Unraveling Gap Selection Process during Discretionary Lane Changing by Vehicle Class

Sharma, Salil; Van Lint, Hans; Tavasszy, Lorant; Snelder, Maaike

DOI
10.1109/ACCESS.2022.3159705

Publication date
2022

Document Version
Final published version

Published in
IEEE Access

IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 4, 2022, accepted March 6, 2022, date of publication March 14, 2022, date of current version March 23, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3159705

Unraveling Gap Selection Process During Discretionary Lane Changing by Vehicle Class

SALIL SHARMA[1], HANS VAN LINT[1], LÓRÁNT TAVASSZY[1], (Member, IEEE), AND MAAIKE SNELDER[1,2]

[1]Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, 2628 CN Delft, The Netherlands
[2]TNO, 2595 DA The Hague, The Netherlands

Corresponding author: Salil Sharma (s.sharma-4@tudelft.nl)

This work was supported by the Netherlands Organization for Scientific Research (NWO), TKI Dinalog, Commit2data, Port of Rotterdam, SmartPort, Portbase, TLN, Deltalinqs, Rijkswaterstaat, and TNO under the Project "ToGRIP-Grip on Freight Trips."

ABSTRACT This paper studies and compares the gap selection process of multiple vehicle classes (passenger cars, delivery vans, and trucks) within their discretionary lane changing activities. Given a trajectory or a sequence of gap selection decisions, we aim to predict whether a vehicle will change or keep a lane. For this purpose, we use a large trajectory dataset, collected for the Netherlands, consisting of 3,647 trajectories of passenger car drivers, 1,080 trajectories of delivery van drivers, and 2,226 trajectories of truck drivers. We apply gated recurrent unit neural networks to separately model their gap selection processes. These three models can not only handle class imbalance but also capture long-term interdependencies. The models can predict gap selection of three vehicle classes with geometric mean accuracies of 84% or higher. To obtain insights into their gap selection processes, we apply a gradient-based technique to analyze what neural networks have learned. Our results suggest that there exist significant differences between vehicle classes in terms of the importance of historical information and features. Trucks seem to value a relatively long period, recent 6 seconds, of driving experience to select gaps compared to passenger cars and delivery vans. In addition, the perception of road topology seems to be a significant factor for delivery vans and trucks, contrary to passenger cars which highly value their kinematic features and interactions with surrounding vehicles, to select gaps. These insights offer a novel contribution towards better understanding and modeling of the driving behavior of multiple vehicle classes.

INDEX TERMS Driving behavior, discretionary lane-changing, gap selection, trajectory data, gated recurrent unit neural network, class imbalance, explainable AI.

I. INTRODUCTION

Lane-changing is an important aspect of driving behavior that has a significant influence on road capacity [1], safety [2], and emissions [3]. Two main categories of lane changing can be distinguished: mandatory and discretionary. Mandatory lane changes arise from either infrastructural or traffic control related constraints, or from the drivers' need to follow a path that leads to his or her—for brevity, we will use male adjectives in the ensuing—destination. Discretionary lane changes are associated with the driver's desire to improve his current driving conditions. Discretionary lane-changing (DLC) is typically structured as a hierarchical process [4], [5] where a driver (1) makes a decision-in-principle (that driving conditions on the current lane are below some desired level and can be improved by shifting to another lane); assesses (2) the options for this lane change (which target lane to move to) and (3) the necessary conditions (the suitability of available gaps on potential target lane(s)); and then finally (4) takes action (initiates and executes the lane change) or not (rejects available gaps, or even abandons the entire lane change maneuver).

Gap selection is an important stage of the lane-changing process where drivers explicitly look for a suitable and safe opportunity in order to initiate their desired lane-changing maneuver. This stage has been extensively studied for passenger car drivers [5]–[9] whereas other vehicle classes such as trucks or delivery vans have not received any attention. Although previous research [10], [11] shows that trucks seem to significantly affect traffic operations, only the first two

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

steps of the hierarchical model for DLC decisions of truck drivers have been investigated [4]. Inter-vehicle interactions during lane change have been shown to affect traffic operations; therefore, it is of vital importance to investigate and compare the lane change behavior of multiple vehicle classes to ensure reliable, efficient, and safe traffic operations. To the best of our knowledge, our paper is the first that focuses particularly on the third step of this hierarchy, studies the gap selection process of delivery van and truck drivers within their DLC maneuvers, and compares it with that of passenger car drivers.

The gap selection behavior of vehicles can be formulated as a binary decision problem with two outcomes: lane-changing (accept) and lane-keeping (reject). This problem is solved using a wide range of techniques in the existing literature: rule-based [12], statistical [6], [7], econometrical [8], and artificial intelligence (AI) models [5], [9]. Most of the earlier works assume instantaneous decision-making in the sense that only features or variables at a specific time instant affect the gap selection decision process. Typically, this time instant is taken just before a vehicle starts shifting laterally, which is an indication that a gap has been accepted [5], [13]. Although some literature [8], [14], [15] shows that historical data may also influence the gap selection process, such long-term interdependencies are typically not considered. In this paper, we consider the long sequences (or trajectories) of up to 20 sec to fill this gap.

The objective of this paper is to obtain insights into the gap selection process of multiple vehicle classes in their DLC maneuvers using AI. To this end, we frame the gap selection process of truck drivers as a many-to-one sequence classification problem and train a gated recurrent unit neural network (GRUNN) model to learn and model such temporal dependencies over longer periods. To assess what this neural network, in the end, has learned—and whether this makes sense behaviorally, we apply explainable AI techniques such as a gradient-based technique [16] and variable importance.

Most previous research works calibrate and validate their gap selection models using data from a specific type of topology (e.g., a weaving section [5], [7]–[9]). In this paper, we use a larger trajectory dataset that covers many different topologies situated around 14 different bottlenecks in the Netherlands including on-ramps, off-ramps, and weaving sections [17], [18]. We incorporate these different topologies in the gap selection model as part of the feature set fed to the GRUNN model and consider for example type of topology, length of the infrastructural bottleneck, and the number of lanes on the mainline carriageway.

This paper contributes to the existing literature by:

1. building gated recurrent unit neural network models to capture the gap selection process of multiple vehicle classes (passenger cars, delivery vans, and trucks) during their discretionary lane changing;
2. considering historical sequential data and external factors arising from topologies in the modeling framework of the gap selection process; and
3. comparing the gap selection process of multiple vehicle classes (passenger cars, delivery vans, and trucks) by unraveling their latent gap selection mechanisms and identifying key features that impact their gap selection through explainable AI techniques.

This paper is organized in the following way. It begins by providing a theoretical background on GRU neural networks and related techniques to interpret their predictions. Subsequently, the data generation process is described. The next section presents an experimental setup to model the gap selection process of truck drivers. Then, the subsequent section presents the model performance and its interpretability. Afterward, the findings and their implications are discussed. Finally, the paper concludes with some recommendations for further research work.

II. RELATED BACKGROUND

A gated recurrent unit neural (GRU) network model is selected in this paper to model the gap selection behavior of multiple vehicle classes (passenger cars, delivery vans, and trucks). A major advantage of this approach over other approaches (e.g., rule-based, econometrical, statistical, fuzzy-logic) is that it can learn long-term temporal interdependencies. The first part of this section discusses the inner workings of GRU neural networks. The second part presents strategies to handle class imbalance which is often a case in real-world trajectory datasets. To interpret predictions and learn more about the gap selection behavior, the third part of this section presents explainable AI techniques.

