimation. According to the PeMS website, for the year 2015 (from January 4, 2015 to December 27, 2015), on average 40.5% of the total data collected from detectors is estimated or imputed, which is reasonably high [21]. Moreover, PeMS assumes a constant flow and speed across the entire segment during a particular time interval [22], which is less valid for longer freeway sections. On the contrary, CVT is a promising technology for continuous real-time traffic data collection [2, 6, 23] to estimate density, and disseminating the traffic information to the surrounding vehicles and infrastructures more efficiently. However, few research has been undertaken regarding the use of the integrated CVT-AI method to estimate traffic density, and this knowledge gap has been the motivation for this research. The specific research objectives are: (i) to evaluate the effectiveness of CVT in estimating the freeway Level of Service (LOS) based on density values, (ii) to explore the feasibility of different AI methods to estimate real-time density in a connected vehicle environment, (iii) to identify CV

penetration levels to reliably estimate freeway LOS, and (iv) to compare the freeway density estimated by the CVT-AI density estimation method with the loop detector based density estimation (e.g., PeMS) algorithm. For this study, Interstate 26 (I-26) in South Carolina (between Exit 194 and Exit 187) is chosen as the study area.

In Section II, previous research works related to (i) traffic density estimation methods, and (ii) the application of AI in traffic operational assessments are presented. The discussion on the adopted research method in this research is introduced in Section III. In Section IV, the detailed analysis of the results is presented based on the findings from the CVT-AI method and the PeMS density estimation algorithm. Section V summarizes the research contribution. Section VI concludes the paper by presenting further discussion on the research findings and the future research directions of CV applications in real-time state estimation.
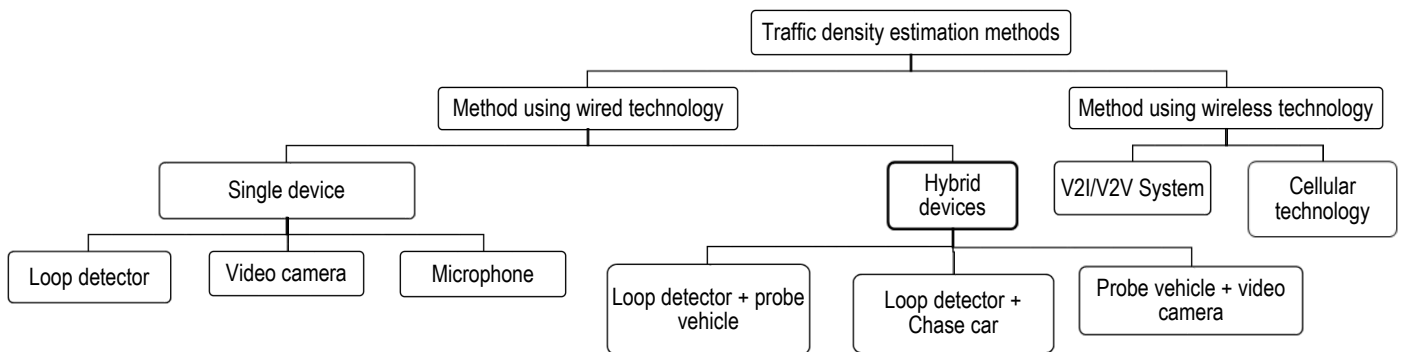

Fig. 2 Classification of traffic density estimation methods

II. LITERATURE REVIEW

The primary goal of this section is to summarize the evolution of the knowledge domain in real-time traffic state assessment using density estimation and the application of AI for traffic operational analysis.

A. Traffic Density Estimation Methods

Different studies have been conducted to determine existing traffic operational conditions. Vehicle speed [24, 25], vehicle spacing (i.e., distance between two vehicles) [26], and density [13] have been used to estimate traffic operational conditions. However, Qui et al. identified traffic density as a requirement to establish a high performance traffic management system, whereas other studies found density to be the single most important factor to identify the traffic congestion [13, 15, 27]. Fig. 2 shows different traffic density estimation methods, which are mainly categorized into two broad classes based on the used technology: wired and wireless density estimation technologies. Wired density estimation technologies can be further classified into two categories: density estimation using (a) single device, and (b) hybrid devices. Single device, such as loop detectors [15, 27, 28], video cameras [29], and microphones [30] are used to estimate density. Hernandez et al. developed a method for real-time density estimation using the travel times of the re-identified vehicles' predicted from a vehicle re-identification algorithm using inductive loop detectors [15]. The authors found less than 4% mean absolute percentage error in both congested and non-congested condition detection after comparing with video camera data. Tyagi et al. used the cumulative acoustic signal collected by microphones to categorize the vehicular traffic states into the jammed, medium-flow, and free-flow condition based on the difference between the various spectral contents of the noise signals [30]. Bayes classifier was used to categorize the acoustic signal segments, where they found high classification accuracy (almost 95%). Further classification accuracy is gained by using a discriminative support vector machine classifier. However, this method is unable to extract microscopic traffic speed.

Apart from these standalone density estimation devices, several hybrid devices have been used for density estimation. Qiu et al. used the integrated loop detector responses and probe vehicle information to assess density for a particular freeway segment at a predefined time interval [13]. The results showed that in comparison to using only detector data (almost 40% error), the hybrid method is more accurate (4% error) in estimating traffic density. In another study, chase-car data was combined with loop detector data to determine the LOS of freeway segments in California [16]. Speed data collected by chase cars (a chase car is an instrumented vehicle that records the distance between the vehicle itself and another target vehicle representing normal driving behaviors in the study area, which is then used to determine speed and acceleration data for each second for the target vehicle [31, 32]) for more than 37 hours covering 250 freeway sections in Los Angeles were used in this study. After matching with the data collected by a loop detector (i.e., speed, count, occupancies, etc.), data is aggregated for each segment for every 15 mins to compute segment density. Anand et al. used digital videos and GPS-equipped probe vehicles to collect flow data and travel time data

respectively [33]. Here, the mean absolute percentage error of the prepared model is varied from 0.9 to 15.5%.

Emerging CVT enabled vehicles uses wireless communication to collect and transfer the traffic data to surrounding vehicles and infrastructures (e.g., traffic signal, RSU). Usability of traffic data collection using V2V and V2I based wireless communication for density estimation has been studied in several studies. Beacon messages received from CVs (i.e., using V2V) and from RSUs (i.e., using V2I) and roadmap topology features were used by Barrachina et al. to estimate the density in a study [34]. In this study, the ns-2 simulator was used to study the traffic conditions in different cities. From the simulated data, a mathematical relationship was derived, which resulted in 1.02% average relative error for V2V-based density estimation. For V2I-bassed density estimation, the average relative error was 3.04%. This algorithm depends on the examination of the corridor maps to derive the street-junction ratio, which was a variable in the regression analysis model used by authors to estimate density. Barrachina et al. concluded that combining both V2V and V2I would give higher accuracy. In another study, Venkata et al. found vehicle clustering to be an effective approach for density estimation using both V2V and V2I communication [35]. In this study, the authors developed an algorithm to estimate density with the vehicle clusters. The proposed algorithm was validated in a simulation environment. Analysis revealed that the algorithm accurately estimated the density in different roadway traffic speeds. Caceres et al. used numerous cellular phones acting as probes to capture traffic volume [36]. To approximate the vehicle number moving from one area (i.e., service area under the coverage of a set of base stations) to other, the authors used the anonymous call phone data. After comparing with loop detector data, the experimental results showed an absolute relative error of 17%. Also, the developed method can be used for non-real-time estimations.

For real-time traffic management, DOTs need to collect accurate data across the state. Many state DOTs use loop detectors to collect traffic data. Washington DOT uses loop detectors embedded in the pavement of the state highways to collect traffic data [37]. South Carolina DOT utilizes side-fire microwave speed detectors as well as automatic traffic recorders; whereas, Ohio DOT has loop detectors to record speed [38, 39]. Also, 15 State DOTs collect real-time data from private companies such as INRIX, Tom-tom, and SpeedInfo [40]. These DOTs allocate a major share of their budget to buy data services from private companies. In connected vehicle environment, the integrated CVT-AI based density estimation method can be used by DOTs to get the reliable and accurate real-time traffic condition information [41].

