Data-Driven Approach for Modeling the Nonlane-Based Mixed Traffic Conditions

Raju, Narayana; Arkatkar, Shriniwas S.; Easa, Said; Joshi, Gaurang

DOI
10.1155/2022/6482326

Publication date
2022

Document Version
Final published version

Published in
Journal of Advanced Transportation

Citation (APA)
Raju, N., Arkatkar, S. S., Easa, S., & Joshi, G. (2022). Data-Driven Approach for Modeling the Nonlane-Based Mixed Traffic Conditions. *Journal of Advanced Transportation*, *2022*, 1-16. Article 6482326. https://doi.org/10.1155/2022/6482326

Hindawi
Journal of Advanced Transportation
Volume 2022, Article ID 6482326, 16 pages
https://doi.org/10.1155/2022/6482326

Research Article

Data-Driven Approach for Modeling the Nonlane-Based Mixed Traffic Conditions

Narayana Raju,[1] **Shriniwas S. Arkatkar**,[2] **Said Easa**,[3] **and Gaurang Joshi**[4]

[1]*Transport & Planning Department, CiTG, Delft University of Technology (TU Delft), Delft, Netherlands*
[2]*Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat 395007, India*
[3]*Department of Civil Engineering, Ryerson University, Toronto, Canada*
[4]*Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat 395007, India*

Correspondence should be addressed to Shriniwas S. Arkatkar; sarkatkar@ced.svnit.ac.in

Received 21 October 2021; Accepted 10 March 2022; Published 11 April 2022

Academic Editor: Nagendra R Velaga

The diverse nature of vehicle categories and the resultant lane discipline in mixed (heterogeneous) traffic cause complex spatial interactions. As a result, the driving behavior process in mixed traffic conditions is meaningfully different, where both longitudinal and lateral movements of the vehicles continuously occur. Under prevailing homogeneous traffic conditions in developed countries, driving behavior is partially discrete, where following longitudinal behavior and outboard lane-change models can model traffic behavior. However, the established car-following and lane-change models cannot be directly used in shaping mixed traffic conditions. Such conditions also warrant the use of high-quality microlevel vehicular trajectory data. Accordingly, realizing this need, vehicular trajectory data for different traffic flow conditions were developed. The data were used to extract the parameters required for modeling the vehicles' positions using machine learning algorithms. Three established supervised machine learning algorithms (k-NN, random forest, and regression tree) and deep learning are selected to model mixed traffic conditions. The parameters which influence longitudinal and lateral movements are identified using Spearman correlation analysis. Furthermore, simulation runs are performed using the python language. The performance of the algorithms is evaluated both at the microscopic and macroscopic levels using relevant traffic indicators. The results show that a deep learning model and k-NN tend to replicate better-mixed traffic conditions than random forest and regression trees.

1. Background

Understanding the traffic performance of the road section is vital for effective utilization. In this direction, traffic flow modeling concepts have proven to be an efficient source in gauging network elements' performance at microscopic and macroscopic levels. Since its inception, researchers conceptualized numerous concepts for traffic flow modeling for understanding the performance of road networks. Primarily, it includes car-following models, such as Pipes [1], General Motors [2], Gipps [3], and Wiedemann [4], among others. To a certain extent, these framed concepts produced a satisfactory performance and can replicate the traffic characteristics.

Further, in this direction, researchers sensed the importance of driving behavior in modeling the traffic, for which significant efforts were put in exploring human factors in developing car-following models, such as weather conditions [5], drivers perspective [6], fatigue driving [7], anticipation [8], among others. By incorporating the human factors and the car's performance, the following models is found to increase. On the other side, researchers also understood the importance of lateral behavior of vehicles in traffic streams. Numerous lane-changing models are conceptualized for capturing the lateral behavior, including research studies [9–11], and are able to model the lane-changing movement of the vehicles. With advancements in technology and the availability of computational tools, numerous traffic microsimulation tools [12] are developed by embedding the car-following and lane-changing models for different road geometry and vehicular characteristics.

This includes a few examples, such as PTV VISSIM [13], AIMSUN [14], PARAMICS [15], and SUMO [16]. It is well established that these microsimulation tools boosted the traffic modeling studies to an exceptional level and facilitated a more considerable extent in uncovering numerous concepts. With the development of NGSIM [17], high-quality trajectory data and driver behavior aspects have further strengthened traffic flow modeling for evaluating policy interventions more comprehensively. Further, in recent times, researchers [18–20] tested the deep learning and reinforcement learning strategies for improving the mixed traffic flow efficiency levels.

It can also be noted that most of the literature mentioned above entirely belongs to a homogeneous traffic environment with lane-based traffic conditions. On the other hand, in the case of mixed traffic conditions with poor lane discipline, traffic streams can result in complex spatial interactions among the vehicles in longitudinal and transverse directions. In this direction, very few studies have been carried out to assess different traffic characteristics, especially driving behavior (vehicle-dependent). Furthermore, researchers applied the concepts mentioned above under mixed traffic conditions, which includes studies such as microsimulation [21], modeling traffic flow on expressways [22], calibration of car-following models [23], to name a few. To a certain extent, these strategies can perform better at the macroscopic level, but their performance at the microscopic level (vehicle-vehicle interaction) is questionable. Many studies [24, 25] conducted under mixed traffic conditions reported the predominant lateral movement of vehicles.

Mainly in case of mixed traffic conditions, with poor lane discipline due to the involvement of different vehicle categories, both longitudinal and lateral movements of vehicles can be observed simultaneously, whereas under homogeneous traffic conditions, lane-changing maneuvers are discrete. This significant difference in driving behavior can be attributed as one of the main reasons for the limited performance of established homogeneous traffic concepts in mixed traffic conditions. Under mixed traffic conditions, due to the predominance of lateral movement factor (driving behavior), the subject vehicle movements are influenced by the presence of surrounding vehicles. As a result, numerous parameters that are not accounted for in following traditional behavior and lane-changing models impact the vehicles' movement. Furthermore, in recent times, researchers [26, 27] highlighted the intricacies of nonlane-based mixed traffic conditions. Given the variation in physical properties, vehicles by virtue of their size acquire any available space that can impact the driving behavior of the surrounding vehicles.

From the literature [28–30], it is inferred that machine learning tools are proven to be productive in understanding complex data patterns supported by quality data. Yu et al. [31] tested the Fixed Radius Near Neighbors (FRNN) for modeling the longitudinal car-following behavior and strongly advocated the usage of data driven approaches in traffic modeling. After identifying the research gaps, in the present work, it was decided to explore the performance of machine learning algorithms to address the need for modeling mixed traffic flow both at microscopic and macroscopic levels. Hence, this study is focused on employing machine learning tools from the branch of artificial intelligence in modeling this complex vehicular behavior using microlevel trajectory data. Finally, the performance of the selected algorithms is evaluated thoroughly at different levels, including microscopic and macroscopic level comparisons. Further, in recent times, researchers [32] are strongly advocating the importance of producing the reproducible research in transportation engineering. In this direction, the trained models in this study can be effectively used/improved with other data sources and supports reproducible research in the domain of traffic modeling.

Given the limitations of the traditional following and lane-change models, this paper aims to model mixed traffic conditions with machine learning algorithms.

The study consists of the following main tasks:

(i) Develop vehicular trajectory data to capture the study section's driving behavior using a semiautomated image processing tool

(ii) Explain mixed traffic conditions using an example and mark the surrounding vehicles, which can influence the vehicles' movements

(iii) Identify the parameters which influence the longitudinal and lateral movements using correlation analysis

(iv) Train three supervised algorithms and one deep learning algorithm with the parameters, such as dependent longitudinal and lateral speeds with independent settings

(v) Conduct simulation runs based on the trained algorithms, evaluate the algorithms' performance, and present the results with some meaningful insights

2. Understanding Mixed Traffic Conditions

2.1. Study Section. For addressing the challenges of traffic flow modeling under mixed traffic conditions, it was decided to design an experiment incorporating diverse roadways and traffic conditions. A segment of the Western Expressway (Mumbai) in India was considered for this purpose. A wide variation in traffic flow was observed, ranging from free-flow to near-capacity traffic conditions, supplemented even with a stop-and-go congestion regime. Considering this as an opportunity, video-graphic surveys were carried out, and macroscopic plots were developed to identify different traffic states. Further, to sense the driving behavior in a detailed manner, high-quality trajectory data is generated at three different flow levels for the section considered under this research work. Considering the incompetence of automated image processing tools under mixed traffic conditions, a semiautomated image processing tool TDE [33] is used to develop trajectory data. Unlike the automated trajectory development tools, the semiautomated tool offers an added flexibility for tracking each vehicle that individually

demands huge human efforts. As a result, near to 100 percent of vehicles can be tracked over the study sections. Simultaneously, to eliminate the perspectival distortions, the developers of the TDE embedded a vanishing point mechanism to convert the perspective trapezoidal images of the study section to the real-world coordinates. To eliminate the occlusion effect in tracking demands good viewing angles. In the present study, video recordings are captured from a reasonable height to limit the occlusion effect. However, in flow 3 in the congested traffic, some instant's smaller vehicles beyond a heavy vehicle are occluded, but the number of heavy vehicles is less at flow 3, resulting in less occlusion. It may also be noted that the authors developed trajectory data of well-accepted sample size of more than 3312 vehicles, including all classes of vehicles, for capturing the possible stochasticity in terms of trajectory data. Keeping in view the complexity in tracking and developing trajectory data under the traffic conditions involving different classes of vehicles (implicitly varying driver behavior), this may be considered as one of the pertinent contributions from the present study.