A. GATED RECURRENT UNIT NEURAL NETWORKS

A recurrent neural network (RNN) is a widely used method that can handle time-series data for prediction purposes. However, an RNN suffers from well-known problems of vanishing and exploding gradients during backpropagation and is not very good at capturing very long-term dependencies. To overcome these obstacles, long short-term memory (LSTM) neural networks were proposed [19]. Cho *et al.* [20] proposed a GRU neural network or GRUNN, which is a variant of LSTM. Compared to LSTM, GRUNNs have simplified connections and a reduced number of parameters. While LSTM contains three gates (input gate, forget gate, and output gate), the GRU comprises two gates, namely the update gate and the reset gate. The update gate controls the extent to which the state information of the previous moment is passed to the current state. While the reset gate controls how much information of the previous state is stored in the current candidate state $\tilde{h}_t$. In this way, GRU can improve upon the training efficiency by relying on the memory ability of neurons and fewer tensor operations. The architecture of a GRU cell is shown in Fig. 1.

The forward propagation process of GRU is as follows (see (1)-(4)):

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (1)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (2)$$

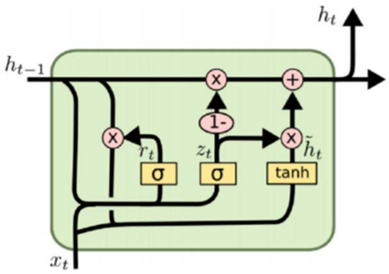


FIGURE 1. Architecture of a GRU cell (adapted from (Cho *et al.* [20])).

$$\tilde{h}_t = \tanh(W_h \cdot [r_t * h_{t-1}, x_t] + b_h) \quad (3)$$
$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (4)$$

where x_t denotes the input vector at time t. h_t denotes the state of the system at time t. $\tilde{h}_t$ denotes the current candidate state. z and r denote the update gate and reset gate, respectively. σ denotes the Sigmoid function. W_z, W_r, and W_h refer to weight matrices. b_z, b_r, and b_h refer to bias vectors.

B. RELEVANCE OF CLASS IMBALANCE TO UNDERSTANDING LANE-CHANGING BEHAVIOR

Most of the current research on lane-changing does not address the problem of class imbalance. However, there seems to be a large gap between the number of lane-changing trajectories and lane-keeping trajectories in most of the collected datasets [5]. This results in class imbalance because of an unequal distribution of instances belonging to target classes. The performance of traditional classifiers is likely to be affected if they are trained on imbalanced datasets. To handle this problem, previous research has used several techniques: data-level methods, algorithmic modifications, and ensemble methods [21]. In this paper, we use cost-sensitive learning (an algorithmic approach) and an ensemble method which are particularly useful in imbalanced classification problems.

1) COST-SENSITIVE LEARNING

It is an algorithmic method where we specify different misclassification costs for instances belonging to different classes. Class-specific weights are computed in (5).

$$W_i = \frac{N}{C * n_i} \quad (5)$$

where W_i denotes the weight for class i, N denotes the total number of instances, C denotes the total number of classes, and n_i denotes the number of instances belonging to the class i.

2) ENSEMBLE METHOD

This approach incorporates the strengths of random undersampling (RUS) and bagging [21]. RUS is a form of data sampling that randomly selects majority class instances and removes them from the dataset until the desired class distribution is achieved. In this way, several balanced training subsets are created by RUS of the majority class. Each subset contains all the minority class instances and an equal number of randomly selected majority class instances. The number of training subsets i.e. M can be chosen equal to the imbalance ratio [21]. In this way, we train M different models and aggregate their output using the majority voting approach to determine the final prediction.

C. INTERPRETING THE TRAINED GRUNN MODEL

The interpretability of AI models is a challenging problem that has been gaining increasing attention for the last few years. In this paper, we consider a gradient-based technique [16], [22] to interpret the trained model. The advantages of this technique over other methods (e.g., shapely values) are the ease of implementation and faster processing time.

The backpropagation-based approach [16], [22] is used to compute the attributions for all input features in a single forward and backward pass through the network and adapted for our use. Given a single target output, the goal is to determine the contribution of each input to the output. Let's define N are total instances present in the test dataset and the input is of shape $(T \times F)$. Here, T denotes the total number of time steps and F denotes the total number of features. Equation 6 presents the contribution of input x_{tf}^n to the output $S(x^n)$ for a single instance n.

$$g_{tf}^n = \left| \frac{\partial S(x^n)}{\partial x_{tf}^n} \right| \quad \forall n \in N,\ t \in T,\ f \in F \quad (6)$$

where g_{tf}^n denotes attributions that are of the same shape as that of input x_{tf}^n.

Then, we run over all the instances present in the test dataset and compute the global-level attribution g_{tf} using (7).

$$g_{tf} = \frac{1}{N} \sum_{n=1}^{N} g_{tf}^n \quad \forall t \in T,\ f \in F \quad (7)$$

The matrix $G \in \mathbb{R}^{T \times F}$ is composed of g_{tf} elements that contain the average value of absolute gradient. A heat map or attribution map generated from the matrix G can reveal the dynamics behind the gap selection process of drivers. Higher values of the elements of the matrix G imply greater importance on the prediction output. Further, we can derive feature importance (G_f) and time-step importance (G_t) using (8) and (9).

$$G_f = \frac{1}{T} \sum_{t=1}^{T} g_{tf} \quad \forall f \in F \quad (8)$$
$$G_t = \frac{1}{F} \sum_{f=1}^{F} g_{tf} \quad \forall t \in T \quad (9)$$

III. DATA PREPARATION TO MODEL GAP SELECTION USING GRU NEURAL NETWORK MODELS

This section is composed of three parts. The first part presents the characteristics of the trajectory dataset. Afterward, the second section describes the feature set used to build gap selection models. Finally, the third section elaborates on creating datasets for training and testing the models.

FIGURE 2. Locations where the trajectory dataset is collected (each location comprises two bottleneck sites on each side of a bi-directional motorway).

A. TRAJECTORY DATA

In this paper, a trajectory dataset is used to develop the gap selection decision models for multiple vehicle classes [17]. This dataset is a freely available resource that has been used in research before to understand driving behavior [18], [23]. The data comprise vehicle trajectories obtained through aerial imaging in the vicinity of 14 infrastructural bottlenecks located in the Netherlands, which include 3 on-ramps, 3 off-ramps, and 8 weaving sections (see Fig. 2).

This dataset was collected using a high-resolution camera attached to a hovering helicopter. The sites represent isolated discontinuities and their lengths are at most 1100 m meaning that the trajectories can be captured using the helicopter method. For each site, 30 min of the video feed was collected at the onset of evening congestion, that is, between 14:00 and 17:00 h. These 14 bottleneck sites included in this dataset sites either have a three-lane or a two-lane mainline carriageway with one auxiliary lane. For further information about the data collection, the reader may refer to van Beinum *et al.* [18]. Note that traffic operates under keep-right regulations in the Netherlands. Further, trucks are forbidden to drive on the left lane on carriageways with more than 2 lanes, except in the case of 2×2 weaving sections (where they are allowed to drive on the left-most lane).

The trajectory dataset is processed as follows.