B. Application of Artificial Intelligence in Traffic Operational Analysis

Different algorithms have been developed for traffic operation analysis, including embedded algorithms in loop detector systems (e.g., the vehicle re-identification algorithm and the direct calculation of the density value from the traffic parameter recorded by the detector), the computer vision-based algorithm, and AI [15, 38, 42-44]. Among these algorithms, AI-based
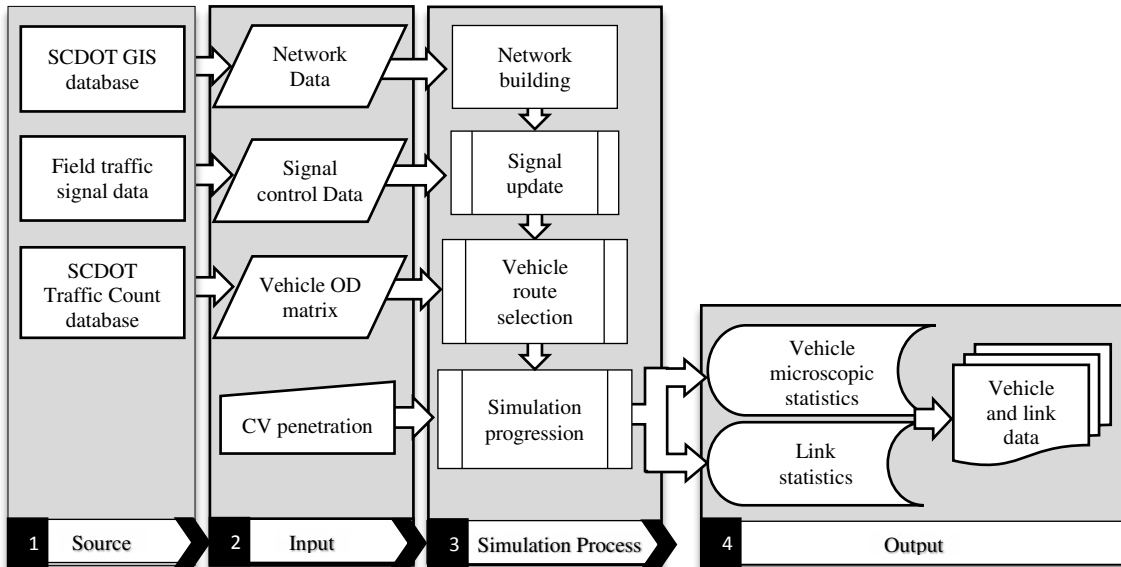
Fig. 3 Simulation method of CV environment

algorithms have the potential to increase the efficacy of the system to assess the density over time due to their inherent capability to adapt to changing traffic conditions. Among different AI paradigms, case-based reasoning (CBR) uses decisions from relevant past experiences to resolve new problems; thus, CBR mimics human behavior to resolve new cases [45]. The key assumption behind finding the solution to a problem is that similar problems have similar answers, and based on this hypothesis, CBR finds the closest or nearest earlier event, and proposes a result. CBR algorithms can work for challenging problems like noisy data and poor similarity functions [46]. Previous studies regarding the application of CBR in transportation engineering problems include collision analysis, real-time traffic diversion during incidents, and traffic control measures' impact evaluation [47-50]. Another widely used AI technique is the support vector machine (SVM), which represents a collection of supervised learning methods that are used to classify (e.g., one-class SVM for one-class classification, and support vector classification or SVC for multi-class classification), or to conduct regression analysis (Support Vector Regression or SVR). SVM achieves nonlinear classification by creating a hyper plane after mapping the training data onto a richer kernel-induced feature space. Moreover, only the cases determining the support vectors require computational cost, which signifies that the SVM is a

computationally efficient. Previous study revealed that SVM has higher prediction capability and learning potential compared against other AI-based paradigms, such as artificial neural network [51]. SVM is applied to predict traffic parameters (e.g., speed, travel time, and traffic flow) and to detect incident [6, 48, 51, 52]. For this research, the feasibility of both AI paradigms (CBR and multi-class SVC) to assess the real-time operational condition of a freeway is evaluated.

III. METHOD

In order to accomplish the research objectives, traffic microsimulation software (i.e., VISSIM) is used to create an AI-incorporated connected vehicle environment for the real-time freeway state (i.e., LOS) estimation, and to compare the CVT-AI method performance with the PeMs density estimation algorithm. This section discusses the basic assumptions and associated research steps.

A. Basic Assumptions for CV Simulation Environment

For the I-26 simulation network, it is assumed that 7 RSUs, equipped with a microprocessor and wireless interfaces are placed along the freeway. Previous studies estimated that a reliable data transmission for V2I communication is possible by placing the RSUs at regular intervals [6], [53]. It is assumed that the CVs communicate with the RSU directly, or relay to the RSU through CVs within RSU's communication range. After collecting microscopic traffic data such as headway, acceleration, speed, number of stops and vehicle maneuver data like lane-changing behavior using vision-based corridor lane marking detection and the onboard camera with time stamp information, CVs can transmit the information to the RSU without any communication latency. Once the data is collected by the RSU, it analyzes the data with the AI technique to estimate traffic density and LOS. Density values corresponding to each LOS, as well as the class label is shown in the following Table 1.
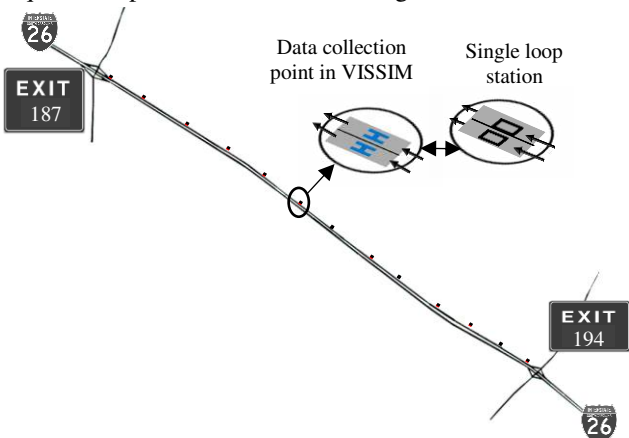

Fig. 4 Data collection point locations for PeMS along I-26

Table 1
Label for different Level of Service (LOS) [54]

LOS	Density (passenger car/mi/ln)	Label
A	<=11	1
B	>11 and <18	2
C	>18 and <26	3
D	>26 and <35	4
E	>35 and <45	5
F	>45	6

In order to get the headway value, it is assumed that the vehicle location data is provided with a combination of the corridor name and the mile marker information, which can be found using the onboard GPS units and geographic information system (GIS) database.

B. Development of Simulation Network

In order to develop the simulation network in VISSIM, the macroscopic software VISUM is used to initially develop the geometric model using the shape file obtained from the SCDOT GIS database warehouse. Once the network is modified according to the field data of the roadway geometry, the network is then exported to VISSIM simulation software. In order to derive the origin-destination matrix, each gateway point Annual Average Daily Traffic (AADT) value is converted to Directional Design Hourly Volume using the 2012 AADT

data [55]. Dynamic Traffic Assignment (DTA) is used in the network to get turning volumes at each interchange. The network is calibrated in an iterative process with the field measured travel time and volume data, link cost (in simulation), vehicle speed distribution (in simulation), and driver behavior parameters (in simulation) to match the volumes and travel times observed on the site. The model is considered calibrated once the volumes and travel times are found to be within 10 percent of the field values. All dynamic routes are converted to static routes using the final DTA paths and cost files. Finally, the calibrated model is used to generate different LOS categories (i.e., LOS A-F categories). Fig. 3 shows the steps for the simulation model development and case generation for a CV environment. For each penetration level of CVs, a total of 50 cases are developed with different seed numbers. Among them, 20 cases are taken to be the training files, whereas 30 cases are considered as the test files. Each case represents a different combination for different LOS categories, where each LOS category has a one hour of representative data, and LOS was estimated for one second time interval. Among the one hour data, initial 1,200 seconds were considered as the simulation warm-up period. Vehicle volume is adjusted in order to simulate different LOS events based on the density values compared to the base scenario. The following Table 2 shows a sample vehicle generated dataset from microscopic simulator VISSIM. In this example, for a 2 second time interval, one