Further, in line with the literature [34], to limit the noise in trajectory data, smoothening techniques were applied, and the details of traffic states, for which trajectory data is developed, are given in Table 1. More details about the data that can be used in the present work can be found from authors' previous studies [35, 36]. The selected study section's snapshots and developed time-space plots of vehicles observed during real field conditions on the western expressway are depicted in Figure 1. Based on the videographic surveys broadly, six types of vehicle categories are found on the selected roadway study section: Motorized three-wheelers (MThW), Motorized two-wheelers (MTW), Buses, Cars, Trucks, and Light commercial vehicles (LCV).

2.2. Overview of Mixed Traffic Conditions. In explaining the mixed traffic conditions in a better manner, an example is presented in Figure 2. It can be noted that the movement of the subject vehicle (MTW as an example in Figure 2) is depicted for different time frames in (a) through (h). It can be observed that the subject's motorized two-wheeler (marked with a yellow star) is largely influenced by its surrounding vehicles, and the subject vehicle tends to maneuver out from its surrounding vehicles to have a better movement and avoid delays. It may be noted that the MTWs are acquiring any available position on a roadway space based on the availability of adequate longitudinal and lateral space simultaneously. Due to this traffic movement nature, even the most established following behavior and lane-changing models from homogeneous traffic conditions tend to underperform under mixed traffic conditions (as depicted in Figure 2). In general, most vehicles following behavior models predict acceleration (a) and speed (v) of the follower vehicle as a function of several variables related to its leader vehicle, as follows:

$$a\,(\text{or})v = f\left(s, \Delta v, s_{\min}, V, \ldots \text{etc.}\right), \tag{1}$$

where s = relative spacing. Δv = relative speed, $s_{\min}$ = minimum spacing, and V = desired speed.

On the other hand, in mixed traffic conditions, the subject vehicle is primarily influenced by its surrounding vehicles present in the traffic stream. Hence, the subject vehicles' movement can be governed by other added parameters discussed in the following sections.

2.3. Assessing Diving Behavior. Considering the vehicles' naturalistic movement in the traffic stream, the subject vehicle's longitudinal and lateral movement is mainly dependent on its surrounding vehicles. In the present work, initially surrounding vehicles for a given subject vehicle are identified accurately using trajectory data. In line with the literature [37, 38], a surrounding zone created by the addition of 40 m distance in front (look-ahead) and 30 m distance behind (look-back) from the center of the subject vehicle is considered, with a total longitudinal distance of 70 m forming a longer side of a rectangle (Figure 3). A lateral distance of 5.5 m from the center position of the subject vehicle to the center position of the surrounding vehicles, including the total width of the subject vehicle (with an overlap of width), is considered over the entire road space (in longitudinal and lateral directions over time), as depicted in Figure 3. As per the developed logic, there can be a possibility of eight combinations of surrounding vehicles for the subject vehicle.

Based on the literature [39, 40] from both homogeneous and mixed traffic conditions, the parameters that influence the longitudinal and lateral speeds are identified. On these lines, for longitudinal speeds, around 16 independent parameters are identified, other 22 independent parameters are identified for lateral speeds, as shown in Table 2, along with a brief description. With the help of developed trajectory data sets, using python code, surrounding vehicles are identified, and all mentioned parameters are evaluated for each vehicle at every instant of time.

Later, to identify the influential parameters for longitudinal movement and lateral movement, the Spearman correlation [41] test is performed between vehicles' instantaneous longitudinal speeds and 16 parameters, which may influence the longitudinal movement. Similarly, for the instant lateral speed of the vehicles along with other 22 parameters, which may influence lateral movement. The parameters identified are presented in Table 3, along with their brief description. After correlation analysis, it is observed that with a change in traffic flow conditions, the correlation values differed substantially, and, in some cases, the correlation nature has even varied. For example, in the case of longitudinal movement, the parameters such as lateral tilt (long_8), the lateral gap with adjacent vehicle (long_10), present lane (long_12), presence of left leading vehicle (long_13), and presence of right leading (long_14), with a change in flow levels the nature of correlation is varied. Mainly it can be noted that the parameters mentioned above are related to the lateral gap. However, it is inferred that with the rise in flow levels, the vehicles' longitudinal movement is constrained. As a result, vehicles tend to find lateral gaps for better maneuverability, particularly flow 2 and flow 3 (higher traffic flow levels). Due to this, with a change in flow levels, the correlation is found to be varying.

Table 1: Details of the study section (multilane urban road).

Trap length (m)	Road width (m)	Traffic flow level	Traffic composition (%)[b]	Avg. speed (km/h)	Avg. flow (pcu/h	V/C	No. of vehicles tracked	Trajectory data duration (min)
120	17.5	Flow 1	15/35/5/40/2/3	65	4800	0.4	1080	15
		Flow 2	20/29/2/45/1/3	42	10120	0.85	1715	15
		Flow 3[a]	17/25/5/45/3/4	20	3500	<1	660	10

[a]Stop-and-go conditions. [b]Traffic composition in order of MThW/MTW/Bus/Car/Truck/LCV.

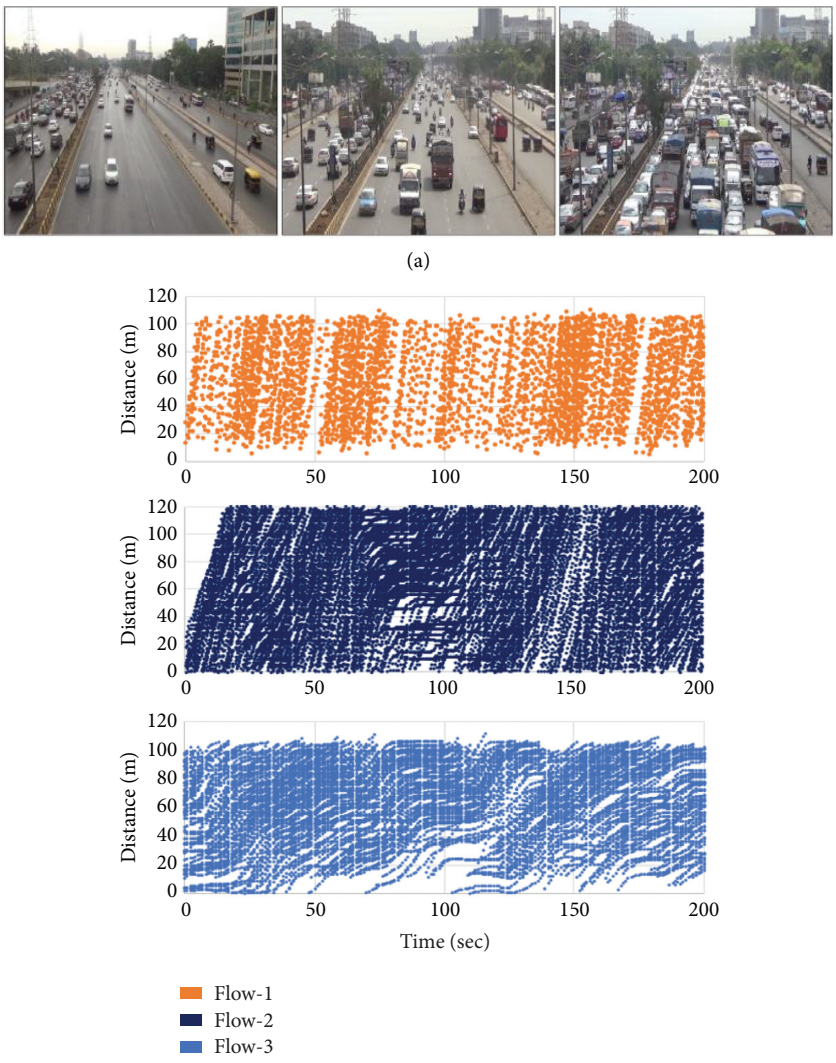

Figure 1: (a) Snapshots of the study section and (b) time-space plots of vehicles.

On the other hand, parameters such as right longitudinal gap (long_7), angle of seeping (long_9), the lateral gap with left adjacent (long_10), lateral gap with right adjacent (long_11), presence of right leading (long_14), TTC (long_15), and Smin/S (long_16) are found to be sparsely active under the present traffic conditions for the longitudinal movements, whereas parameters such as leader presence (long_1), relative distance (long_2), relative speed (long_3), leader vehicle category (long_5), left longitudinal gap (long_6), lateral tilt (long_8), present lane (long_12), and presence of left leading (long_13) are found to play a governing role in the longitudinal movement of the vehicles.

From the correlation analysis on the instant lateral speed with 22-lateral influencing parameters, it is found that, unlike longitudinal movement correlation analysis, the sense of the parameters (+ve/-ve correlation values) is similar for all flow conditions, whereas the parameters, such as present lane (lat_6), left front vehicle (lat_7), right front vehicle (lat_8), left back speed (lat_11), lateral tilt (lat_16), distance from left back (lat_17), distance from right back (lat_18), left

Figure 2: Movement of MTW over the road section at different time frames.

longitudinal gap (lat_21), and right longitudinal gap (lat_22) tend to have a good correlation with lateral speeds.