1. We label vehicles shorter than 5.6 m as cars and longer than 12 m as trucks. Vehicles that fall in between are labeled as delivery vans.
2. Vehicles traveling on mainline carriageways are considered; vehicles entering onto a mainline carriageway from an on-ramp or exiting a mainline carriageway through an off-ramp are not considered since these two maneuvers fall under mandatory lane changing.
3. Vehicles making multiple lane changes on the mainline carriageway are excluded since these are more likely to be mandatory lane changes. Similar consideration is also made in a previous study [5].
4. Only vehicles changing lanes from right to left on mainline carriageways are considered. Vehicles changing lanes to the right side are excluded because traffic operates under keep-right regulations in the Netherlands.
5. In this way, two types of trajectories are considered: lane keeping (drivers do not change lanes) and lane changing (drivers change lanes to their left).
6. The trajectory data contain noisy speed and acceleration estimates. We use a Savitzk–Golay filter [26] to improve these estimates.
7. Concerning lane keeping trajectories, only trajectories pertaining to drivers who show normal or relaxed car following behavior are included. Any aggressive car following behavior, if detected, is used to exclude that trajectory. Aggressiveness in the car-following behavior is assessed through a time-to-collision-based (TTC) indicator. The aggressiveness is assumed to be present if the TTC value is less than 4 s, which is also used in previous research [24], [25].
8. To account for the effect of historical information, a maximum span of 20 s is considered to analyze the gap selection decision process since more than 95% of truck drivers are observed to changes lanes within the first 20 s in the trajectory dataset.
9. For every vehicle in the trajectory dataset, we have position (lateral and longitudinal coordinates), length, dynamics (speed, acceleration), relative measurements with the surrounding vehicles (distance gap and relative speed) available at a resolution of 10 frames/second.
10. Relevant data are sampled at a frequency of 0.5 s which means we collect two data points per second. The average parameter values for a sampling instant t are computed by averaging data for the interval $[t-0.4, t]$ [5]. The reasons for taking the average value over 0.5 s are a) to maintain consistency with the previous research [4, 5, 7, 27; (b) to reduce the error caused by using instantaneous values in the trajectory data; and (c) to be in line with driver's perception time [5].
11. The lane-changing process for a vehicle begins when it starts to drift laterally and ends when it stabilizes its lateral position after changing to a neighboring lane. The time instances are marked as lane change initiation and lane change completion, For a lane changing vehicle, the lane change initiation point is termed as T which refers to a relative increase in the lateral position of a vehicle with respect to time. 40 data points within a span of $[T-19.5, T]$ are considered for a lane changing trajectory as shown in Fig. 3.
12. For a lane keeping vehicle, T refers to the last point in its observed or recorded trajectory. As shown in Fig. 3, 40 data points are considered within a span of $[T-19.5, T]$.

A total of 3,647 trajectories of passenger car drivers are obtained out of which 2,803 are lane-keeping and 844 are lane-changing trajectories. For delivery van drivers, a total of 1,080 trajectories are extracted out of which 898 are lane-keeping and 182 are lane-changing trajectories. A total of

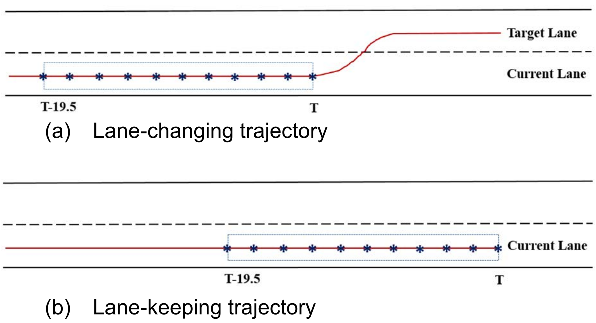


FIGURE 3. Figure showing span and data sampling for (a) lane-changing and (b) lane-keeping trajectories of drivers (Note: * denotes a data sampling instant at a resolution of 0.5 s.)

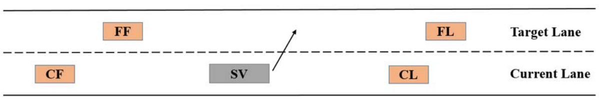


FIGURE 4. A driver during the gap-selection process (Note: SV: subject vehicle; CL: current leader; CF: current follower; FL: future leader; and FF: future follower).

2,226 trajectories of truck drivers are obtained out of which 2,103 are lane-keeping and 123 are lane-changing trajectories. Further, vehicle class-specific datasets suffer from class imbalance due to the presence of fewer lane-changing labels compared to the dominant lane-keeping labels.

B. FEATURE SELECTION

Trajectories can be viewed as sequences of decisions (lane-changing/lane-keeping) taken by drivers over time. At each time step, drivers can consider different features to make a decision. A typical gap selection scenario at a specific instant of time is presented in Fig. 4. Here, the subject vehicle (SV) is trying to move to the target lane from his current lane. During this process, the SV might be involved in interactions with up to four vehicles in its surroundings, as also considered by Balal *et al.* [5]. In the current lane, the SV interacts with its current leader (CL) and current follower (CF). Whereas in the target lane, its interactions are with the future leader (FL) and future follower (FF).

In this paper, we consider three dimensions that are hypothesized to affect this decision process. These capture the characteristics of the subject vehicle, its interaction with surrounding vehicles, and its perception of a topology (see Table 1). Typically only the first two dimensions are considered in previous research works [5]–[9].

Please note that it might not always be the case for the SV to be involved with four other vehicles during its lane changing process. For such situations where a surrounding vehicle is not observed or recorded in the trajectory dataset, we use a default value of 250 m for the distance gap spacing. A higher value such as 250 m also suggests that a vehicle is not affected by an unobserved surrounding vehicle. During the data collection, the camera captures more of the area than just the bottleneck section; therefore, 250 m seems to be a justified assumption in this respect. Similarly, for the speed of an unobserved surrounding vehicle, we assume its speed to be 0 m/s to compute the speed difference with the SV.

TABLE 1. Features describing vehicle's interaction during lane-changing.

Features	Unit	Description
Dimension 1: Characteristics of the subject vehicle		
v_{SV}	m/s	Speed of the vehicle SV
a_{SV}	m/s^2	Acceleration of the vehicle SV
l_{SV}	m	Length of the vehicle SV
Dimension 2: Interaction of the subject vehicle with surrounding vehicles		
d_{CL}	m	Distance gap between the vehicle SV and the vehicle CL
d_{CF}	m	Distance gap between the vehicle SV and the vehicle CF
d_{FL}	m	Distance gap between the vehicle SV and the vehicle FL
d_{FF}	m	Distance gap between the vehicle SV and the vehicle FF
Δv_{CL}	m/s	The speed difference between the vehicle SV and the vehicle CL, i.e. $v_{SV} - v_{CL}$
Δv_{CF}	m/s	The speed difference between the vehicle SV and the vehicle CF, i.e. $v_{SV} - v_{CF}$
Δv_{FL}	m/s	The speed difference between the vehicle SV and the vehicle FL, i.e. $v_{SV} - v_{FL}$
Δv_{FF}	m/s	The speed difference between the vehicle SV and the vehicle FF, i.e. $v_{SV} - v_{FF}$
t_{CL}	-	Type of the vehicle CL (truck or non-truck)
Dimension 3: Subject vehicles' perception of a topology		
n_{lanes}	-	Number of lanes on the mainline carriageway (2 or 3)
l_{top}	m	Length of the topology
t_{top}	-	Type of the topology (on-ramp, off-ramp, weaving section)

TABLE 2. Data split for vehicle classes.

Vehicle class	Label	Training dataset (80%)	Validation dataset (10%)	Test dataset (10%)
Passenger cars	Lane-changing	675	84	85
	Lane-keeping	2242	280	281
Delivery vans	Lane-changing	158	20	20
	Lane-keeping	705	88	89
Trucks	Lane-changing	98	12	13
	Lane-keeping	1682	210	211

C. CONSTRUCTING DATA FOR TRAINING AND TESTING

Having identified key features, we will now prepare data for our GRUNN model.