(a) Density estimation with CVT-SVM method

(b) Density estimation with CVT-CBR method

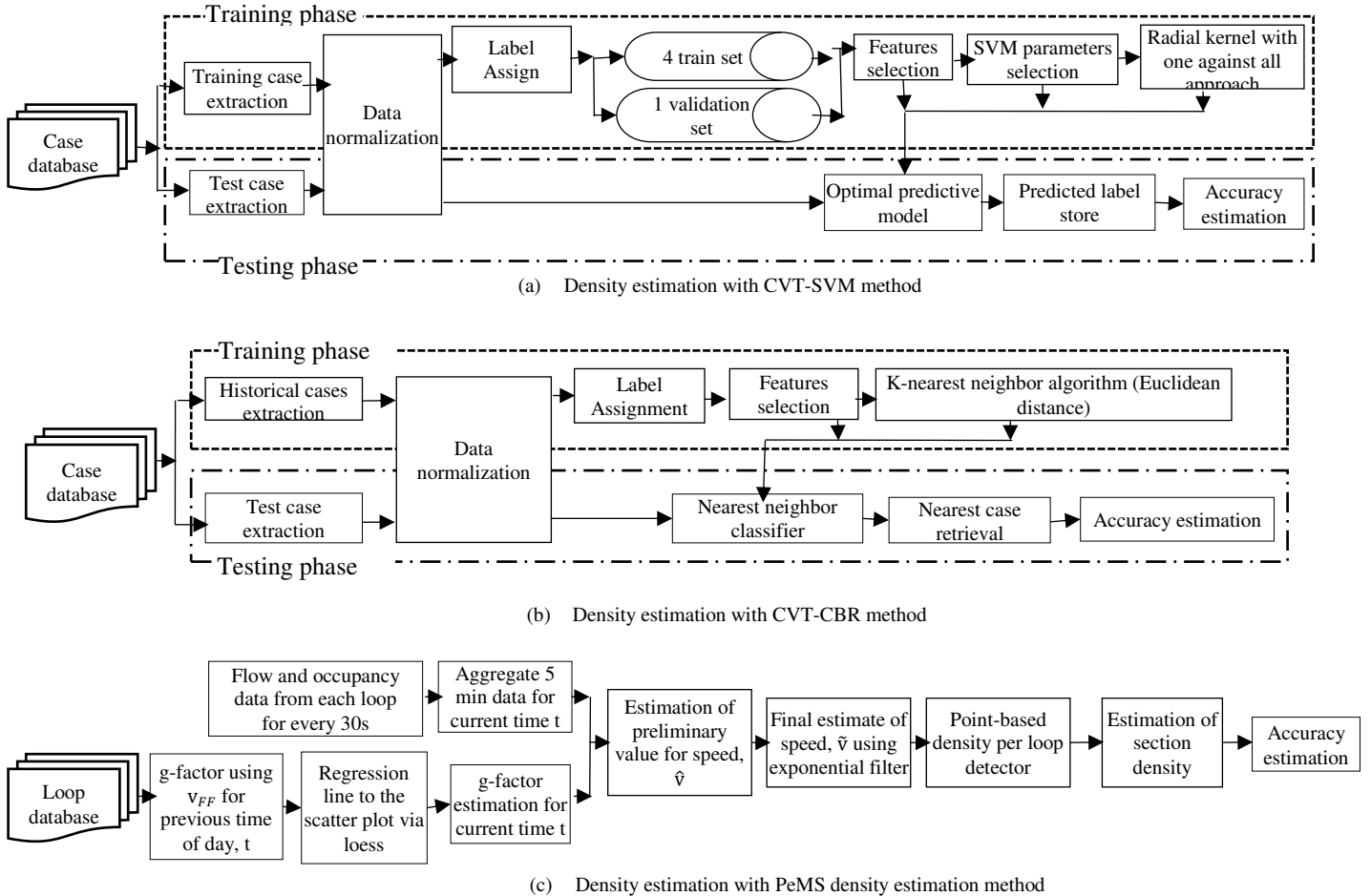(c) Density estimation with PeMS density estimation method

Fig. 5 Density estimation using CVT-AI method and PeMS density estimation algorithm

vehicle (vehicle ID. 8) on a particular corridor (the study corridor, Link: 181) generated data for different variables: time, headway (ft.), lane change status, number of stops, acceleration (ft./s²) and speed (mph).

TABLE 2
Sample Vehicle Generated Data for 2 Seconds Interval

Time (sec.)	Link	Vehicle ID.	Headway (ft.)	Lane Change	No. of Stop	Acceleration (ft/s²)	Speed (mph)
232	181	8	37375.8	-	0	0.00	77.31
233	181	8	37262.4	-	0	0.00	77.31

The corridor specific dataset includes data for a particular time for all vehicles and all links. The corridor dataset includes variables such as: time (sec), average speed of all vehicles on the link (mph) and density (veh/mi/lane). The Table 3 illustrates a sample corridor specific data for Link 181 between the Exit 187 and Exit 194 on I-26.

TABLE 3
Sample Corridor Specific Data for 2 Seconds Interval

Time (sec.)	Link	Avg. Speed (mph).	Density (Veh/mi/lane)
232	181	74.68	5.65
233	181	74.59	5.73

In real-life, the microscopic data will be collected from CVs and roadside units with Dedicated Short Range Communication (DSRC)-enabled devices. DSRC devices will transmit the basic safety messages (BSMs) [56] , which include the data (e.g., headway, speed, brake system status) required for the CV safety applications. Each RSU will have a processor to process the microscopic vehicle-generated data required for the density estimation. Once the RSUs collect BSMs from all CVs and the data processing task is completed, the freeway LOS will be estimated in the real-life CV environment using the CVT-AI method.

To replicate the PeMS density estimation method using loop detectors, two data collection points placed 6 ft. apart in simulation to represent a 6 ft. long single loop detector, shown in Fig. 4. Two successive data collection points capture the detector occupancy and number of vehicle crossing the loop detector, which is a surrogate measure of a single loop detector system. Following the loop detector spacing in PeMS, the spacing interval between the loop stations is 0.5 mile [57]. As there are two lanes in study area on I-26, a total of 28 data collection points are installed which represents 28 loops in the real world. The raw data collection interval for the PeMS density estimation algorithm is typically 30 seconds which is aggregated to 5-minute average [58]. According to the literature, the summation of the individual 30-second data samples for flow from the loop detectors gives the flow value for a 5-minute interval. For occupancy, the 5-minute value is the average of the 30-second data samples received. A similar 5-minute time interval is considered for the AI-aided CVT system for comparison with the PeMS loop detector based density estimation algorithm. 25 microscopic traffic simulator generated files are used to derive the g-factor for any particular time, and a different set of 25 microscopic traffic simulator files are used to check the accuracy of the PeMS LOS estimation algorithm.

C. Feature Selection

In order to select initial features, the Spearman coefficient value is calculated for each feature to estimate the LOS (i.e., headway, speed, acceleration, number of stops, and vehicle lane-changing maneuver) [59]. Spearmen coefficient can be applied when the data type is ordinal (i.e., the features can be divided in multiple categories and they can be ranked or ordered). In this research, the features associated with the different density values can be classified in six categories (i.e., LOS A to F). Also, the input features are monotonically related to density (which is the indicator of LOS) as (i) when the value of one feature increases (i.e., number of stops, number of lane changing), the freeway density increases; or (ii) when the value of one features decreases (i.e., speed, headway, acceleration), the freeway density increases. The following Equation 1 shows the formula of Spearmen coefficient between each individual feature and LOS [60].

$$\rho = 1 - \frac{6 \sum d_i^2}{n.(n^2-1)} \qquad (1)$$

Where, ρ is the Spearman rank correlation, d_i is the difference between the ranks of input feature and corresponding LOS at a particular time i, and n is the number of data point in each data set. Using the above formula, Spearman coefficient values for LOS and input features (i.e., headway, speed, acceleration, number of stops, and vehicle lane-changing maneuver) are calculated at a 5% CV penetration level. The estimated Spearman coefficient for headway, speed, acceleration, number of stops, and vehicle lane-changing maneuver are 0.8, 0.7, 0.006, 0.4 and 0.2, respectively. When the coefficient value is less than 0.3, it means there exists small association between input feature and LOS [61]. Medium association between the input variable and LOS is observed when the coefficient values are within the range of 0.3 and 0.5. When the coefficient value is more than 0.5, it represents strong association between the input variable and LOS. Based on the calculated value of the Spearmen coefficient, three input features (i.e., headway, speed and number of stops) are considered for further analysis in this study.