Based on the correlation analysis, it is observed that in the case of flow 2 and flow 3 (higher flow levels), where longitudinal movement is constrained, vehicles tend to find the lateral movement to escape the delay in the traffic stream. As a result, parameters associated with lateral gaps are better correlated with instantaneous longitudinal speeds. Similarly, parameters associated with longitudinal gaps are better correlated with instantaneous lateral speeds. In most of these cases, it may be noted that the correlation range of the parameters is found to be within 0.5. Given the stochastic nature of driving behavior, in line with the literature [42, 43], this can be treated as an acceptable correlation.

3. Machine Learning Modeling

In line with the work objectives, it is attempted to model the mixed traffic flow conditions. From the literature (Evan Lutins, 2017), it is inferred that numerous car-following behavior models are conceptualized. The following behavior models stood out for homogeneous traffic conditions and proved their potential in traffic flow modeling. On the other hand, in mixed traffic involving different vehicle categories and lack of lane discipline in the traffic, the vehicles' spatial interactions will be more involved. Even from correlation analysis, it is identified that numerous parameters affect the longitudinal and lateral movement of the vehicles. Considering this, modeling mixed traffic conditions with established car-following and lane-changing models from homogeneous traffic conditions may not be prudent.

In overcoming this challenge in the present work, to model the mixed traffic, machine learning from artificial intelligence is considered. Three established supervised machine learning algorithms, such as k-NN, random forest, and regression tree algorithms, are selected. Along with that, deep learning is also explored for modeling mixed traffic conditions.

3.1. k-NN Algorithm. In general k-NN algorithm (Min-Ling Zhang & Zhi-Hua Zhou, 2005) works on the principle of pattern recognition and learns the data patterns. To better explain this, let $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ be the data points

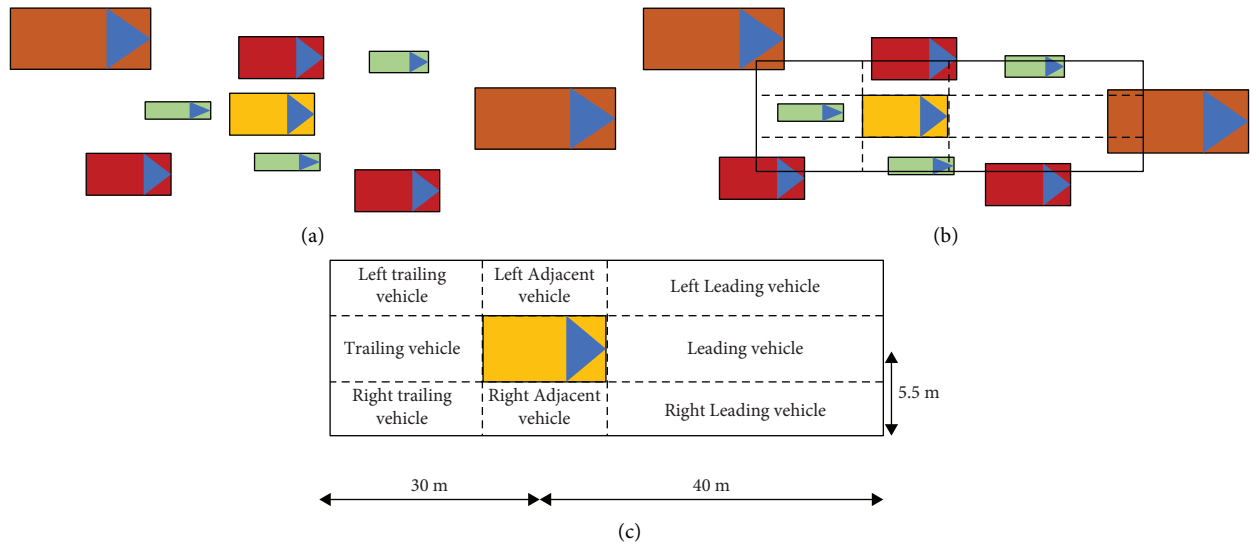

Figure 3: Explaining the surrounding vehicle combination for the subject vehicle.

from a sample space 'S' that belong to two classes, Class-I and Class-II, as follows:

$$\text{Class} - I = \begin{cases} (x_1, y_1) \\ (x_2, y_2) \\ . \\ . \\ . \\ (x_m, y_m) \end{cases}, \quad \text{Class} - II = \begin{cases} (x_{m+1}, y_{m+1}) \\ (x_{m+2}, y_{m+2}) \\ . \\ . \\ . \\ (x_n, y_n) \end{cases}. \tag{2}$$

Let the class of data point (x_t, y_t) be the point of interest from another sample space. To identify the class of the data, k-NN adopts the nearest neighboring approach. For example, say in the present case, k-NN adopts 3-neighbors. Initially, by means of Euclidean distance, the three nearest neighbors will be identified. Further, based on the majority class of the neighbors, the data point (x_t, y_t) class will be predicted. On these lines, by changing the number of neighbors and distances measures, the performance of the algorithm can be improved. In the present trial and error strategy adopted, 5-nearest neighbors are found to be optimized values.

3.2. Random Forest. In general, random forest [44] learns the data with a constructive multitude decision tree framework. Machine learning models the target outcomes in the form of categorization with ascertained probabilities. If the dependent variable is a categorical one, the category with maximum probability is given as an outcome by the machine learning models. On the other hand, if the target outcome is a continuous variable, in that case, by means of the weighted mean approach, the outcome is predicted. To explain the basic framework of the random forest algorithm in a better manner, let us consider the 'N' number of classes, with 'M' input variables or features. A number 't' is specified ($t < M$) such that at each node, t-variables will be selected at random out of M. The best split on these 't' is used to split the node. The value of 't' is held constant when the forest is developed. Further, each tree will be grown to the most substantial extent possible.

Let the training set $X = x_1, ..., x_n$ with responses $Y = y_1, ..., y_n$, bagging repeatedly (N times) selects a random sample with replacement of the training set and fits trees to these samples. In the next sample, t training examples are selected from X, Y as X_t, Y_t. Later by means of the random forest framework, trees f_t is trained on X_t, Y_t. After training the samples, the predictions can be given as follows:

$$\hat{f} = \frac{1}{t} \sum_{b=1}^{t} f_b(x_t), \tag{3}$$

where the number of trees and t are independent parameters that can be optimized using different cross-validation strategies.

3.3. Regression Tree. Decision tree learning [45] is a predictive model, where the decision tree is framed as branches (inputs) and leaves of the tree (outputs), in which the decision variable is categorized into subsets. A tree can be learned using recursive partitioning, in which the trained data is split into subsets until the trained data is matched with an observed target value. This process of Top-Down Induction of Decision Trees (TDIDT) [46] is generally applied in developing the decision trees. The independent

Table 2: Parameters influencing the movement of mixed vehicles.

ID	Longitudinal movement parameters	Description
long_1	Leader presence	The presence of leader vehicle is taken, 0 is assigned when its absent and 1 is taken if this is present
long_2	Relative distance (m)	Relative distance from leader vehicle
long_3	Relative speed with leader (m/s)	Relative speed (subject vehicle minus leader vehicle)
long_4	Subject vehicle category	Vehicle class of the subject vehicle.
long_5	Leader vehicle category	Vehicle class of the leader vehicle.
long_6	Left longitudinal gap(m)	Available left longitudinal gap for the subject vehicle
long_7	Right longitudinal gap (m)	Available right longitudinal gap for the subject vehicle
long_8	Lateral tilt with leader vehicle (m)	Lateral incline of the subject vehicle towards the leader vehicle in terms of lateral overlap (let side is taken negative value and right side as positive).
long_9	Angle of seeping(deg)	Available angle for seeping
long_10	Lateral gap with left adjacent vehicle(m)	Lateral gap clearance from the left adjacent vehicle
long_11	Lateral gap with right adjacent vehicle(m)	Lateral gap clearance from the right adjacent vehicle
long_12	Present lane of subject vehicle (1-median)	Present lane id in the order as median side lane -1, shoulder side lane -5
long_13	Presence of left leading	The presence of left leader vehicle is taken, 0 is assigned when its absent and 1 is taken if this present
long_14	Presence of right leading	The presence of the right leader vehicle is taken, 0 is assigned when its absent and 1 is taken if this is present
long_15	TTC (sec)	Time to collision
long_16	Smin/S	The ratio of minimum relative distance to relative distance
Id	Lateral movement parameters	Description
lat_1	Leader presence	The presence of leader vehicle is taken, 0 is assigned when its absent and 1 is taken if this is present.
lat_2	Leader vehicle category	Vehicle class of the leader vehicle.
lat_3	Subject vehicle category	Vehicle class of the subject vehicle.
lat_4	Relative speed with leader (m/s)	Relative speed (subject vehicle minus leader vehicle)
lat_5	Subject vehicle longitudinal speed (m/s)	Longitudinal speed of the subject vehicle
lat_6	Present lane	Present lane id in the order as median side lane -1, shoulder side lane -5
lat_7	Left front vehicle	Presence of left front vehicle, 0 is assigned when its absent and 1 is taken if this present.
lat_8	Right front vehicle	Presence of right front vehicle, 0 is assigned when its absent and 1 is taken if this is present.
lat_9	Left lateral clearance	Available lateral clearance in the left side
lat_10	Right lateral clearance	Available lateral clearance in the right side
lat_11	Left back vehicle speed (m/s)	Left back vehicle longitudinal speed
lat_12	Right back vehicle speed (m/s)	Right back vehicle longitudinal speed
lat_13	Left back vehicle acceleration (m/s^2)	Left back vehicle longitudinal acceleration
lat_14	Right back vehicle acceleration (m/s^2)	Right back vehicle longitudinal acceleration
lat_15	No. Of surrounding vehicles	Number of vehicles in the surrounding vicinity of the subject vehicle.
lat_16	Lateral tilt with leader vehicle (m)	Lateral incline of the subject vehicle towards the leader vehicle in terms of lateral overlap (let side is taken negative value and right side as positive).
lat_17	Distance from left back vehicle (m)	Longitudinal distance from left back vehicle
lat_18	Distance from right back vehicle (m)	Longitudinal distance from right back vehicle
lat_19	Area occupancy the vehicles ahead of subject vehicle (m^2)	Area occupied by the vehicles in the frontal surrounding vicinity
lat_20	Subject vehicle longitudinal acceleration (m/s^2)	Instant longitudinal acceleration of the subject vehicle
lat_21	Left longitudinal gap(m)	Available left longitudinal gap for the subject vehicle
lat_22	Right longitudinal gap (m)	Available right longitudinal gap for the subject vehicle