1) DATA SPLIT

We split the whole dataset into three parts: 80% training dataset, 10% validation dataset, and 10% test dataset. Table 2 shows the instances belonging to lane-keeping and lane-changing classes for every considered split and every vehicle class. The model is trained on the training dataset.

TABLE 3. Confusion matrix.

		Predicated class	
		Positive	Negative
Actual class	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

Hyperparameters of the model are tuned using the validation dataset. Finally, the performance of the model is tested on a test dataset.

2) FEATURE ENGINEERING

We consider 12 continuous and 3 categorical features (see Table 1). To support faster learning, the continuous features are standardized [28] where the values of each feature vector component x_i are centered around the mean with a unit standard deviation (see (10)):

$$\tilde{x}_i = \frac{x_i - \mu_{x_i}}{s_{x_i}} \tag{10}$$

where $\tilde{x}_i$ depicts the normalized feature vector component, and μ_{x_i} and s_{x_i} the mean and standard deviation of x_i.

Categorical features are converted into numerical forms via one-hot encoding [28], which represents categorical features in k possible categories as a binary feature vector of length k. The binary vector marks the class label with a value of 1 and all other positions with a value of 0. Consequently, a total of 19 features are considered in this paper. The target is a binary variable that comprises two labels: lane-changing (LC) and lane-keeping (LK). These labels are encoded as integer variables where 1 and 0 refer to lane-changing and lane-keeping, respectively. First, we learn a standardization function on the training dataset. Then, we transform validation and test datasets using the already learned standardization function to ensure that the model is not peaking at these two datasets.

3) PADDING THE TRAJECTORY DATA

When processing sequence data, it is common for individual samples to have different lengths. In our case, not all trajectories contain sufficient data for the desired span of 20 s. This is where padding is used to make all sequences in a batch of a given standard length (i.e., 40 time steps in our case) before one starts training the network. In this paper, trajectories are pre-padded so that they all are of the same size, i.e., 40 time steps.

IV. EXPERIMENTAL SETUP

This section begins by specifying the evaluation metric that is used to assess the model performance. Subsequently, the architecture of the proposed GRUNN model and its parameters that need to be optimized are presented.

A. EVALUATION METRIC

The confusion matrix is widely used to evaluate the performance of a classifier as shown in Table 3. For the binary classification, a confusion matrix is represented as a 2×2 matrix, which comprises four elements: TP (true positives), the number of correctly predicted positive instances; TN (true negatives), the number of correctly predicted negative instances; FP (false positives), the number of incorrectly predicted negative instances; and FN (false negatives), the number of incorrectly predicted positive instances.

These elements are used to derive traditional evaluation metrics such as accuracy $\left(= \frac{TP+TN}{TP+TN+FP+FN}\right)$. Having discussed previously that our dataset is an imbalanced dataset with more instances of lane-keeping (majority or negative class) than lane-changing (minority or positive class) ones, traditional evaluation metrics might provide biased results [29], [30]. Therefore, we consider geometric mean accuracy or G-mean that integrates recalls of both classes and is used in previous research to classify imbalanced datasets. G-mean can be expressed by (11):

$$\text{G} - \text{mean} = \sqrt{TPR \cdot TNR} \tag{11}$$

where $TPR \left(= \frac{TP}{TP+FN}\right)$ denotes true positive rate or accuracy on the minority class and $TNR \left(= \frac{TN}{TN+FP}\right)$ denotes true negative rate or accuracy on the majority class. G-mean tries to maximize the accuracy of each class while keeping these accuracy values balanced. Thus, a higher G-mean value indicates that the comprehensive performance of a classifier is better.

B. MODEL SPECIFICATION

The model is specified using several layers which take an input, i.e., a trajectory, and outputs the target; i.e., a label denoting either a lane-changing or lane-keeping decision. The following layers are considered in this paper.

1. Input layer: The input is of the shape (time steps, number of features). For this paper, the input is of the shape (40, 19).
2. Masking layer: A masking layer is added on top of the input layer so that model knows that missing time steps of an input should be skipped when processing the data. These missing time steps can be identified using the padded values which we have described earlier during data processing.
3. A gated recurrent (GRU) layer: A GRU layer followed by a dropout layer is added on top of the masking layer. The dropout layer is used to avoid overfitting and can subsequently improve the model generalization. The GRU layer can memorize previous information and feed the same to next time-steps using the activation function. In the case of a multi-layered GRUNN, more GRU layers can be stacked here. Each of which is followed by a dropout layer.
4. Dense layer: The output of the (final) GRU layer is collected at the latest time-step T using a dense layer. We use the Sigmoid activation function here to output the probability of classifying a trajectory as a lane-changing one.

C. SELECTION OF MODEL PARAMETERS

In this section, we discuss the model parameters of our GRUNN model that include hyperparameters of the GRUNN model and a choice of strategy to deal with class imbalance. A selection of optimal hyperparameters is shown to improve the GRUNN model performance [31]. Therefore, The following hyperparameters are considered in this paper.

1. The number of hidden layers: The multi-layer neural network architectures are shown to improve model performance and can achieve better realization than a single-layer architecture [32]. Therefore, we investigate the model performance with more than one hidden layer.
2. The number of units: The number of hidden units is a very important parameter of our GRUNN model, as the different number of hidden units may greatly affect the prediction precision. To choose the best value, we experiment with different hidden units and select the optimal value by comparing the predictions.
3. Dropout rate: Dropout is a technique to reduce overfitting. Its central idea is to take a model that is overfitting and train sub-models derived from it by randomly removing units for each training batch. The number of units to retain is controlled by a hyperparameter known as the dropout rate.
4. Learning rate: Learning rate is a parameter related to the optimization algorithm used while training the neural network. It controls how quickly the algorithm updates the weights at each iteration. A larger learning rate makes the model learn faster.
5. Batch size: A neural network is trained in batches. A batch is defined as the number of samples used for each iteration during the training process. Therefore, it is important to find the optimal batch size to achieve a good model performance.

Two strategies are used to handle the class imbalance problem in our case where the majority class (LK) contains 17 times more instances than the minority class (LC). The first strategy deals with class imbalance by assigning different classes with different weights, which are in proportion to their corresponding misclassification costs. The second strategy deals with training a model on a balanced dataset that has equal distribution of both majority (or LK) and minority (or LC) classes. An ensemble classifier is developed which aggregates the results by training models on several balanced datasets. The number of balanced datasets is equal to the proportion of instances in the majority class to the minority class.

In this paper, the grid-search method is used to search for the optimal values of parameters. As shown in Table 4, the tuning range is the range out of which the most appropriate value is selected. For each combination of hyperparameters and a choice of strategy to deal with class imbalance, we train the GRUNN on the training dataset. The proposed model uses the Adam algorithm [33] as our optimizer. The maximum number of epochs is set as 100 and the early stopping criterion is adopted to prevent overfitting. In this process, the training is stopped if its performance does not improve over 10 consecutive epochs. All the experiments are coded with Keras 2.4.0, TensorFlow 2.3.0, scikit-learn 0.24.1, NumPy 1.19.2, and pandas 1.2.3 in python 3.8.5.

TABLE 4. Parameter selection for GRUNN models.