D. AI Methods for Traffic Operational Analysis

1) Support Vector Machine (SVM)

Fig. 5 (a) shows a flow chart with the steps of SVM followed in this study. For the multiclass classification problem (LOS A to F), the LIBSVM library tool is used for the C-support vector classification (C-SVC) algorithm with the radial-basis kernel function [62]. Earlier study identified radial basis kernel function generally performs well compared to other kernel functions [63]. For this research, the LOS estimation accuracy achieved with various kernel functions (i.e., linear, polynomial, sigmoid and radial basis kernel functions) are evaluated. It was found that highest LOS estimation accuracy is achieved with the radial basis kernel function. A desktop computer with a 3.4 GHz Intel® Core i7 and 24 GB of RAM is used for running C-SVC algorithm. Various combinations of CV generated data are tested which include combinations of distance headway, speed, acceleration, number of stops, and vehicle lane-changing maneuver data. Using cross validation (with 4 train sets and 1 validation set) the combination of distance headway, number of stops, and speed data are determined that give the highest

accuracy in estimating traffic density. After finding the values of the three input features (i.e, distance headway, number of stops, and speed) from the CV generated data, cost coefficient (C), a parameter to avoid misclassification of the training sets, and the kernel function parameter (γ), a parameter to define the influence area of a single training example, are estimated for SVM classifications. To find the optimal value of C and γ to maximize the classification accuracy, a grid-search method is used with a 10-fold cross validation technique.

2) Case-based Reasoning (CBR)

Fig. 5 (b) outlines the k-nearest neighbor classification method used for CBR analysis. In this study, the Euclidean distance measurement formula (Equation 2, for all instances corresponding to the points in the *n*-dimensional space) is applied to find the closest case for each test case. Once the nearest cases are identified from the search, the LOS is determined for the respective case.

$$\text{Euclidean Distance } (a, b) = \sqrt[2]{\sum_{f=1}^{n} w_f.(a_f - b_f)^2} \qquad (2)$$

where w_f is the weight for each parameter of each case, a_f and b_f are the components of case *a* and *b* [48], [64]. To get the optimal nearest neighbor number and weight values associated with each feature, the k-fold cross validation method was applied. The training data is grouped into k subsets in this validation strategy [65]. In total, trials were conducted for 'k' times. For each scenario, test data contains data from one of the k subsets, and training data is considered as the other k-1 subsets. The average error across all trials is estimated for both scenarios: (i) with different values of nearest neighbors, and (ii) with different weight combinations for three input features (i.e., headway, speed and number of stops). The number of nearest neighbors and weight combinations, which produce the minimum average error, were considered to be the optimal value. Once the optimal values were found, an analysis was conducted on the test cases. For the test cases, LOS was estimated based on the minimum expected classification cost.

E. PeMS Density Estimation Algorithm

For the PeMS density estimation algorithm, 25 VISSIM generated files are used to estimate the g-factor for time interval *t* (i.e., 5 minutes) using the Equation 3 [19]. The g-factor is estimated, when the traffic is experiencing the free-flow condition, as the basic assumption is that vehicles move with a free flow speed when the occupancy is low.

$$g_{factor} = \frac{v_{FF}.\rho(t)}{N(t)} \qquad (3)$$

where v_{FF} is the corridor free flow speed where the detector is placed, $\rho(t)$ and $N(t)$ are the occupancy and flow values for the loop detector for the time interval, *t* respectively. To calculate *g-factor* for the test files, the regression line via loess is used following the PeMS density estimation algorithm, as shown in Fig. 5 (c) [19]. Using the derived g-factor for time interval *t*, detector determines the occupancy, and flow value from the test files. The preliminary speed, $\hat{v}$ is estimated using Equation 4.

$$\hat{v}(t) = \frac{N(t).g_{factor}}{\rho(t)} \qquad (4)$$

Later, the initial estimate of speed is passed through an exponential filter with weights that vary as a function of flow. Based on the exponential filter, the final estimate of speed, $\tilde{v}$ for w(t) is derived using Equation 5. w(t) can be estimated with Equation 6.

$$\tilde{v}(t) = w(t).\hat{v}(t) + (1 - w(t)).\tilde{v}(t - 1) \qquad (5)$$

$$w(t) = \frac{N(t)}{N(t) + C} \qquad (6)$$

After cross validation, the value of the smoothing parameter *C* is derived. Once $\tilde{v}$ is calculated, the density value for particular time interval *t* is assessed using the flow value for that loop detector. This estimated density represents a point- density. For a particular time interval *t*, estimates from all loop detector stations can be averaged together to estimate the freeway section-density [15].

F. Data Formatting and Normalization

For AI based density estimation methods, 20 training cases and 30 test cases are generated using VISSIM simulation software for each CV penetration level. The microscopic traffic simulator VISSIM generated vehicle record file contains different microscopic data. For a certain time stamp, the CV-generated microscopic data (speed, acceleration, number of stops and lane change) are captured from all CVs. Later, for same time stamp, the average values of these parameters are matched with the corresponding density data collected from the simulator link evaluation file. For the training files of the integrated CVT-AI method, the 6 different LOS values are labeled from 1 to 6, where 1 means LOS A, 6 means LOS F, and the values 2, 3, 4, and 5 represent LOS B, C, D, and E, respectively. Similar labels are used for the PeMS density estimation algorithm to examine the accuracy of the algorithm. Literature shows that both AI techniques (SVM and CBR) are highly sensitive to the input data variance and scale, which necessitates the use of data normalization [66, 67]. Once the training data is normalized and used to develop the classifier, test files are normalized using the same scale from 0 to 1, as suggested in the literature [62].

G. Accuracy Estimation of Traffic Operational Assessment

According to a real-time system management information program under Title 23 of the Code of Federal Regulations, state DOTs and other responsible agencies must report real time traffic and travel information at a minimum of 85% accuracy [68, 69]. In this research, 85% accuracy is considered to be the threshold in examining the suitability of different algorithms. For both the CVT-AI methods and the PeMS density estimation method, classification accuracy, A is measured using the Equation 7.

$$A = \left(\frac{s}{n}\right).100\% \qquad (7)$$

where, *s* is the number of accurately classified density events, and *n* is the total event number. In this research, the accuracy is measured with respect to density data from the VISSIM link

Fig. 6 Density estimation accuracy by CVT-AI methods
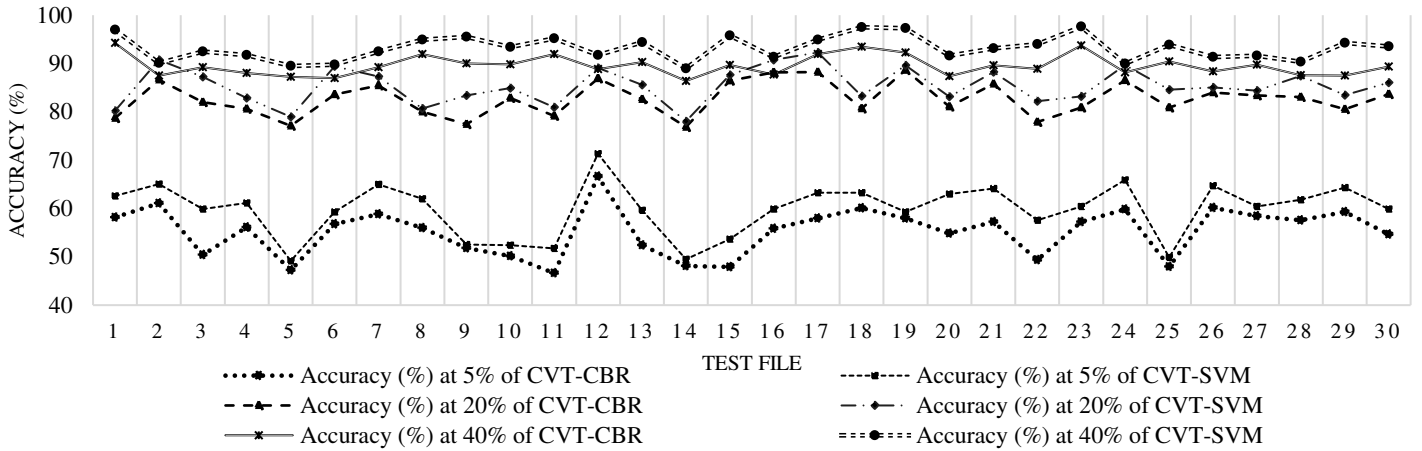
evaluation output, which is considered to be the 'ground truth' value of the density. This link density is the density data for the study corridor (i.e., the simulated I-26 freeway section from Exit 194 to Exit 187).