variables are best riven for the target variable; on this basis, the decree is selected to split the node. The same process is repeated until all the target values are sorted to either of the nodes.

Further, every branch of the decision tree dismisses a target value. Each target falls into one and exactly one terminal node, and each terminal node is uniquely defined by a set of rules [47]. Based on the class of the output decision, variables decision trees are classified as classification trees and regression trees.

Further regression trees employ Gini impurity [48] as a measure to check the accuracy of the tree labeling. The Gini impurity is nothing but the sum of the probability p_i of a data point with class i being chosen times the probability $\sum_{t\neq i}^{n} p_k = 1\text{-}p_i$ of error in selecting the class. The Gini impurity is given by the following:

Table 3: Spearman correlation coefficient for different flow levels.

ID	Parameter	Spearman correlation coefficient		
		Flow 1	Flow 2	Flow 3
(a) Longitudinal movements				
long_1	Leader presence	−0.31	−0.52	−0.60
long_2	Relative distance (m)	0.20	0.45	0.65
long_3	Relative speed with leader (m/s)	0.10	0.43	0.28
long_4	Subject vehicle category	−0.14	−0.08	−0.01
long_5	Leader vehicle category	−0.11	−0.12	−0.25
long_6	Left longitudinal gap(m)	−0.30	−0.16	0.15
long_7	Right longitudinal gap (m)	0.23	0.22	−0.01
long_8	Lateral tilt with leader vehicle (m)	−0.27	0.45	0.55
long_9	Angle of seeping(deg)	0.01	−0.22	−0.12
long_10	Lateral gap with left adjacent vehicle(m)	−0.08	0.03	0.19
long_11	Lateral gap with right adjacent vehicle(m)	0.02	0.13	0.06
long_12	Present lane of subject vehicle (1-median)	−0.50	−0.38	0.46
long_13	Presence of left leading	0.31	0.10	−0.12
long_14	Presence of right leading	−0.07	−0.16	0.12
long_15	TTC (sec)	−0.02	−0.02	−0.17
long_16	Smin/S	0.00	-0.17	-0.06
(b) Lateral movements				
lat_1	Leader presence	0	0	0.17
lat_2	Leader vehicle category	0.13	0.06	0.25
lat_3	Subject vehicle category	0.24	0.01	0.15
lat_4	Relative speed with leader (m/s)	−0.1	0.30	0.13
lat_5	Subject vehicle longitudinal speed (m/s)	0.06	0.46	0.32
lat_6	Present lane	−0.6	−0.70	−0.81
lat_7	Left front vehicle	0.18	0.59	0.59
lat_8	Right front vehicle	−0.6	−0.5	−0.52
lat_9	Left lateral clearance	0	−0.2	−0.19
lat_10	Right lateral clearance	0.32	0.17	0.15
lat_11	Left back vehicle speed (m/s)	0.48	0.42	0.47
lat_12	Right back vehicle speed (m/s)	−0.2	−0.21	−0.25
lat_13	Left back vehicle acceleration (m/s^2)	0.07	0	0.02
lat_14	Right back vehicle acceleration (m/s^2)	0.01	0	0.09
lat_15	No. Of surrounding vehicles	0	0.15	0.14
lat_16	Lateral tilt with leader vehicle (m)	−0.5	−0.3	−0.61
lat_17	Distance from left back vehicle (m)	−0.51	−0.4	0.65
lat_18	Distance from right back vehicle (m)	0.23	0.28	0.69
lat_19	Area occupancy the vehicles ahead of subject vehicle (m^2)	0	0.07	0.24
lat_20	Subject vehicle longitudinal acceleration (m/s^2)	0	0.04	0.07
lat_21	Left longitudinal gap(m)	−0.6	−0.61	−0.7
lat_22	Right longitudinal gap (m)	0.53	0.43	0.58

$$I_G(p) = \sum p_i(1 - p_i). \quad (4)$$

3.4. Deep Learning. Typically, deep learning is developed based on the neuron's architecture in the human brain cells. In which, the way electrical signals travel across the cells of living, each subsequent layer of nodes is activated when it receives stimuli from its neighboring neurons. Given this, the accuracy from deep learning models predictions can be increased with the right amount of training data. Deep learning: there will be three different layers, as input layers, hidden layers, and output layers, as shown in Figure 4. Specifically, the input layers are provided with the input vectors as $x_1, x_2, \ldots..x_n$. to map the final outcomes in the output layer. Given this, the input data is filtered through a series of hidden layers. The hidden layers are sandwiched between the input and output layers.

The deep learning models are developed with the help of python programming [49] using the Google TensorFlow [50] library environment. Later, the input parameters and the output velocities are mapped over deep learning models with numerous combinations of hidden layers, neuron activation functions, and many epochs. Applying the trial and error approach to limit the overfitting, for the present case, three hidden layers with 128, 64, and 16 nodes were adopted with 250 epochs by sequential modeling [51]. At the same time, ReLU [52] activation functions are used other than the final SoftMax [53] layer. The details of the trained deep learning models are presented in Table 4.

In the present study, the authors attempted to model the vehicles' instant longitudinal and lateral speeds instead of instant acceleration. It can be noted that, in normal traffic conditions, the acceleration values are in the range of -3.5 to $3.5\ m/s^2$, the longitudinal speeds are in the range of 0 to 25 m/s. In comparison, the range and variation of speeds are

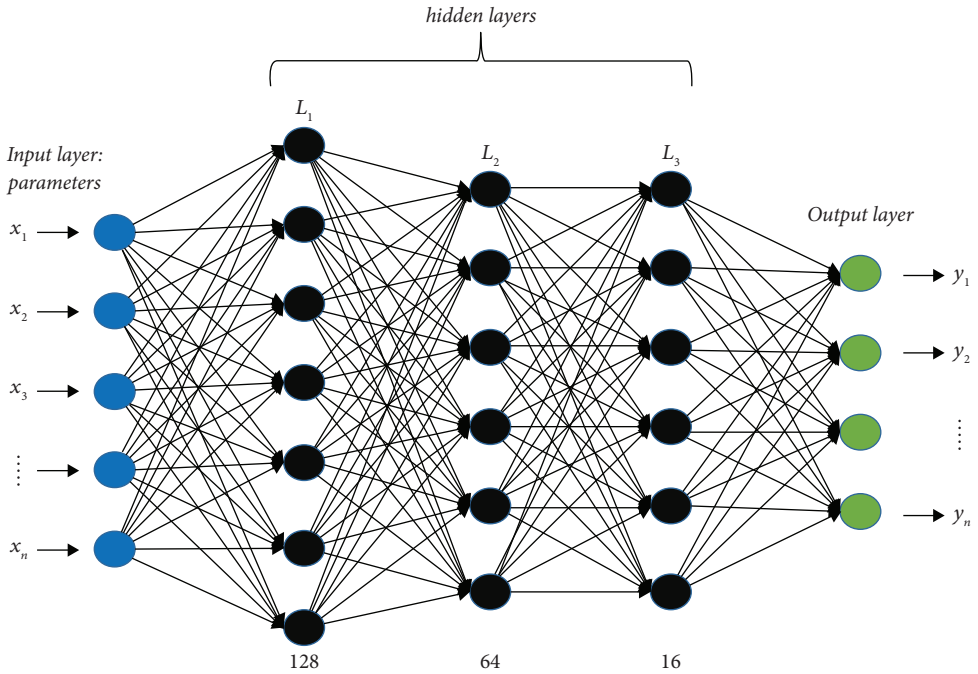

Figure 4: Deep learning architecture.

Table 4: Details of the trained deep learning model.