Model parameters	Tuning range	Selected value: Passenger cars	Delivery vans	Trucks
		GRUNN-PC	GRUNN-DV	GRUNN-T
Hyperparameters of the GRUNN model				
Number of hidden layers	1, 2	2	2	2
Number of units	32, 64, 128	128	64	64
Dropout rate	0.1, 0.2, 0.3, 0.4, 0.5	0.1	0.1	0.3
Learning rate	0.01, 0.005, 0.001, 0.0005, 0.0001	0.001	0.01	0.0005
Batch size	32, 64, 128, 256, 512	256	128	64
Handling class imbalance				
Approach	Cost-sensitive learning, ensemble learning	Cost-sensitive learning (LC weight: 2.16, LK weight: 0.65)	Cost-sensitive learning (LC weight: 2.74, LK weight: 0.61)	Cost-sensitive learning (LC weight: 9.11, LK weight: 0.53)

Three separate GRUNN models are built: GRUNN-PC (passenger cars), GRUNN-DV (delivery vans), and GRUNN-T (trucks). After comparing the model's performance (or G-mean) on the validation dataset under different parameter values, the optimal values of hyperparameters are selected. Table 4 shows the selected hyperparameters for these three models.

V. RESULTS

Having identified the best model parameters, this section will focus on applying these models to respective test datasets. The performance of the trained GRUNN models (GRUNN-PC, GRUNN-DV, and GRUNN-T) are evaluated on the respective test datasets that have been kept aside. This section is further divided with respect to the vehicle classes considered in this paper. In each subsection, the prediction performance of models is discussed in the first part. After that, model interpretability (or explainable AI) techniques are used to explain what models have learned to gain insights into the gap selection process.

A. PASSENGER CARS

Table 5 shows the performance of the trained GRUNN model for passenger cars, i.e., GRUNN-PC. It can be observed that this model achieves the G-mean of 88.35%. This model

TABLE 5. **Model performance on the test dataset for passenger cars (GRUNN-PC).**

		Predicated class		Classification accuracy (%)
		LC	LK	
Actual class	LC	74	11	87.05
	LK	29	252	89.67

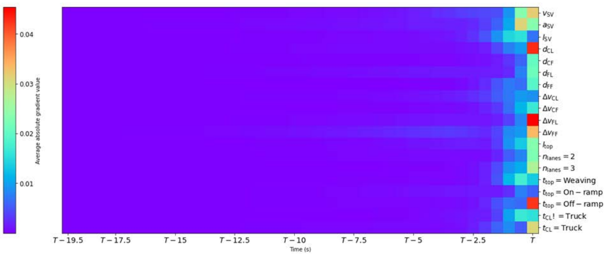
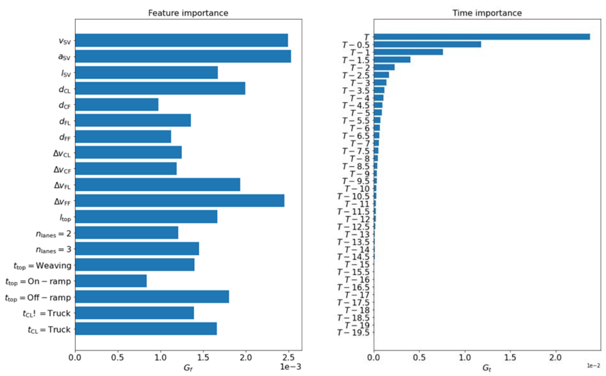


FIGURE 5. **Heat map showing the dynamic gap selection process of passenger cars.**



(a) Feature importance (b) Time importance

FIGURE 6. **Variable importance for passenger cars.**

can accurately predict 87.05% of LC and 89.67% of LK trajectories.

Now, we will use model interpretability (or explainable AI) techniques to discover new knowledge about the gap selection process of passenger cars. Fig. 5 shows the dynamics behind the gap selection process of passenger cars by using a heat map. This heat map is a 2D representation, consisting of average absolute gradient values, that can be used to determine what parts of the input contribute to the classification, and how important are these parts to the result. It can be observed that passenger car drivers seem to dynamically vary their attention over the feature set. For instance, they seem to consider their characteristics (speed and acceleration) to be more important than other features at the time instant $T-0.5$ s. On the contrary, at the time instant T, which is closer to the instance of their decision-making, they seem to shift their attention to their interactions with surrounding vehicles (distance gap of the vehicle SV with the vehicle CL (d_{CL}) and the speed difference between the vehicles SV and FL (Δv_{FL})) and topological related features (the type of topology (t_{top}= Off − ramp)).

Let us now turn to the variable importance, which captures its contribution to the target activation using gradient values. The higher the value of the gradient, the higher will be the contribution of that variable on the target activation. We consider both feature importance and time importance, at a global level, to explain the gap selection process of passenger car drivers (see Fig. 6). Looking at the feature importance, the three features that contribute most to the target activation are a passenger car driver's speed (v_{SV}), his acceleration (a_{SV}), and the speed difference with the vehicle FF (Δv_{FF}). Their reliance on their interactions with their respective future followers might indicate that they consider safety during lane changing. The feature importance plot suggests that topological features might not play a significant role in the gap selection process of passenger car drivers since their characteristics and their interactions with surrounding vehicles seem to dominate their gap selection process. If we now consider the time importance, it is observed that around three-fourths of the contribution can be captured by the time interval $[T-1.5, T]$. This suggests that passenger car drivers do not rely only on instantaneous information at T rather they also consider historical information when it comes to gap selection. Nevertheless, a large value of the average gradient at the time instance T indicates the passenger car drivers seem to place higher weights on the instantaneous information to decide whether to accept or reject gaps.

TABLE 6. **Model performance on the test dataset for delivery vans (GRUNN-DV).**

		Predicated class		Classification accuracy (%)
		LC	LK	
Actual class	LC	17	3	85.00
	LK	15	74	83.14

B. DELIVERY VANS

Table 6 presents the model performance for delivery vans. The GRUNN-DV model is able to achieve the G-mean of 84.06%. The accuracies with which this model can predict LC and LK trajectories are balanced. The model can accurately predict 85% of LC and 83.14% of LK trajectories.

Fig. 7 shows the dynamics behind the gap selection process of delivery van drivers by using a heat map. It can be observed that delivery van drivers also seem to dynamically vary their attention over the feature set similar to passenger car drivers. However, noticeable gradients for delivery van drivers encompass more time instants than passenger car drivers. This suggests that delivery van drivers utilize more historical information than passenger car drivers towards selecting gaps.

Let us now turn to the variable importance, which captures its contribution to the target activation using gradient values. We consider both feature importance and time importance, at a global level, to explain the gap selection process of delivery van drivers (see Fig. 8). Looking at the feature importance, the three features that contribute most to the target activation are the interactions of delivery van drivers with surrounding vehicles, captured by the distance gap with the vehicle CL (d_{CL}) and the vehicle FL (d_{FL}), and

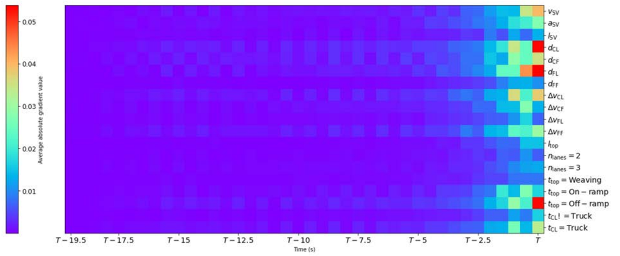
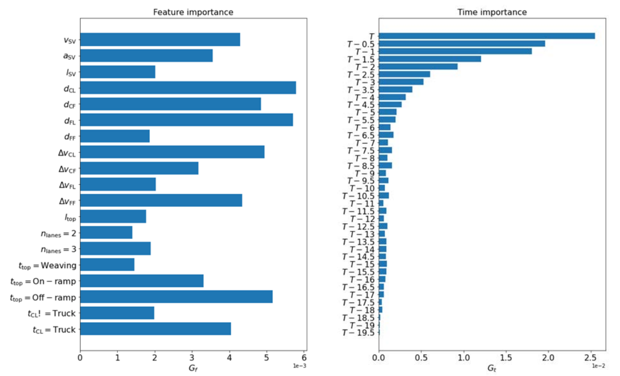


FIGURE 7. Heat map showing the dynamic gap selection process of delivery vans.