IV. RESULT AND ANALYSIS

This section is organized into the following subsections to present research findings from the freeway density estimation analysis based on the CVT-AI framework, and the PeMS density estimation algorithm.

A. Parameter adjustment for SVM and CBR

The accurate estimation of the SVM parameters is needed for the accuracy of the multi-class SVM. Using the grid search method, both the cost coefficient, C, and the kernel parameter, γ, are estimated. For each CV penetration level the optimal value of the C and γ parameters are selected depending on the highest cross-validation accuracy with the 20 training cases. Using these values of these SVM parameters, the trained SVM model files are generated to estimate density for the test cases. Similarly, weight parameters and number of nearest neighbors are also estimated for CBR. Table 4 contains the optimal values of the SVM parameters for various CV penetration levels in the I-26 network. Table 5 shows the optimal weight parameters and number of nearest neighbors for CBR.

TABLE 4
Estimated SVM Parameters

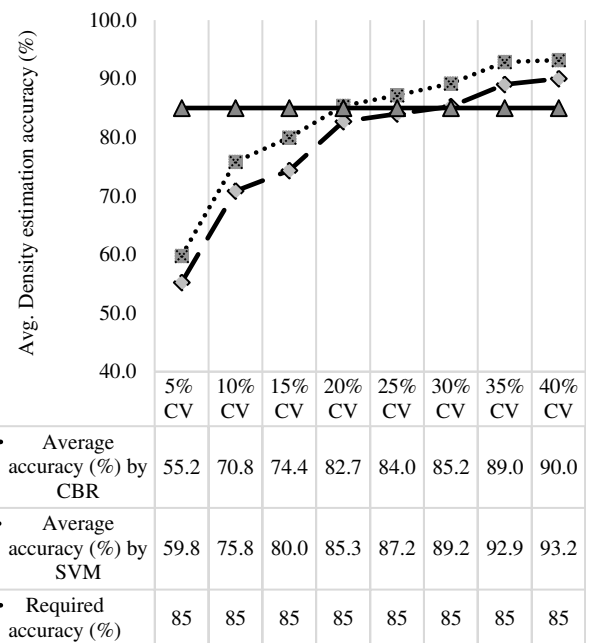| CV penetration (%)	5	10	15	20
C	4096	1024	256	1024
Υ	0.25	0.25	0.025	0.25
CV penetration (%)	25	30	35	40
C	1024	1024	1024	1024
Υ	0.25	0.25	0.25	0.25

B. Comparison of density estimation by CVT-CBR and CVT-SVM method

In MATLAB, the C-support vector classification (C-SVC) method from LibSVM library tool is used to analyze CV test case data with SVM. Fig. 6 shows the accuracy of both AI methods for different penetration levels of CVs (i.e., 5%, 20%

TABLE 5
Estimated CBR Features

CBR Features	Input features for CVT-CBR	CV Penetration (%)							
		5	10	15	20	25	30	35	40
Weight	Headway	0.21	0.79	0.79	0.64	0.64	0.32	0.64	0.03
	Number of Stops	0.32	0.06	0.06	0.22	0.22	0.54	0.22	0.32
	Speed	0.47	0.15	0.15	0.14	0.14	0.14	0.14	0.65
Number of nearest neighbors	All features (i.e., headway, number of stops. speed)	13	15	7	11	11	7	11	5

		5% CV	10% CV	15% CV	20% CV	25% CV	30% CV	35% CV	40% CV
➤	Average accuracy (%) by CBR	55.2	70.8	74.4	82.7	84.0	85.2	89.0	90.0
⊞	Average accuracy (%) by SVM	59.8	75.8	80.0	85.3	87.2	89.2	92.9	93.2
△	Required accuracy (%)	85	85	85	85	85	85	85	85

Fig. 7 Average density estimation accuracy by CVT-AI methods

and 40%) along the I-26 study corridor. From Fig. 6, it is observed that, for the tested penetration level of CVs, SVM

TABLE 6
Level of service (LOS) estimation accuracy by different AI methods

CV penetration (%)-AI method	Accuracy for LOS Events (%)						Average Accuracy for all LOS events (%)
	A	B	C	D	E	F	
5% CV-SVM accuracy (%)	73.5	46.7	46.3	54.1	49.0	87.1	59.8
5% CV-CBR accuracy (%)	56.8	48.6	43.6	47.5	46.9	84.9	55.2
10% CV-SVM accuracy (%)	88.4	65.7	68.8	68.9	68.0	92.0	75.8
10% CV-CBR accuracy (%)	82.7	63.8	62.4	59.9	62.1	90.2	70.8
15% CV-SVM accuracy (%)	88.2	75.3	73.2	74.2	74.1	92.1	80
15% CV-CBR accuracy (%)	82.4	70.3	63.2	67.6	68.1	91.6	74.4
20% CV-SVM accuracy (%)	96.6	82.0	80.2	79.1	76.4	94.7	85.3
20% CV-CBR accuracy (%)	94.4	80.9	75.9	76.9	70.4	93.4	82.7
25% CV-SVM accuracy (%)	96.0	86.3	81.0	82.2	78.9	95.8	87.2
25% CV-CBR accuracy (%)	93.8	82.2	76.1	78.6	74.1	95.7	84
30% CV-SVM accuracy (%)	96.2	86.5	83.3	85.0	87.7	95.6	89.2
30% CV-CBR accuracy (%)	95.2	81.9	77.2	79.7	80.9	94.6	85.2
35% CV-SVM accuracy (%)	98.5	90.8	90.7	89.2	88.9	97.7	92.9
35% CV-CBR accuracy (%)	98.7	90.9	85.9	82.7	85.4	96	89
40% CV-SVM accuracy (%)	99.2	92.6	89.7	89.6	89.5	97.1	93.2
40% CV-CBR accuracy (%)	98.3	90.1	86	85.1	80.2	94.9	90

gives higher accuracy in density estimation over CBR for each test case. The difference between the accuracy of the CBR and SVM estimated density varies up to 18.7% for 5% CV penetration, whereas for the 40% penetration level, it varies up to 7.7%. Fig. 7 shows that the average estimated density accuracy for the CVT-SVM method is lower than the Federal regulation defined threshold (i.e., 85% accurate) for a 15% or less penetration level of CVs, whereas for the CVT-CBR method, it is lower than the threshold for a 25% or less penetration level of CVs. One significant observation is that for each CV penetration level, the average density estimation accuracy for the radial kernel based SVM algorithm is higher than the k-nearest neighbor based CBR algorithm. Moreover, statistical analysis revealed that the density estimation accuracy difference between the CVT-CBR and CVT-SVM methods is significant for each CV penetration level at a 95% confidence level. It is also evident from Fig. 7 that the average density estimation accuracy increases with the increase in CV penetration levels for both AI methods. Also, with increasing penetration levels of CVs, the difference between the average accuracy of CBR and SVM decreases. From the observed trend, it is expected that with higher penetration levels (i.e., more than 40%) of CVs, the accuracy of the density estimation will further increase.

Table 6 shows the detailed density estimation accuracy for different classes of LOS (LOS A to F) for both CBR and SVM. At the lowest penetration level (5%), AI methods cannot estimate LOS B, C, D and E with sufficient accuracy (i.e., minimum 85% accurate) in comparison to higher CV penetration levels. Moreover, SVM performs better than CBR in classifying all LOS events for any penetration level.