Layer (type)	No. of nodes	Activation function	
		Input	Output
L1 (dense)	128	ReLU	
L2 (dense)	64	ReLU	
L3 (dense)	16	ReLU	SoftMax
Model	Sequential		
Epochs	300		
Optimizer	RMSprop		
Loss function	Mean square error		
Metrics	Mean absolute error		

higher compared to the acceleration. Given the less range and variation, it is envisaged that the models will be underfitted if they are trained with acceleration as a dependent parameter. Considering this, the speeds are taken as a dependent parameter over the acceleration. In the present work, to improve the precision of the training of the algorithms, the dependent variables (instantaneous longitudinal and lateral speeds) were rounded off to 0.5 m/s and 0.01 m/s, respectively. Due to this, the variable classes decrease, and the data correlation patterns can be refined. For both the longitudinal and lateral movements, correlation coefficients with values equal to or greater than 0.4 at any of the flow levels were considered as influential parameters.

Using a similar approach, the preceding algorithms were trained for the dependent variables as instantaneous longitudinal speeds and instantaneous lateral speeds. In the case of k-NN, based on trial and error, the five-nearest neighbors were considered. For random forest, 15 trees were selected for training. The regression trees and deep learning were trained with their exact formulations. The training and setup process is carried out in python 3.7.0 programming language ("A primer on scientific programming with python," 2013). For training and testing the data, the entire trajectory data from the three different traffic flow conditions are divided into two equal halves. One-half is used for training the data and the other for testing purposes.

4. Simulation of Mixed Traffic

Further, based on the trained algorithms, vehicle movements are simulated again in python 3.7.0, as shown in Figure 5. According to their correct positions observed from field conditions, the vehicles were generated one after another, according to the initial time stamps and the positions, as shown in Figure 6. For that, the initial positions of all the vehicles are taken, and the vehicles are generated one after another according to the initial time stamps. At the same time, to sort the initial movement, the vehicles are placed with the true speeds (not by models, in the present study: 7 Vehicles). Later the vehicles are generated according to the initial start time and the positions. Upon their generation, the trained algorithms governed the subject vehicle's movement concerning its surrounding vehicles and derived the influential parameters in the traffic stream. On these lines, the simulation of mixed traffic is performed. In the present work, with the trained models along with surrounding vehicle combination, the vehicles next time instant longitudinal and lateral speeds will be predicted. For say at given the data of time step t_n, the speeds of t_{n+1} will be predicted again with the combination data of t_{n+1}; next t_{n+2} will be predicted and the process goes on till all predictions are made.

To assess whether the calibrated models' performance mimics the traffic behavior, time-space plots of vehicles were plotted one after another and compared with field extracted

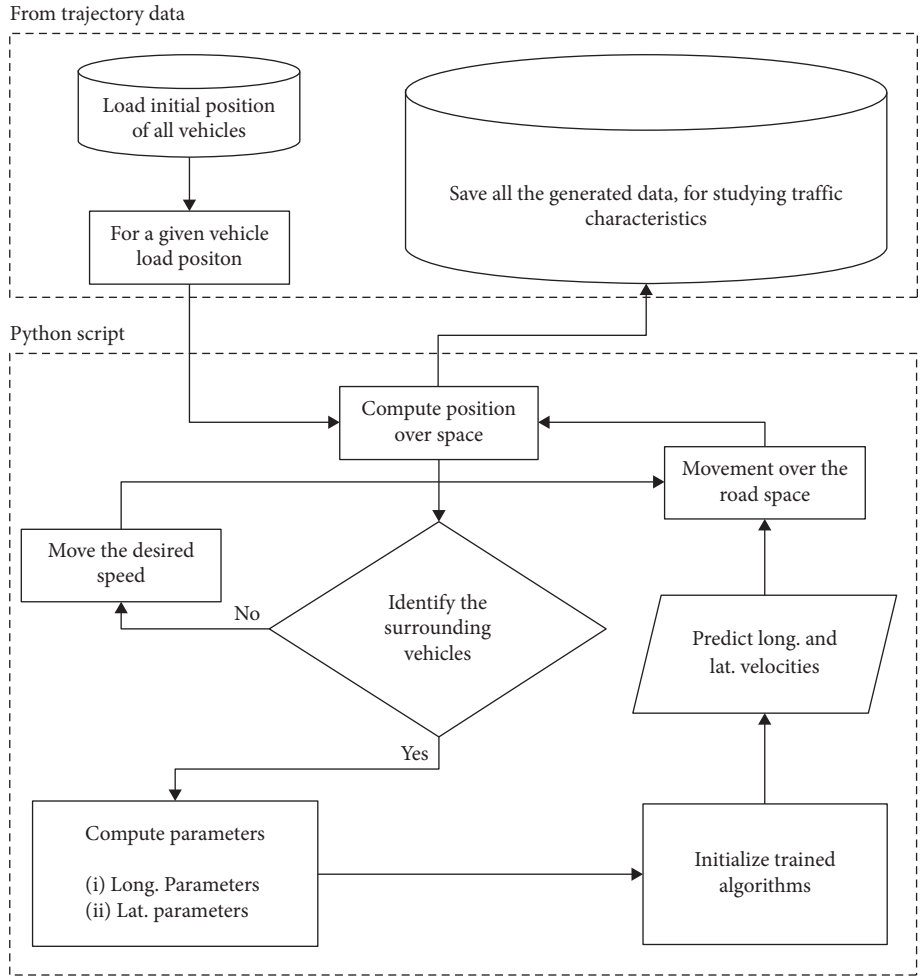

Figure 5: Simulation framework.

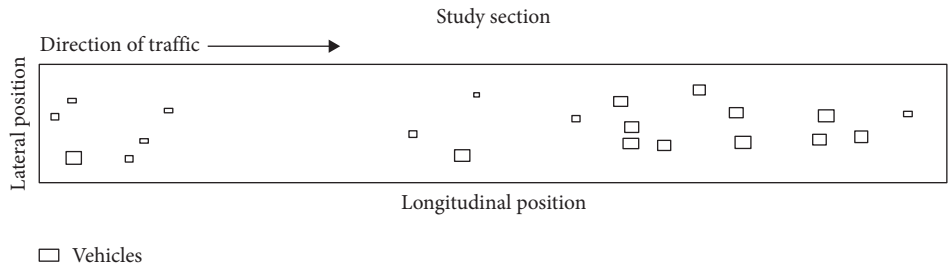

Figure 6: Snapshot of simulated movement of vehicles under mixed traffic environment.

vehicles, as shown in Figure 7. From the primary visualization of the time-space scenarios, it is observed that the simulated time-space plots using k-NN and deep learning algorithms tended to match the field observed time-space plots reasonably well. However, with random forest and regression trees, significant variation in the time-space plots was found. To assess the performance, the combined mean absolute percentage error (MAPE) was computed among the vehicle longitudinal and lateral positions for the three flow levels, and the results are as shown in Table 5. The MAPE values show that as the traffic flow level increases, the MAPE for each of the algorithms increases. In the case of k-NN, the MAPE was well within the range of 10%. For deep learning, MAPE was about 5%. On the other hand, in the random forest and regression tree case, the error was 7% to 17%.

Similarly, based on vehicles' simulated movement, macroscopic traffic characteristics, such as stream speed, density, and flow, are computed at every instant time frame using simulation. As speed, density, and flow are computed

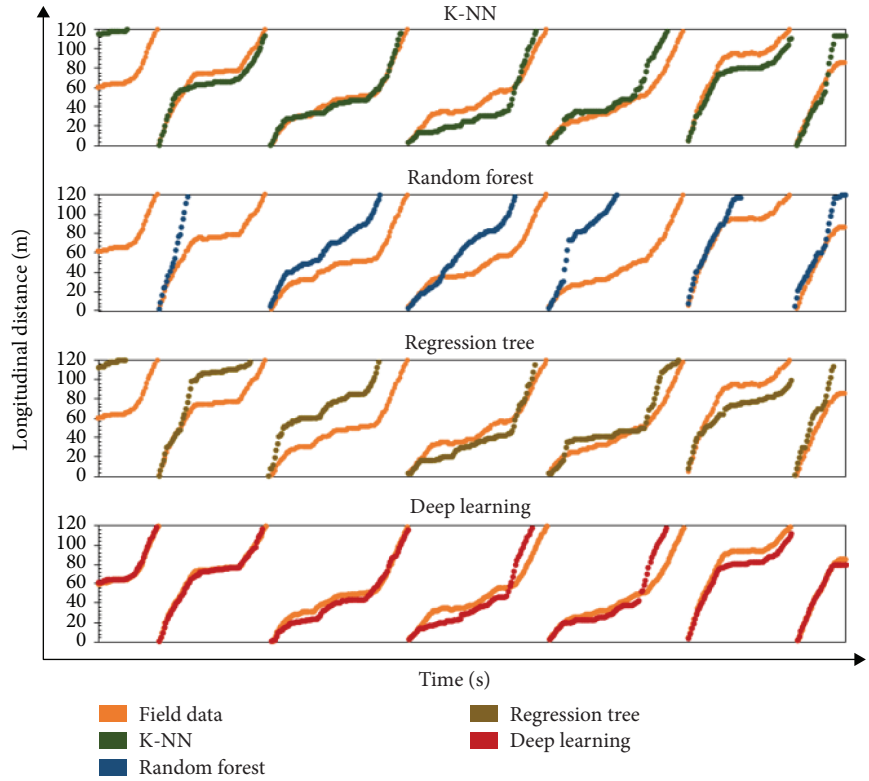

Figure 7: Time-space plots of the vehicles.