(a) Feature importance (b) Time importance

FIGURE 8. Variable importance for delivery vans.

type of topology (t_{top}= Off − ramp). The feature importance suggests that delivery van drivers consider their interactions and their perception of a topology more salient than their characteristics. Further, their reliance on their interactions with the current leader and future leader might indicate they consider them as a trade-off to evaluate traffic conditions and maneuverability on both lanes. If we now consider the time importance, it is observed that around three-fourths of the contribution can be captured by the time interval $[T-3, T]$. Interestingly, most of the effect is produced by the gradient values computed at the time instance T as also highlighted in the heat map. This suggests that delivery van drivers not only consider historical information from previous time-steps but also rely on current information to decide on the gap selection.

C. TRUCKS

Third, the performance of the model for trucks (GRUNN-T) is discussed in Table 7. This model can achieve the G-mean of 87% on the respective test dataset of trucks. Further, the model can accurately predict 92.31% of LC and 82% of LK trajectories.

Fig. 9 shows the dynamics behind the gap selection process of truck drivers by using a heat map. Truck drivers, similar to passenger car and delivery van drivers, dynamically vary their attention on the considered feature set. Yet, truck drivers differ in the manner how salient they consider historical information. The heat map in Fig. 9 indicates that they seem to be more anticipatory [34] than passenger car or delivery van drivers as noticeable gradient values cover more time span than passenger car and delivery van drivers.

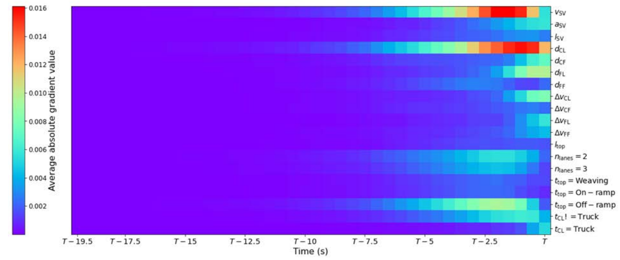
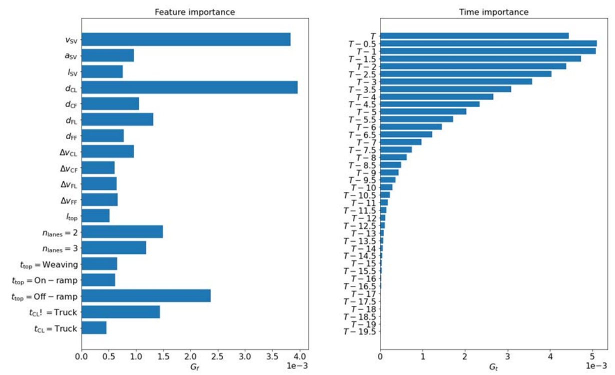


FIGURE 9. Heat map showing the dynamic gap selection process of trucks.



(c) Feature importance (d) Time importance

FIGURE 10. Variable importance for trucks.

TABLE 7. Model performance on the test dataset for trucks (GRUNN-T).

		Predicated class		Classification
		LC	LK	accuracy (%)
Actual class	LC	12	1	92.31
	LK	38	173	82.00

Moving to the variable importance (Fig. 10), the three features that contribute most to the target activation are the truck driver's speed (v_{SV}), the distance gap he maintains with the vehicle CL (d_{CL}), and the type of topology he is driving on (t_{top}= Off − ramp). These top three features also encompass all three dimensions (subject vehicle's characteristics, its interaction with surrounding vehicles, and its perception of the topology) considered while developing the feature set and thus showing entirely different behavior than both passenger car and delivery van drivers. Consequently, these features have much more impact on the classification score than other features. If we now consider the time importance, it is observed that around three-fourths of the contribution can be captured by the time interval $[T-6, T]$. This suggests that truck drivers do not rely only on instantaneous information at T rather they also consider previous time-steps or historical information when it comes to gap selection.

VI. DISCUSSION

The discussion section is composed of four parts. In the first part, the uniqueness of the trajectory dataset is described. Next, the gap selection behavior of multiple

TABLE 8. Comparison of the gap selection process of multiple vehicle classes.

Vehicle class	Salient time span (s)	Salient dimensions (features)		
		SV characteristics	Interaction of SV with surrounding vehicles	Perception of topology
Passenger cars	[T-1.5, T]	++	+	o
Delivery vans	[T-3.0, T]	o	++	+
Trucks	[T-6.0, T]	+	++	+

Note: SV: Subject vehicle;
Feature importance scale: ++ very important; + important; o less important

vehicle classes is compared. Then, the performance of our models (GRUNN-PC, GRUNN-DV, and GRUNN-T) is compared with the state-of-the-art. The last part discusses the implications and possible applications of our models.

This paper has used a large trajectory dataset that contains multiple vehicle classes. 3,647 trajectories of passenger car drivers, 2,226 trajectories of truck drivers, and 1,080 trajectories of delivery van drivers are present in this dataset. Unlike the widely used NGSIM dataset [35], our dataset contains significant numbers of trucks and delivery vans. This has a significant advantage over the earlier study focused on modeling lane-changing motivations of trucks [4], which used limited data of only 39 trajectories of truck drivers. Moreover, this dataset is also unique in terms of analyzing the behavior of delivery van drivers as other available trajectory datasets (e.g., NGSIM [35] and highD [36]) may not contain delivery vans.

Three GRUNN models are proposed in this paper: GRUNN-PC (passenger cars), GRUNN-DV (delivery vans), and GRUNN-T (trucks). These three models show that there exist significant differences with respect to the gap section process among these three vehicle classes (see Table 8). Passenger car drivers seem to be more concerned about their kinematic features and the motion of the lag vehicle in the target lane during gap selection. Whereas delivery van drivers give more weight to the traffic conditions in the current and target lane by looking at their gap spacing with the respective leading vehicles along with their perception of the topology. Truck drivers, on the other hand, consider a three-dimensional view that includes their vehicle kinematics (speed), their interactions with the surroundings (gap spacing with the current leader), and their perception of topology. Our findings demonstrate that topological factors are important to consider while analyzing the lane-changing behavior especially of commercial vehicles such as delivery vans and trucks. This fills a gap in the literature, as highlighted by Rahman *et al.* [37]. An advantage of our models is that they can be used on a general road network consisting of multiple topologies (e.g., on-ramps, off-ramps, and weaving sections). If we compare the time importance for multiple vehicle classes, we observe that gap selection is not an instantaneous process but a sequential one. This finding is consistent with previous research [5], [9], [13] which noted that passenger car drivers seem to consider lagged information for a specific time instant in their gap selection process. We show that the effect of memory or historical information is more salient for trucks than passenger cars or delivery vans. Especially, the last 6 seconds largely influence the gap selection process of truck drivers. Time importance plots also suggest that historical information generally has a fading effect on the gap selection process of vehicles which means that memory is not always constant. Previous research has also noted similar fading effects for the car following behavior [38].