C. Computational time requirement for AI methods

As the real-time density estimation application requires instant analysis of the collected data, the computational time needs to be negligible so that it does not delay the decision making process. Table 7 shows the average processing time of both AI techniques in analyzing each test file. In the analysis, individual

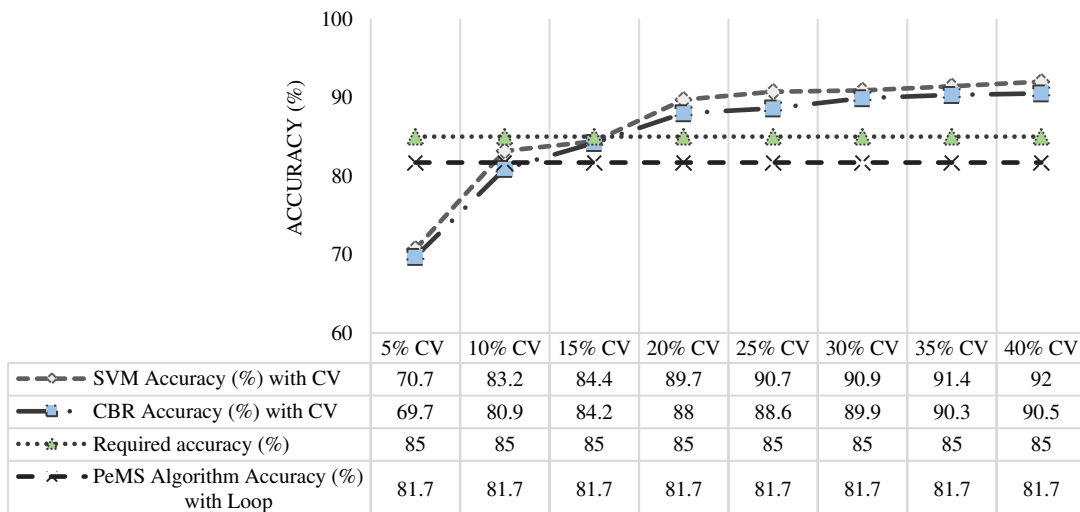|	5% CV	10% CV	15% CV	20% CV	25% CV	30% CV	35% CV	40% CV
SVM Accuracy (%) with CV	70.7	83.2	84.4	89.7	90.7	90.9	91.4	92
CBR Accuracy (%) with CV	69.7	80.9	84.2	88	88.6	89.9	90.3	90.5
Required accuracy (%)	85	85	85	85	85	85	85	85
PeMS Algorithm Accuracy (%) with Loop	81.7	81.7	81.7	81.7	81.7	81.7	81.7	81.7

Fig. 8 LOS estimation accuracy by CVT-AI methods and PeMS algorithm

vehicle generated data is aggregated for one second time interval to assess the study corridor traffic density. The adopted density estimation methods take only few milliseconds to compute the link density, which stays within the tolerable limit for real-time traffic applications [2]. CVT-SVM requires more time compared to the CVT-CBR method. However, the computational time reduces with increasing CV penetration for both CVT-AI methods.

TABLE 7
Average computational time requirement for CVT-AI methods

CV Penetration (%)	Processing time for CBR (sec)	Processing time for SVM (sec)
5	0.17	65.76
10	0.13	52.99
15	0.15	56.55
20	0.12	29.07
25	0.11	25.98
30	0.11	23.97
35	0.13	26.02
40	0.09	14.30

D. Comparison of density estimation by CVT-AI method and PeMS Density Estimation method

When comparing the CVT-AI density estimation method and PeMS density estimation method, a 5-minute time interval is adopted to compute freeway section density. Each method is compared against the ground truth data- link density from VISSIM simulation. It is observed that with the PeMS density estimation algorithm, 81.7% accuracy can be achieved (Fig. 8), which is less than the required accuracy level (85%) according to the real-time system management information program under Title 23 of the Code of Federal Regulations [68, 69].

In this study, the loop detector-based PeMS performance is estimated based on the assumption that all the loop detectors are functioning properly. However, in real-life, loop detectors often malfunction and require frequent maintenance. Earlier, Hernanadez et al. collected loop detector data for a segment of I-405N in Irvine, California and achieved an overall density estimation accuracy of 87% applying PeMS algorithm [15]. The difference in the estimation accuracy of this study with our research could be attributed to the variation in traffic pattern and geometric characteristics of study areas. For example, I-405N corridor used in Hernanadez et al. study is 0.66 mile long, and 7 lane (one-direction) freeway segment including two high-occupancy vehicle (HOV) lanes, where the study corridor used in this study is a 7-mile segment of I-26 with two lanes (each direction).

Findings from the statistical analysis reveal that, at a 95% confidence level the differences between the CVT-AI methods and the PeMS density estimation method accuracy are significant for a 20% or more CV penetration level. One probable cause of the relatively poor performance of the PeMS algorithm compared to CVT-AI method can be the use of a fixed g-factor value for 5-minute periods. A fixed g-factor is not sensitive to the changes in traffic conditions in the freeway section. The changes in traffic conditions could be estimated by the g-factor values for shorter period of time. One important issue here is to ensure the continuous flow of reliable and accurate data from all the loop detectors. Otherwise, similar accuracy cannot be achieved in real-world. For any missing and/or inaccurate data in the PeMS system, several data

imputation methods are available, which are used to replace the loop detector collected bad data in real-world applications. Accuracy of the PeMS density estimation algorithm including the data from these imputation methods are not tested in this paper, since data is available from the simulation network for all the induction loop detectors.

Fig. 9 shows the accuracy of the PeMS algorithm for density estimation for different LOS events compared to SVM and CBR method. The accuracy is almost 96% for LOS A, while it is as low as only 56% for LOS E.
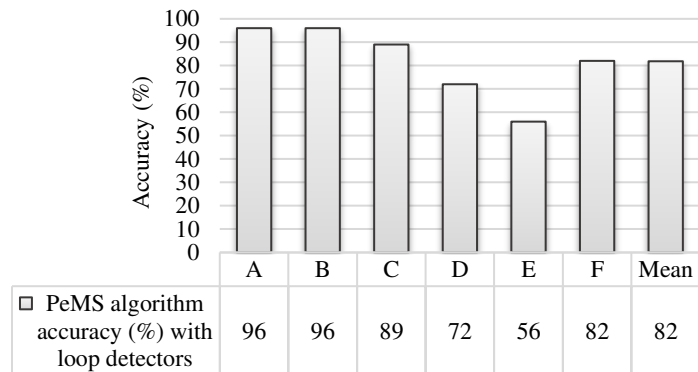

	A	B	C	D	E	F	Mean
PeMS algorithm accuracy (%) with loop detectors	96	96	89	72	56	82	82

Fig. 9 LOS event detection accuracy by PeMS algorithm

V. RESEARCH CONTRIBUTIONS

Prior studies investigated the accuracy of different traffic density estimation algorithms (such as, polynomial and non-polynomial function based density estimation algorithm and vehicle clustering-based density estimation algorithm) using data collected through V2V and V2I connectivity on urban streets. In contrast, this study is the first attempt to study an integrated CVT-AI method to estimate the LOS/traffic density on freeways. In this study, the three CV generated data sets (i.e., distance headway, number of stops, and speed) were identified as significant variables in feature selection step for freeway LOS estimation. Using these data, an increase in LOS estimation accuracy was observed with the gradual increase of CV penetration level applying both AI methods (i.e., CBR and SVM). This study also demonstrated the advantage of integrated CVT-AI method compared to the loop based LOS estimation algorithm used by the PeMS. Earlier studies did not compare CVT-AI based density estimation methods with the loop detector based density estimation methods. It was found that the developed CVT-AI method can provide better density estimation accuracy with a 20% or more CV penetration level compared to the loop detector based LOS estimation algorithm.

VI. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

A. Conclusions

Intelligent transportation systems enabled by CVT will substantially improve safety, mobility, and the environment. This study examined a mobility application, specifically the real-time traffic state assessment using CV data. Among various combinations of microscopic traffic data that can be collected from CVs, space headway, number of stops, and speed were found to be the best parameter combination to predict traffic density with AI methods. CV-based traffic condition assessment strategy provides more than 85%

accuracy with a 20% or more CV penetration level. The radial-kernel based CVT-SVM algorithm consistently provides higher traffic density estimation accuracy compared to the k-nearest neighbor based CVT-CBR algorithm at any CV penetration level. Moreover, both AI methods' estimation accuracy improves with higher CV penetration levels. When comparing to a traditional loop detector based density estimation algorithm (i.e., PeMS algorithm), it is observed that the estimated LOS by the PeMS system is much lower (~80%) compared to true LOS values. However, both CVT-AI methods showed significantly higher LOS estimation accuracy at 20% or more CV penetration levels, which shows the effectiveness of the CVT-AI framework developed in this research over the traditional loop detector based LOS estimation system. While some state DOTs have deployed extensive road sensors to monitor traffic conditions, other DOTs have contracts with private companies for receiving real-time traffic speeds and travel times data. The CVT application evaluated in this research will create a unique opportunity to collect real-time traffic condition data, which will reduce and eliminate the current use of traditional loop detector-based data collection practices or data purchased from private vendors that cost taxpayer dollars. Moreover, data collected from smart-phones can be used to estimate density, which will act as surrogate data when the penetration level of CVs is low. Apart from providing a more accurate alternative to the traditional traffic loop detector, CVT-AI system has the potential to assess traffic conditions where traffic sensors are not installed.