Table 5: Comparison of MAPE (%) in vehicle positions among the algorithms at the microscopic level.

Flow type	MAPE of positions (%)			
	k-NN	Random forest	Regression tree	Deep learning
Flow 1	5.2	7.5	8.2	3.9
Flow 2	7.6	14.5	13.2	4.7
Flow 3	9.7	17.7	16.9	5.2

every immediate time frame, a large amount of data is developed for developing fundamental macroscopic plots for traffic conditions ranging from the free-flow regime to the congested regime. Further to comprehensively develop the macroscopic plots, it is planned to adopt the Rakha model [54], as previously it is proven that it works well under traffic conditions considered in this study. The model formulation is briefly given by the following:

$$h = c_1 + c_3 u_n + \frac{c_2}{u_f - u_n}, \tag{5}$$

where h = headway, u_n = speed of the nth vehicle, and u_f = free-flow speed. c_1, c_2, c_3 = constants.

The density of the traffic stream k is given by the following:

$$k = \frac{1}{h} = \frac{1}{c_1 + c_3 u_n + c_2/u_f - u_n}, \tag{6}$$

where

$$c_1 = \frac{u_f}{k_j u_c^2}\left(2u_c - u_f\right), \tag{7}$$

$$c_2 = \frac{u_f}{k_j u_c^2}\left(u_f - u_c\right)^2, \tag{8}$$

$$c_3 = \frac{1}{q_c} - \frac{u_f}{k_j u_c^2}. \tag{9}$$

Using Equations (8)–(10), c_1, c_2 and c_3 parameters are estimated using the simulated macroscopic data and speed-density plots are developed for each of the algorithms, as shown in Figure 8. Further, speed-flow plots are also compared with observed speed-flow data plots, as shown in Figure 9. Additionally, boundary conditions such as free-flow speed, optimum speed, capacity, and jam density parameters derived from the macroscopic plots developed using algorithms are evaluated and are depicted in Table 6. Similar to microscopic outcomes, again, it is witnessed that deep learning tends to match the field conditions in a better manner. It is observed that with deep learning capacity is found to be around 11,810 pcu/h/direction, free-flow speed 57 kmph, and jam density as 815 pcu/km/direction. With observed field conditions having respective values of capacity, free-flow speed, and jam density as 11,860 pcu/h/direction, 61 kmph, and 810 pcu/km/direction. This further proves that the deep learning algorithm is fairly working well under mixed traffic conditions, contributing a mean

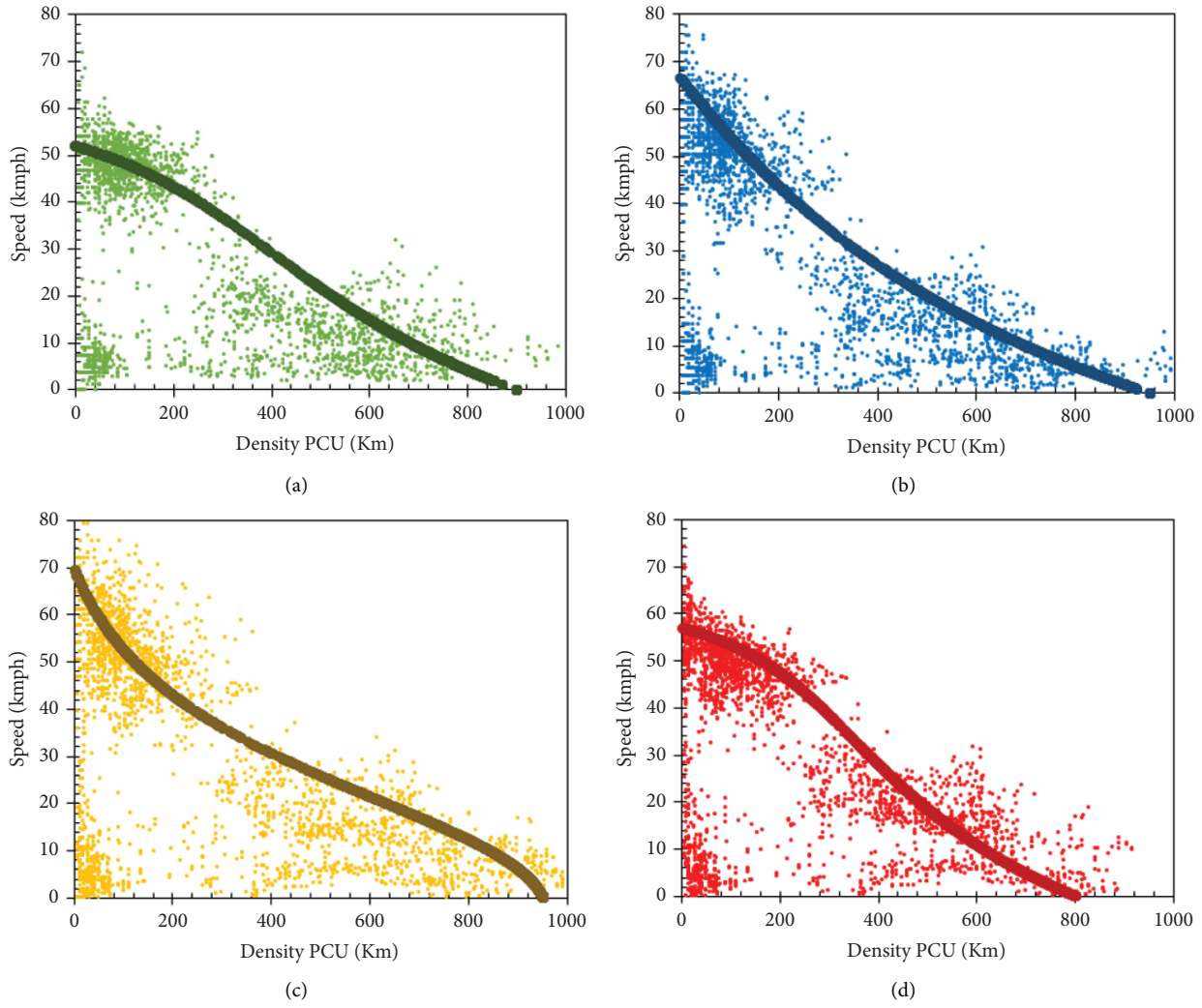

Figure 8: Macroscopic speed-density plots from different algorithms. (a) k-NN. (b) Random forest. (c) Regression tree. (d) Deep learning.

absolute percentage error (MAPE) of less than 5%. On the other hand, in the case of random forest and regression trees, mainly free-flow speeds are higher (about 70 kmph), even the jam densities are overestimated as 1,000 pcu/km/direction, contributing to a MAPE value of more than 10%. In addition to this, from the shape of macroscopic plots, it may be further corroborated that these plots (developed using regression trees and random forests) tend to deviate significantly from the forms of macroscopic plots developed using actual field data. Further, the macroscopic data is generated from the trajectory data for every time instant from the study section with both empirical data and the simulation data. While simulating the congestion at its end times, the vehicles in that traffic exited with their time stamps. There are no vehicles to enter the section, with vehicles exiting at lower speeds and no vehicles to enter the road space. The density levels in the traffic stream fell near the end timestamps of the simulation. As a result, fewer speeds are observed at lower density levels.

$$\text{MAPE} = \frac{1}{\text{sample size}} \sum_{\text{position}=1}^{\text{sample size}} \frac{\left|\text{simulated}_{\text{position}} - \text{observed}_{\text{position}}\right|}{\text{observed}_{\text{position}}} * 100. \tag{10}$$

5. Practical Aspects

In the present study, the models' core logic in learning the data patterns played a huge part in revealing the models' performance. For example, the random forest and regression trees apply ensemble learning methods for learning the data. Given the multitude of decision trees in handling the data, the target outcome speeds have deviated. Whereas k-NN works on the nearest neighbor approach, with less variation in the dependent variables in the close vicinity, k-NN performance tends to show better results. On the other hand,