This paper used imbalanced datasets to analyze the gap selection behavior of multiple vehicle classes. A large body of research is concentrated on passenger cars where most of these works built balanced datasets retrieved from widely used NGSIM data. In lieu of any reference study on the gap selection behavior of delivery van and truck drivers, the performance of our three models is compared with previous research on passenger car drivers which uses imbalanced datasets [5], [15], [39], [40]. These studies report G-mean in the range of 66.39-90.50% which is in accordance with the performances of our models. It is also encouraging to compare the performance of our models with other traffic-related studies [29], [30] on imbalanced classification which report G-mean in the range of 70.40-88.50%. It seems that the performance of our models (GRUNN-PC, GRUNN-DV, and GRUNN-T) is on par with these earlier studies.

Due to the dataset used, the empirical findings presented in this paper seem to be valid for drivers operating in the Netherlands. We expect that these findings may apply to all European countries with similar driving regulations. For other countries with different driving rules, further research is recommended so that the gap selection process of international drivers can be compared. This paper is an important step in improving our understanding regarding the microscopic phenomena of multiple vehicle classes that give rise to macroscopic or observable effects. Overall, these findings hold the potential to improve current models, to perform improved traffic and safety assessments, and eventually to support the design of advanced autonomous systems, for example aiming at guidance for the lane changing process. These advanced systems may use sensors instrumented in the subject vehicle or vehicle-to-vehicle (V2V) communication technology to feed inputs to the GRU neural network model. The input variables related to the perception of a road topology may be transmitted via digital mapping services.

VII. CONCLUSION

Gap selection is an important part of the discretionary lane changing activity. To understand and unravel the latent gap selection mechanisms of multiple vehicle classes, we use gated recurrent unit neural network (GRUNN) models on a large and unique trajectory dataset, collected for the Netherlands, that comprises lane changing trajectories of passenger cars, delivery vans, and trucks. The proposed vehicle class specific models are able to handle the class imbalance observed in the trajectory dataset. Moreover, these models can capture temporal interdependencies, by incorporating historical information, and the effect of external factors arising from the perception of a road topology.

The proposed models are interpreted using explainable AI techniques in order to obtain insights into the gap selection process of multiple vehicle classes. We show that gap selection is a sequential process governed by the impact of historical information or decisions and this impact fades over time. Passenger cars and delivery vans mostly utilize up to 3 seconds of recent driving experiences towards selecting gaps in contrast to trucks which rely on a longer duration of nearly up to 6 seconds. Using feature importance, we find that the most important factors associated with gap selection differ by vehicle classes. Passenger cars focus on their kinematic features (speed, acceleration) and their interactions with surrounding vehicles (the speed difference with the lag vehicle in the target lane). Whereas delivery vans utilize their interactions (gap spacing with current and future leading vehicles) and the type of topology. Trucks, on the other hand, consider a three-dimensional view that includes their kinematics (speed), their interactions with the surroundings (gap spacing with the current leading vehicle), and their perception of topology during their gap selection process.

The issue of lane-changing is an intriguing one that could be usefully explored in further research by using recent advances in AI and newly available trajectory datasets. Future research might explore the lane-changing execution of multiple vehicle classes. More broadly, a further study could also build integrated models, by also accounting for the inter-driver difference, to capture full lane-changing dynamics.

ACKNOWLEDGMENT

The authors would like to thank Dr. Taoufik Bakri from TNO for his inputs that greatly helped the model development.

REFERENCES

[1] M. J. Cassidy and J. Rudjanakanoknad, "Increasing the capacity of an isolated merge by metering its on-ramp," *Transp. Res. B, Methodol.*, vol. 39, no. 10, pp. 896–913, 2005, doi: 10.1016/j.trb.2004.12.001.

[2] Z. Li, S. Ahn, K. Chung, D. R. Ragland, W. Wang, and J. W. Yu, "Surrogate safety measure for evaluating rear-end collision risk related to kinematic waves near freeway recurrent bottlenecks," *Accident Anal. Prevention*, vol. 64, pp. 52–61, Mar. 2014, doi: 10.1016/j.aap.2013.11.003.

[3] X. Li and J. Q. Sun, "Studies of vehicle lane-changing dynamics and its effect on traffic efficiency, safety and environmental impact," *Phys. A, Statist. Mech. Appl.*, vol. 467, pp. 41–58, Feb. 2017, doi: 10.1016/j.physa.2016.09.022.

[4] S. Moridpour, M. Sarvi, G. Rose, and E. Mazloumi, "Lane-changing decision model for heavy vehicle drivers," *J. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 24–35, Jan. 2012.

[5] E. Balal, R. L. Cheu, and T. Sarkodie-Gyan, "A binary decision model for discretionary lane changing move based on fuzzy inference system," *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 47–61, Jun. 2016, doi: 10.1016/j.trc.2016.02.009.

[6] K. I. Ahmed, *Modeling Drivers' Acceleration and Lane Changing Behavior*. Cambridge, MA, USA: Massachusetts Institute of Technology, 1999.

[7] T. Toledo, H. Koutsopoulos, and M. Ben-Akiva, "Modeling integrated lane-changing behavior," *Transp. Res. Rec.*, vol. 1857, no. 1, pp. 30–38, 2003, doi: 10.3141/1857-04.

[8] M.-Y. Pang, B. Jia, D.-F. Xie, and X.-G. Li, "A probability lane-changing model considering memory effect and driver heterogeneity," *Transportmetrica B, Transp. Dyn.*, vol. 8, no. 1, pp. 72–89, Jan. 2020, doi: 10.1080/21680566.2020.1715310.

[9] D.-F. Xie, Z.-Z. Fang, B. Jia, and Z. B. He, "A data-driven lane-changing model based on deep learning," *Transp. Res. C, Emerg. Technol.*, vol. 106, pp. 41–60, Sep. 2019, doi: 10.1016/j.trc.2019.07.002.

[10] M. Saeednia and M. Menendez, "A consensus-based algorithm for truck platooning," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 404–415, Feb. 2017, doi: 10.1109/TITS.2016.2579260.

[11] J. A. Laval and C. F. Daganzo, "Lane-changing in traffic streams," *Transp. Res. B, Methodol.*, vol. 40, no. 3, pp. 251–264, Mar. 2006, doi: 10.1016/j.trb.2005.04.003.

[12] P. G. Gipps, "A model for the structure of lane-changing decisions," *Transp. Res. B, Methodol.*, vol. 20, no. 5, pp. 403–414, Oct. 1986, doi: 10.1016/0191-2615(86)90012-3.

[13] V. Punzo, M. T. Borzacchiello, and B. Ciuffo, "On the assessment of vehicle trajectory data accuracy and application to the next generation simulation (NGSIM) program data," *Transp. Res. C, Emerg. Technol.*, vol. 19, pp. 1243–1262, Dec. 2011, doi: 10.1016/j.trc.2010.12.007.

[14] T. Toledo and R. Katz, "State dependence in lane-changing models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2124, no. 1, pp. 81–88, 2009, doi: 10.3141/2124-08.

[15] L. Tang, H. Wang, W. Zhang, Z. Mei, and L. Li, "Driver lane change intention recognition of intelligent vehicle based on long short-term memory network," *IEEE Access*, vol. 8, pp. 136898–136905, 2020, doi: 10.1109/ACCESS.2020.3011550.

[16] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," in *Proc. Workshop Int. Conf. Learn. Represent.*, 2014, pp. 1–8.

[17] A. van Beinum, "Motorway turbulence—Empirical trajectory data," 4TU.ResearchData, Jun. 14, 2018, doi: 10.4121/uuid:6be1aefa-0803-4ce2-91b2-caaa7982abcd.

[18] A. van Beinum, H. Farah, F. Wegman, and S. Hoogendoorn, "Driving behaviour at motorway ramps and weaving segments based on empirical trajectory data," *Transp. Res. C, Emerg. Technol.*, vol. 92, pp. 426–441, Jul. 2018, doi: 10.1016/j.trc.2018.05.018.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997, doi: 10.1162/neco.1997.9.8.1735.