B. Future Research Directions

The real-world field tests are required to assess the reliability, feasibility and performance of the CVT-AI framework, developed in this research, to estimate traffic states. Also, the real-world DSRC communication reliability is needed to be tested. CVs will be on public roads within the next few years, and this offers challenges as well as opportunities for transportation agencies in terms of transportation planning, design, operations, and management. This study did not incorporate non-CV data, such as data from social media and/or news feeds. Future studies to estimate LOS should be done incorporating these non-CV data sources. Data collected in the CV environment will also permit the incorporation of non-real time applications, for example transportation planning.

REFERENCES

[1] Wang, F. Y. (2010). Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications. Intelligent Transportation Systems, IEEE Transactions on, 11(3)

[2] Ma, Y., Chowdhury, M., Jeihani, M., & Fries, R. (2010). Accelerated incident detection across transportation networks using vehicle kinetics and support vector machine in cooperation with infrastructure agents. IET intelligent transport systems, 4(4), 328-337.

[3] Claes, R., Holvoet, T., & Weyns, D. (2011). A decentralized approach for anticipatory vehicle routing using delegate multiagent systems. Intelligent Transportation Systems, IEEE Transactions on, 12(2), 364-373.

[4] Cheu, R. L., Qi, H., and Lee., D. H. (2002). "Mobile Sensor and sample-based algorithm for freeway incident detection." Transportation Research Record, 1811, Transportation Research Board, Washington D.C., 12-20.

[5] Crabtree, J. D., & Stamatiadis, N. (2007). Using dedicated short-range communication (DSRC) technology for freeway incident detection: A performance assessment based on traffic simulation data. In 87th Annual. Meeting Transportation Research Board (pp. 13-17).

[6] Ma, Y., Chowdhury, M., Sadek, A., & Jeihani, M. (2009). Real-time highway traffic condition assessment framework using vehicle–infrastructure integration (VII) with artificial intelligence (AI). IEEE Transactions on Intelligent Transportation Systems, 10(4), 615-627.

[7] Qi, H., Cheu, R. L., & Lee, D. H. (2002). Freeway incident detection using kinematic data from probe vehicles. In 9th World Congress on Intelligent Transport Systems.

[8] He, Y., Chowdhury, M., Ma, Y. and Pisu, P., (2012). Merging mobility and energy vision with hybrid electric vehicles and vehicle infrastructure integration. Energy Policy, 41, 599-609.

[9] Cregger, J., Brugeman, V.S., and Wallace, R. (2013). International Survey of Best Practices in Connected and Automated Vehicle Technologies: 2013 Update.

[10] MDOT (2011). Using Data to Improve Safety, Manage Traffic and Manage Assets. Newsletter available at: http://www.michigan.gov/documents/mdot/MDOT_ORBP_Newsletter_Oct_2 011_365778_7.pdf

[11] USDOT (2015) Connected Vehicle Research in the United States. Available at: http://www.its.dot.gov/connected_vehicle/connected_vehicle_research.htm

[12] VDTO (2015). Vehicle Data for Traffic Operations. Available at: http://www.iteris.com/cvria/html/applications/app87.html#tab-3

[13] Qiu, T. Z., Lu, X. Y., Chow, A. H., & Shladover, S. E. (2010). Estimation of freeway traffic density with loop detector and probe vehicle data. Transportation Research Record: Journal of the Transportation Research Board, 2178(1), 21-29.

[14] HCM (2010). Highway Capacity Manual. Transportation research board. National Research Council, Washington, DC.

[15] Hernandez, S., Tok, A., & Ritchie, S.G. (2013). UCI - ITS - WP - 13 - 4 Density Estimation using Inductive Loop Signature based Vehicle Re-identification and Classification. Available at: http://www.its.uci.edu/its/publications/papers/ITS/UCI-ITS-WP-13-4.pdf

[16] Choe, T., Skabardonis, A., & Varaiya, P. (2002). Freeway performance measurement system: operational analysis tool. Transportation Research Record: Journal of the Transportation Research Board, 1811(1), 67-75.

[17] Mao, R., & Mao, G. (2013). Road traffic density estimation in vehicular networks. In IEEE Wireless Communications and Networking Conference (WCNC), pp. 4653-4658.

[18] Performance Measurement System (PeMS) website. Available at: http://pems.dot.ca.gov.

[19] E. V. Zwet, C. Chen, Z. Jia, and J. Kwon (2003) "Statistical methods for estimating speed using single-loop detectors." Available at: https://drive.google.com/file/d/0B5wZ4dLpgONnT1NxbG9SQ21rR1k/view?u sp=sharing

[20] P. Varaiya "Creating transportation system intelligence using PeMS" Available at: http://cvrr.ucsd.edu/events/PeMS-Pravin.pdf

[21] Performance Measurement System (PeMS) % Observed Website. Available at: http://pems.dot.ca.gov/?report_form=1&dnode=State&content=fidelity&tab=fi d_compare&export=&s_time_id=1420156800&s_mm=1&s_dd=2&s_yy=201 5&e_time_id=1451692799&e_mm=1&e_dd=1&e_yy=2016&gn=week&html .x=51&html.y=13

[22] Performance Measurement System (PeMS) Level of Service website. Available at: http://pems.dot.ca.gov/?dnode=State&content=los.

[23] Bhavsar, P., Chowdhury, M., Sadek, A., Sarasua, W., & Ogle, J. (2008). Decision support system for predicting traffic diversion impact across transportation networks using support vector regression. Transportation Research Record: Journal of the Transportation Research Board, 2024(1), 100-106.

[24] Shi, W., & Liu, Y. (2010). Real-time urban traffic monitoring with global positioning system-equipped vehicles. IET intelligent transport systems, 4(2), 113-120.

[25] Tao, S., Manolopoulos, V., Rodriguez, S., & Rusu, A. (2012). Real-time urban traffic state estimation with A-GPS mobile phones as probes. Journal of Transportation Technologies, 2(01), 22.

[26] Seo, T., Kusakabe, T., & Asakura, Y. (2013). Traffic State Estimation Method Using Probe Vehicles Equipped With Spacing Measurement System. In International Symposium on Recent Advances in Transport Modelling.

[27] Ni, D. (2007) Determining Traffic-Flow Characteristics by Definition for Application in ITS. IEEE Transactions on Intelligent Transportation Systems, 8(2), 181-187.

[28] Ramezani, A., Moshiri, B., Abdulhai, B., & Kian, A. R. (2011). Estimation of free flow speed and critical density in a segmented freeway using missing data and Monte Carlo-based expectation maximisation algorithm. IET control theory & applications, 5(1), 123-130.

[29] Ajitha, T., Vanajakshi, L., & Subramanian, S. C. (2013). Real-Time Traffic Density Estimation without Reliable Side Road Data. Journal of Computing in Civil Engineering.

[30] Tyagi, V., Kalyanaraman, S., & Krishnapuram, R. (2012). Vehicular traffic density state estimation based on cumulative road acoustics. Intelligent Transportation Systems, IEEE Transactions on, 13(3), 1156-1166.

[31] Yu, L, Wang, Z.and Shi, Q (2010). PEMS-Based Approach to Developing and Evaluating Driving Cycles for Air Quality Assessment. No. SWUTC/10/169300-1. Center for Transportation Training and Research, Texas Southern University. Available at: http://static.tti.tamu.edu/swutc.tamu.edu/publications/technicalreports/169300-1.pdf

[32] Klauer, S.G., Guo, F., Sudweeks, J. and Dingus, T.A., (2010). An analysis of driver inattention using a case-crossover approach on 100-car data: Final report (No. HS-811 334). Available at: http://www.nhtsa.gov/DOT/NHTSA/NVS/Crash%20Avoidance/Technical%20Publications/2010/811334.pdf

[33] Anand, A., Ramadurai, G., & Vanajakshi, L. (2014). Data fusion-based traffic density estimation and prediction. Journal of Intelligent Transportation Systems, 18(4), 367-378.