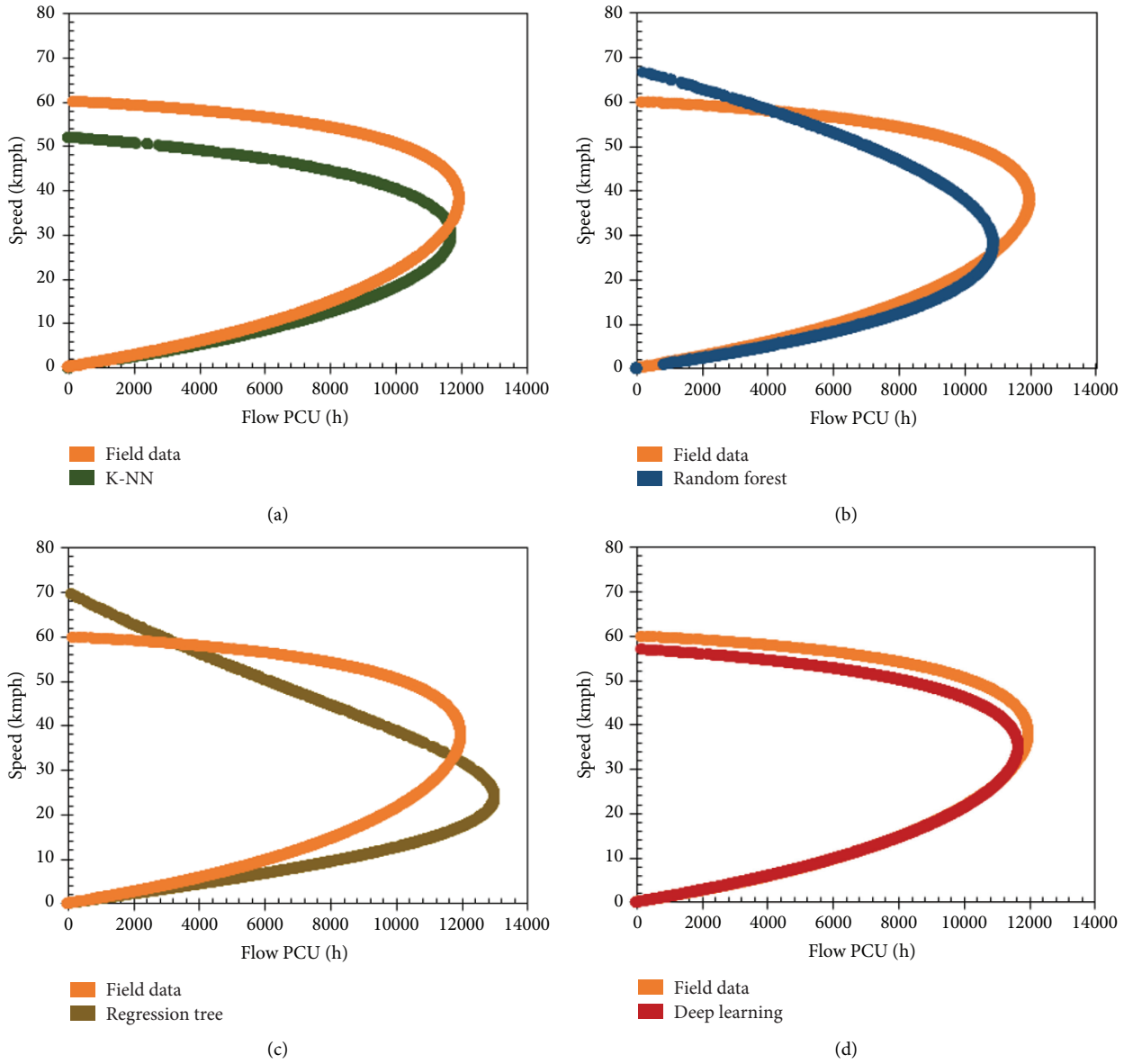

Figure 9: Macroscopic speed-flow plots from different algorithms. (a) k-NN. (b) Random forest. (c) Regression tree. (d) Deep learning.

Table 6: Performance comparison of observed and predicted traffic flow measures for selected algorithms at the macroscopic level.

Measure	Observed	Predicted			
		k-NN	Random forest	Regression tree	Deep learning
Free-flow speed (km/h)	61	52	69	70	57
Optimum speed (km/h)	41	31	27	21	40
Capacity (pcu/h)	11860	11500	12514	13125	11810
Jam density (pcu/km)	810	912	987	980	815

the deep learning model depicts the speed variations given the internal neural layered structure.

Presently modeling the car-following, lane-changing, and lateral movements of the vehicles are challenging aspects for the researchers and practitioners working under mixed traffic conditions. With the help of the depicted methodology, simulation modeling can be done with ease. Simultaneously, the accuracy in the present simulation packages can be improved by embedding the illustrated algorithms. Presently, in mixed traffic, real trajectory datasets are very scarce in the present context. As a result, very few driving behavior studies are attempted in this direction. With the study methodology, naturalistic trajectory data can be predicted for carrying driving behavior instincts. Along with that, the vehicle's

driving behavior can be quantified, and the level of aggression can be checked with the modeled driving data.

Currently, trajectory prediction is one of the significant hurdles remaining to achieve safe and reliable autonomous driving. There have been many proposed metrics for evaluating the quality of forecasts on static datasets. However, trajectory prediction for autonomous driving inherently must run in real-time, in conjunction with other driving pipeline components, such as planning. We have discussed why algorithm runtime, environment complexity, and frequency of predictions should also be considered when evaluating a prediction algorithm. To do this, we implemented several state-of-the-art prediction models and evaluated their behavior in a realistic simulation.

6. Conclusions

In a mixed traffic stream, vehicular movement is primarily influenced by its surrounding vehicles in the flow. Thus, both longitudinal and lateral movement continuously varied over a given road space based on the traffic conditions. This continuous longitudinal and lateral movement phenomenon is the principal root cause of the limited performance of the established car-following and lane-changing models developed under the perfect lane-discipline environment prevailing under homogeneous traffic conditions.

The correlation analysis shows that the vehicles' instantaneous longitudinal speeds are reasonably correlated with lateral gap parameters under mixed traffic conditions. Similarly, lateral speeds are also correlated with longitudinal distance parameters. Interestingly, in the case of flow 1, that is, free-flow traffic conditions, the correlation nature (positive/negative) for certain parameters has differed compared to flow 2 and flow 3, both for longitudinal and lateral speeds. This analysis carried out in this research work indicates that initially, at free-flow conditions, the vehicle tends to move continuously over the road space with less lateral amplitude. But vehicles tend to find the lateral movement for escaping the delay in the traffic stream when the vehicles' longitudinal movement is constrained.

Based on the study, it is well established that by deploying advanced computational tools such as machine learning tools, mixed traffic conditions can be modeled with better accuracy. Based on the analysis conducted here, it is observed that k-NN and deep learning algorithms mimic the mixed traffic conditions better with a MAPE of 3 to 9 percent at microscopic and macroscopic levels.

The algorithm performances can be attributed to their core model stability in handling complex data patterns. On the other hand, with a random forest and regression tree, the results tend to deviate substantially from the actual field observed traffic conditions. It can be noted that in both algorithms, data is trained in a tree assembly. Due to this, the predefined form (less flexibility) and multiple causal parameters, the algorithms' performance is limited.

Further, the methodology adopted in the present work addresses the challenge of reasonably replicating mixed traffic conditions and will be useful in modeling such mixed traffic scenarios effectively to develop viable, practical applications. Interestingly, these tested algorithms can also be used in traffic microsimulation tools to replicate mixed traffic and boost traffic simulation studies in these conditions.

Furthermore, it is inevitable that, due to variation in the driving behavior, the microscopic traffic simulation studies are limited under mixed traffic conditions, given the better accuracy of modeling the mixed traffic conditions using machine learning algorithms. These algorithms can be embedded in simulation tools as a substitute for following behavior and lane-changing models and could have a strong potential to boost the simulation studies from these traffic conditions (mixed traffic). Adopting this strategy for modeling homogeneous traffic conditions, the simulation models' accuracy may also be taken to the next level. [55–57].

7. Limitations and the Future Scope

Along with the research findings, the present study has certain limitations, which can act as the future scope of the work.

(i) Driving behavior is a stochastic phenomenon. All the drivers might behave differently by their own choice, which is highly discrete. Nevertheless, at times, the machine learning models and the deep learning model will remain ineffective due to interdriver variability but observed to perform better than the present limitations of conventional models of traffic flow modeling under mixed traffic conditions.

(ii) In the present work, the models are tested at three different flow conditions from the study section. However, the present study framework must be tested over the study sections with even more different flow conditions with variations in vehicles' proportion. This can undoubtedly help in comprehensively gauging machine learning and deep learning models for traffic flow modeling.

(iii) In the present study, the simulation analysis is performed by generating the vehicles at the same observed time stamps and the positions and governing the movement with models. As a result, the variation in the vehicle's arrival and the impact due to the composition of the traffic is not much tested in the simulation process. However, this can act as the future scope of the work.

Data Availability

The data that support the findings of this study can be made available upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The present research work was supported by Mitacs Global Research Award by Mitacs, Canada.

References

[1] L. A. Pipes, "An operational analysis of traffic dynamics," *Journal of Applied Physics*, vol. 24, no. 3, pp. 274–281, 1953.

[2] C. Laws, "Characterization of traffic oscillation propagation under nonlinear," *Environmental Engineering*, vol. 17, 1981.

[3] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological*, vol. 15, no. 2, pp. 105–111, 1981.

[4] R. Wiedemann, *Simulation des Straßenverkehrsflusses. Schriftenreihe des IfV, 8*, Institut für Verkehrswesen. Universität Karlsruhe, Karlsruhe, Germany, 1974.

[5] M. Saifuzzaman and Z. Zheng, "Incorporating human-factors in car-following models: a review of recent developments and research needs," *Transportation Research Part C: Emerging Technologies*, vol. 48, 2014.

[6] E. R Boer, "Car following from the driver's perspective," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 2, 1999.