[20] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1–11.

[21] R. Choudhary and S. Shukla, "A clustering based ensemble of weighted kernelized extreme learning machine for class imbalance learning," *Expert Syst. Appl.*, vol. 164, Feb. 2021, Art. no. 114041, doi: 10.1016/j.eswa.2020.114041.

[22] H. van Lint, "Reliable travel time prediction for freeways: Bridging artificial neural networks and traffic flow theory," Ph.D. dissertation, Dept. Transp. Planning, Fac. Civil Eng. Geosci., Delft Univ. Technol., Delft, The Netherlands, 2004.

[23] S. Sharma, M. Snelder, L. Tavasszy, and H. van Lint, "Categorizing merging and diverging strategies of truck drivers at motorway ramps and weaving sections using a trajectory dataset," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 9, pp. 855–866, Sep. 2020, doi: 10.1177/0361198120932568.

[24] R. Horst, *Time-to-Collision as a cue for Decision-Making in Braking*. Amsterdam, The Netherlands: Vision in Vehicles–III, 1991.

[25] T. Faber, S. Sharma, M. Snelder, G. Klunder, L. Tavasszy, and H. van Lint, "Evaluating traffic efficiency and safety by varying truck platoon characteristics in a critical traffic situation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2674, no. 10, pp. 525–547, Oct. 2020, doi: 10.1177/0361198120935443.

[26] S. Ahn, S. Vadlamani, and J. Laval, "A method to account for non-steady state conditions in measuring traffic hysteresis," *Transp. Res. C*, vol. 34, pp. 138–147, Sep. 2013, doi: 10.1016/j.trc.2011.05.020.

[27] S. Siuhi and M. Kaseko, "Parametric study of stimulus-response behavior for car-following models. Paper 10–1179," in *Proc. 89th Annu. Meeting Transp. Res. Board Compendium Papers*, 2010. [Online]. Available: https://trid.trb.org/view/910163

[28] A. Zheng and A. Casari, *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. Sebastopol, CA, USA: O'Reilly Media, 2018.

[29] Y. Peng, C. Li, K. Wang, Z. Gao, and R. Yu, "Examining imbalanced classification algorithms in predicting real-time traffic crash risk," *Accident Anal. Prevention*, vol. 144, Sep. 2020, Art. no. 105610, doi: 10.1016/j.aap.2020.105610.

[30] Z. E. A. Elassad, H. Mousannif, and H. A. Moatassime, "A proactive decision support system for predicting traffic crash events: A critical analysis of imbalanced class distribution," *Knowl.-Based Syst.*, vol. 205, Oct. 2020, Art. no. 106314, doi: 10.1016/j.knosys.2020.106314.

[31] S. Zhang, M. Abdel-Aty, Y. Wu, and O. Zheng, "Modeling pedestrians' near-accident events at signalized intersections using gated recurrent unit (GRU)," *Accident Anal. Prevention*, vol. 148, Dec. 2020, Art. no. 105844, doi: 10.1016/j.aap.2020.105844.

[32] P. Utgoff and D. Stracuzzi, "Many-layered learning," *Neural Comput.*, vol. 14, no. 10, pp. 2497–2529, Oct. 2002.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015, pp. 1–15.

[34] Z. Zheng, "Recent developments and research needs in modeling lane changing," *Transp. Res. B, Methodol.*, vol. 60, pp. 16–32, Feb. 2014, doi: 10.1016/j.trb.2013.11.009.

[35] B. Coifman and L. Li, "A critical evaluation of the next generation simulation (NGSIM) vehicle trajectory dataset," *Transp. Res. B, Methodol.*, vol. 105, pp. 362–377, Apr. 2017, doi: 10.1016/j.trb.2017.09.018.

[36] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2118–2125, doi: 10.1109/ITSC.2018.8569552.

[37] M. Rahman, M. Chowdhury, Y. Xie, and Y. He, "Review of microscopic lane-changing models and future research opportunities," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1942–1956, Dec. 2013, doi: 10.1109/tits.2013.2272074.

[38] X. Huang, J. Sun, and J. Sun, "A car-following model considering asymmetric driving behavior based on long short-term memory neural networks," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 346–362, Oct. 2018, doi: 10.1016/j.trc.2018.07.022.

[39] C. Ng, S. Susilawati, M. A. S. Kamal, and I. M. L. Chew, "Development of a binary logistic lane change model and its validation using empirical freeway data," *Transportmetrica B, Transp. Dyn.*, vol. 8, no. 1, pp. 49–71, Jan. 2020, doi: 10.1080/21680566.2020.1715309.

[40] J. Zheng, K. Suzuki, and M. Fujita, "Predicting driver's lane-changing decisions using a neural network model," *Simul. Model. Pract. Theory*, vol. 42, pp. 73–83, Mar. 2014, doi: 10.1016/j.simpat.2013.12.007.

SALIL SHARMA received the B.Tech. degree in civil engineering from IIT Guwahati, in 2010, and the M.Sc. degree in transportation systems from the Technical University of Munich, Germany, in 2016. He is currently pursuing the Ph.D. degree in transportation engineering with the Delft University of Technology (TU Delft), The Netherlands. His research interests include traffic flow theory, traffic management and control, cooperative and intelligent transportation systems, and transportation and freight data analytics.

HANS VAN LINT is currently a Full Professor in traffic simulation & computing with the Transport and Planning (T&P) Department, Faculty of Civil Engineering and Geosciences, Delft University of Technology, where he is the Director of the Data Analytics and Simulation Laboratory (dittlab.tudelft.nl). He coauthored more than 70 articles in peer-reviewed international journals, supervise(s/d) 21 Ph.D. students, and served as an External Examiner in Ph.D. Committees for more than 30 other Ph.D. students in The Netherlands and abroad. His research interests include multi-scale estimation, prediction and simulation of traffic supply and demand on large transport networks using an interdisciplinary mix of traffic flow theory, AI, and computer science. In the last few years, he worked with his team for example on new OD estimation methods, traffic state estimation and prediction, and data driven methods to unravel large-scale network dynamics and—at the other end of the spectrum—on new microscopic models for driving behavior that include human factors.

He is on the Editorial Board of *Transportation Research Part C: Emerging Technologies*, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS (T-ITS), and involved in many international collaborations (projects, keynotes, and program boards) in the EU, USA, China, India, and Australia.

LÓRÁNT TAVASSZY (Member, IEEE) received the Ph.D. degree in civil engineering from the Delft University of Technology (TU Delft), in 1996. He is currently a Full Professor in freight transport & logistics systems with TU Delft. Between 1996 and 2016, he worked with the Dutch National Research Institute, TNO, as a Researcher, the Team Manager, and a Principal Scientist. He held the part-time Chair of the University of Nijmegen and TU Delft. His research involves freight transportation modeling at urban, national and global level. Besides the textbook *Modelling Freight Transport* he has published over 80 papers in various journals and over 150 articles in books and conference proceedings. He is also the Chair of the Scientific Committee of the World Conference for Transport Research Society (WCTRS) and is an active member of various international professional committees, including from the U.S. Transportation Research Board and the European Technology Platform for Logistics ALICE.

MAAIKE SNELDER received the degree in econometrics from Erasmus University, in 2003, and the Ph.D. degree *(cum laude)* in robust road network design from the Delft University of Technology, in 2010. She is currently a part-time Assistant Professor with the Department of Transport and Planning, Civil Engineering, Delft University of Technology and a Principal Scientist at TNO.