[34] Barrachina, J., Sanguesa, J. A., Fogue, M., Garrido, P., Martinez, F. J., Cano, J. C., Calafate, C. T. & Manzoni, P. (2013). V2X-d: a Vehicular Density Estimation System that combines V2V and V2I Communications. In IEEE Wireless Days (WD), 2013 IFIP (pp. 1-6).

[35] Venkata, M. D., Pai, M. M., Pai, R. M., & Mouzna, J. (2011). Traffic monitoring and routing in VANETs—A cluster based approach. In ITS Telecommunications (ITST), 2011 11th IEEE International Conference on (pp. 27-32).

[36] Caceres, N., Romero, L. M., Benitez, F. G., & Del Castillo, J. M. (2012). Traffic flow estimation models using cellular phone data. IEEE Transactions on Intelligent Transportation Systems, 13(3), 1430-1441.

[37] Bremmer, D., Cotton, K. C., Cotey, D., Prestrud, C. E., & Westby, G. (2004). Measuring congestion: Learning from operational data. Transportation Research Record: Journal of the Transportation Research Board, 1895(1), 188-196.

[38] Coifman, B., & Kim, S. (2013). Assessing the Performance of the SpeedInfo Sensor (No. FHWA/OH-2013/21).

[39] SCDOT (2006). South Carolina Department of Transportation Annual Accountability Report Fiscal Year 2005-2006, Available at: http://www.scstatehouse.gov/archives/aar2006/u12.pdf

[40] UPWP (2011). Private Probe Vehicle Data for Real - Time Applications: FINAL REPORT. Available at: https://www.azmag.gov/Documents/ITS_2011-10-27_Private-Probe-Vehicle-Data-for-RealTime-Applications-Final-Report.pdf

[41] USDOT (2014) Connected Vehicle Research. Available at: http://www.its.dot.gov/connected_vehicle/connected_vehicle.htm

[42] Ma, Y., Chowdhury, M., Sadek, A., & Jeihani, M. (2012). Integrated traffic and communication performance evaluation of an intelligent vehicle infrastructure integration (VII) system for online travel-time prediction. IEEE Transactions on Intelligent Transportation Systems , 13(3), 1369-1382.

[43] Mossi, J. M., Albiol, A., & Ornedo, V. N. (2011). Real-time traffic analysis at night-time. 18th IEEE International Conference on Image Processing (ICIP), pp. 2941-2944.

[44] Rong, Y. U., Guoxiang, W. A. N. G., Zheng, J., & Haiyan, W. A. N. G. (2013). Urban Road Traffic Condition Pattern Recognition Based on Support Vector Machine. Journal of Transportation Systems Engineering and Information Technology, 13(1), 130-136.

[45] Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI communications, 7(1), 39-59.

[46] Aha, D. W. (1991). Case-based learning algorithms. In Proceedings of the 1991 DARPA Case-Based Reasoning Workshop. Vol. 1, pp. 147-158.

[47] Andersen, O. J. (2012). How Case-based Reasoning can be used to predict and improve Traffic Flow in Urban Intersections. Thesis for Master of Science in Informatics, Norwegian University of Science and Technology. Available at: http://www.diva-portal.org/smash/get/diva2:621991/FULLTEXT01.pdf

[48] Chowdhury, M., Sadek, A., Ma, Y., Kanhere, N., & Bhavsar, P. (2006). Applications of artificial intelligence paradigms to decision support in real-time traffic management. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1968(1), 92-98.

[49] De Schutter, B., Hoogendoorn, S. P., Schuurman, H., & Stramigioli, S. (2003) A multi-agent case-based traffic control scenario evaluation system. In Proceedings of the IEEE Intelligent Transportation Systems, Vol. 1, pp. 678-683.

[50] Li, K., & Waters, N. M. (2005). Transportation networks, case-based reasoning and traffic collision analysis: A methodology for the 21st century. In Methods and Models in Transport and Telecommunications Springer Berlin Heidelberg. pp. 63-92.

[51] Vanajakshi, L., & Rilett, L. R. (2004). A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. In IEEE Intelligent Vehicles Symposium, pp. 194-199.

[52] Wu, Ting-Fan, Chih-Jen Lin, and Ruby C. Weng. (2004) "Probability estimates for multi-class classification by pairwise coupling." The Journal of Machine Learning Research 5: 975-1005.

[53] Co-pilot (2015). Cost Overview for Planning Ideas and Logical Organization Tool (CO-PILOT): Overview and User's Guide. Available at: https://co-pilot.noblis.org/CVP_CET/resources/noblis/csv/CO-PILOT_UsersGuide_1_15.pptx

[54] HCM (2010). Highway Capacity Manual. Transportation research board. National Research Council, Washington, DC.

[55] SCDOT (2015), SCDOT Traffic Count 2013, Available at: http://scdot.maps.arcgis.com/apps/OnePane/basicviewer/index.html?appid=d3a7078374ac4f28a201b7ce3b03eb8c

[56] Cronin, B. (2012). Vehicle Based Data and Availability. Available at: http://www.its.dot.gov/itspac/october2012/pdf/data_availability.pdf

[57] Varaiya, P. (2001). Freeway performance measurement system, PeMS V3, phase 1: final report. California Partners for Advanced Transit and Highways (PATH). Available at: http://escholarship.org/uc/item/20p1j2w7

[58] Performance Measurement System (PeMS) System Calculations website. Available at: http://pems.dot.ca.gov/?dnode=Help&content=help_calc.

[59] A. Saltelli, K. Chan, E. M. Scott, "Sensitivity Analysis" Vol. 1. New York: Wiley, 2000.

[60] Spearman's Rank-Order Correlation (2016). Available at: https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php

[61] Statistics Solution (2016), "Correlation (Pearson, Kendall, Spearman)" Available at: http://www.statisticssolutions.com/correlation-pearson-kendall-spearman/

[62] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 27.

[63] Vanschoenwinkel, B. and Manderick, B. (2005). Appropriate kernel functions for support vector machine learning with sequences of symbolic data. In Deterministic and statistical methods in machine learning, pp. 256-280, Springer Berlin Heidelberg.

[64] Mitchell, T.M., (1997). Machine learning. WCB, Available at: http://personal.disco.unimib.it/Vanneschi/McGrawHill_-_Machine_Learning_-Tom_Mitchell.pdf

[65] Schneider, J. (1997). Cross Validation. Available at: https://www.cs.cmu.edu/~schneide/tut5/node42.html

[66] Attig, A., & Perner, P. (2011). The Problem of Normalization and a Normalized Similarity Measure by Online Data. Tran. CBR, 4(1), 3-17.

[67] Graf, A. B., Smola, A. J., & Borer, S. (2003). Classification in a normalized feature space using support vector machines. IEEE Transactions on Neural Networks, 14(3), 597-605.

[68] FHWA (2010). Data Quality White Paper. Available at: http://ops.fhwa.dot.gov/publications/fhwahop08038/04infoshare.htm

[69] FHWA (2013). Real-Time System Management Information Program [1201] Fact Sheet. Available at: http://ops.fhwa.dot.gov/1201/factsheet/

Sakib Mahmud Khan is a Ph.D student at Clemson University, Clemson, SC, USA. He is conducting research on Intelligent Transport System with a focus on Connected Vehicle Technology.

Kakan C. Dey is an assistant professor at West Virginia University (WVU), and Director of Connected and Automated Transportation Systems (CATS) Lab. He received his Ph.D. in Civil Engineering with Transportation Systems major in 2014 from Clemson University, SC. He is a member of IEEE.

Mashrur Chowdhury received the Ph.D. degree in civil engineering from the University of Virginia, Charlottesville. He is the Eugene Douglas Mays Professor of Transportation Engineering in Clemson University, SC. He is the co-director of Complex Systems Analytics and Visualization Institute (CSAVI) at Clemson University. Dr. Chowdhury is a Registered Professional Engineer in Ohio, USA. He is an Associate Editor for the IEEE Transactions on Intelligent Transportation Systems and Journal of Intelligent Transportation Systems