[7] H. Zhang, C. Wu, X. Yan, and T. Z. Qiu, "The effect of fatigue driving on car following behavior," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 43, 2016.

[8] H. Lenz, C. K. Wagner, and R. Sollacher, "Multi-anticipative car-following model," *European Physical Journal B*, vol. 85, 1999.

[9] M. Ben-Akiva, C. Choudhury, and T. Toledo, "Lane changing models," *Geo Connexion*, vol. 2, 2006.

[10] T. Toledo, C. Choudhury, and M. Ben-Akiva, "Lane-changing model with explicit target lane choice," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1934, 2007.

[11] T. Toledo and D. Zohar, "Modeling duration of lane changes," *Transportation Research Record: Journal of the Transportation Research Board*.vol. 1999, 2007.

[12] N. Raju and H. Farah, "Evolution of traffic microsimulation and its use for modeling connected and automated vehicles," *Journal of Advanced Transportation*, vol. 2021, Article ID 2444363, 29 pages, 2021.

[13] P. Ag, "Ptv VISSIM. In data diperoleh melalui situs internet," 2017, https://en.wikipedia.org/wiki/PTV_VISSIM.

[14] J. Casas, J. L. Ferrer, D. Garcia, J. Perarnau, and A. Torday, "Traffic simulation with aimsun," *International Series in Operations Research and Management Science*, vol. 145, 2010.

[15] P. Kachroo, K. Ozbay, P. Kachroo, and K. Ozbay, "Paramics," *Feedback Ramp Metering in Intelligent Transportation Systems*, Springer Science & Business Media, Berlin, Germany, 2003.

[16] D. Krajzewicz, G. Hertkorn, P. Wagner, and C. Rössel, "Sumo (simulation of urban mobility)," in *Proceedings of the 4th Middle East Symposium on Simulation and Modelling (MESM2002)*, Berlin-Adlershof, Germany, September, 2002.

[17] Ngsim, "Next Genaration Simultaion, FHWA," 2007, https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.

[18] W. Lu, Y. Rui, and B. Ran, "Lane-level traffic speed forecasting: a novel mixed deep learning model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, 2020.

[19] B. Peng, M. F. Keskin, B. Kulcsár, and H. Wymeersch, "Connected autonomous vehicles for improving mixed traffic efficiency in unsignalized intersections with deep reinforcement learning," *Communications in Transportation Research*, vol. 1, no. October, Article ID 100017, 2021.

[20] H. Shi, Y. Zhou, K. Wu, X. Wang, Y. Lin, and B. Ran, "Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment," *Transportation Research Part C: Emerging Technologies*, vol. 133, 2021.

[21] S. M. P. Siddharth and G. Ramadurai, "Calibration of VISSIM for Indian heterogeneous traffic conditions," *Procedia - Social and Behavioral Sciences*, vol. 104, pp. 380–389, 2013.

[22] M. S. Bains, B. Ponnu, and S. S. Arkatkar, "Modeling of traffic flow on Indian expressways using simulation technique," *Procedia - Social and Behavioral Sciences*, vol. 43, 2012.

[23] T. Mathew and K. Ravishankar, "Car-following behavior in traffic having mixed vehicle-types," *Transportation Letters*, vol. 3, 2011.

[24] G. Asaithambi and J. Joseph, "Modeling duration of lateral shifts in mixed traffic conditions," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, 2018.

[25] C. R. Munigety and T. V. Mathew, "Towards behavioral modeling of drivers in mixed traffic conditions," *Transportation in Developing Economies*, vol. 2, 2016.

[26] S. Patil, N. Raju, S. Arkatkar, and S. Easa, "Modeling Vehicle Collision Instincts over Road Midblock Using Deep Learning," in *Proceedings of the TRB Annual Conference*, Washington, D.C, USA, January, 2021.

[27] N. Raju, S. Arkatkar, and G. Joshi, "Modeling following behavior of vehicles using trajectory data under mixed traffic conditions: an Indian viewpoint," *Transportation Letters*, vol. 13, pp. 1–15, 2020b.

[28] N. Raju, S. Arkatkar, S. Easa, and G. Joshi, "Addressing challenges of modeling mixed traffic through machine learning supported with vehicular trajectory data," in *Proceedings of the Transportation Research Board 100th Annual MeetingTransportation Research Board*, 2021.

[29] J. Stilgoe, "Machine learning, social learning and the governance of self-driving cars," *Social Studies of Science*, vol. 48, 2018.

[30] H. Wakabayashi, Y. Matsumoto, T. Tsubota et al., "Route CHoice behaviour by travel time," *Transportation Research Part C: Emerging Technologies*, vol. 5, 2015.

[31] Y. Yu, Z. He, and X. Qu, "On the impact of prior experiences in car-following models: model development, computational efficiency, comparative analyses, and extensive applications," *IEEE Transactions on Cybernetics*, 2021, In press.

[32] Z. Zheng, "Reasons, challenges, and some tools for doing reproducible transportation research," *Communications in Transportation Research*, vol. 1, 2021.

[33] V. Vicraman, C. Ronald, T. Mathew, and K. V. Rao, "Traffic data extractor. IIT, bombai," 2014, http://www.civil.iitb.ac.in/tvm/tde2.

[34] N. Raju, P. Kumar, C. Reddy, S. Arkatkar, and G. Joshi, "Examining smoothening techniques for developing vehicular trajectory data under heterogeneous conditions," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 12, pp. 1549–1568, 2017.

[35] N. Raju, P. Kumar, A. Jain, S. S. Arkatkar, and G. Joshi, "Application of trajectory data for investigating vehicle behavior in mixed traffic environment," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 97, 2018.

[36] N. Raju, S. Arkatkar, and G. Joshi, "Effect of construction work zone on traffic stream parameters using vehicular trajectory data under mixed traffic conditions," *Journal of Transportation Engineering, Part A: Systems*, vol. 146, no. 6, Article ID 5020002, 2020a.

[37] E. Lehtonen, O. Lappi, H. Kotkanen, and H. Summala, "*Look-ahead fixations in curve driving*," *Ergonomics*, vol. 56, 2013.

[38] A. E. Patla and J. N. Vickers, "How far ahead do we look when required to step on specific locations in the travel path during locomotion?" *Experimental Brain Research*, vol. 148, 2003.
[39] G. Asaithambi, V. Kanagaraj, and T. Toledo, "Driving behaviors: models and challenges for non-lane based mixed traffic," *Transportation in Developing Economies*, vol. 2, 2016.
[40] D. D. Salvucci, "Modeling driver behavior in a cognitive architecture," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48, 2006.
[41] J. H. Zar, "Spearman rank correlation," *Encyclopedia of Biostatistics*, Wiley & Sons, Hoboken, NJ, USA, 2005.
[42] E. R. Dahlen, R. C. Martin, K. Ragan, and M. M. Kuhlman, "Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving," *Accident Analysis And Prevention*, vol. 37, 2005.
[43] T. A. Ranney, "Models of driving behavior: a review of their evolution," *Accident Analysis And Prevention*, vol. 26, 1994.
[44] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
[45] T. M. Mitchell, "Decision tree learning," in *Lecture Slides for Textbook Machine Learning, T. M. Mitchell, McGraw Hill*, pp. 52–80, New York, NY, USA, 1997.
[46] N. A. B. Gray, "Capturing knowledge through top-down induction of decision trees," *IEEE Expert*, vol. 5, no. 3, pp. 41–50, 1990.
[47] W. Buntine and T. Niblett, "A further comparison of splitting rules for decision-tree induction," *Machine Learning*, vol. 8, no. 1, pp. 75–85, 1992.
[48] J. H. Kwakkel and M. Jaxa-Rozen, "Improving scenario discovery for handling heterogeneous uncertainties and multinomial classified outcomes," *Environmental Modelling and Software*, vol. 79, 2016.
[49] G. Van Rossum and F. L. Drake Jr, *Python Tutorial*, Centrum voor Wiskunde en Informatica, Amsterdam, Netherlands, 1995.
[50] Google, "TensorFlow. An End-To-End Open Source Machine Learning Platform," 2020, https://www.tensorflow.org/.
[51] Y. Bengio, I. J. Goodfellow, and A. Courville, "Sequence modeling : recurrent and recursive nets," in *Lecture Slides for Chapter 10 of Deep Learning*, McGraw Hill, New York, NY, USA, 2015.
[52] S. Pattanayak, *Pro deep learning with TensorFlow A mathematical approach to advanced artificial intelligence in Python*Apress, New york, NY, USA, 2018.
[53] F. Nelli, "Deep learning with TensorFlow," in *Python Data Analytics*, pp. 349–407, Apress, Berkeley, CA, USA, 2018.
[54] H. Rakha and B. Crowther, "Comparison of greenshields, Pipes, and van aerde car-following and traffic stream models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1802, no. 1, pp. 248–262, 2007.
[55] A primer on scientific programming with Python, "Choice Reviews Online," 2013, https://en.wikipedia.org/wiki/Choice_Reviews.
[56] E. Lutins, "Towards data science," 2017, https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f.
[57] M.-L. Zhang and Z.-H. Zhou, "A k-nearest neighbor based algorithm for multi-label classification," in *Proceedings of the 2005 IEEE International Conference on Granular Computing*, Beijing, China, July, 2005.