RESEARCH ARTICLE

Modeling Lane-Changing Behaviors for Autonomous Vehicles Based on Vehicle-to-Vehicle Communication

EUNTAK LEE [ID] 1, YOUNGJUN HAN [ID] 2, JU-YEON LEE 3, AND BONGSOO SON 1

[1]Department of Urban Planning and Engineering, Yonsei University, Seoul 03722, South Korea
[2]The Seoul Institute, Seoul 06756, South Korea
[3]Korea Transport Institute, Sejong 30147, South Korea

Corresponding author: Bongsoo Son (sbs@yonsei.ac.kr)

ABSTRACT With the advent of autonomous vehicles (AVs) and advanced driving assistance systems (ADAS), there has been a growing interest in studying driving behaviors within the field of transportation science. Given that the transition period of mixed traffic is expected to continue for more than 30 years, it is crucial to evolve AV technology to resemble human driving, especially in the freeway weaving sections. Lane-changing (LC) maneuvers in these sections could cause problems for traffic flow, such as traffic break-down, oscillation, or bottleneck activation. This study proposes an interpretable LC implementation model for naturalistic driving behaviors of AVs based on vehicle-to-vehicle (V2V) communication. To achieve this objective, a systematic selection process is adopted to find optimal V2V features that resemble how human drivers assess LC situations. Based on the minimum redundancy maximum relevance (mRMR) algorithm, seven V2V features have been selected out of 25 candidates. Then, a support vector machine (SVM) is employed to investigate how these features exhibit in each of LC and lane-keeping (LK) situations. The proposed model was applied in a field case of a weaving section on freeway US 101. Performance measures of simple accuracy, precision, recall, and F1-score show high accuracy of 0.9814, 0.9150, 0.7955, and 0.8511, respectively. Subsequently, a strategy for naturalistic LC behaviors of AVs was simulated. The proposed model outperforms high prediction accuracy compared to other existing models. Particularly, errors in the lateral movements have significantly improved. These results suggest that the proposed model effectively simulates naturalistic LC behaviors based on V2V communication.

INDEX TERMS Autonomous vehicles, vehicle-to-vehicle communication, lane-changing behavior.

I. INTRODUCTION

As vehicle driving systems have advanced, interest in autonomous vehicle (AV) technology based on wireless communication has increased. Most previous studies on wireless multi-access technology have been conducted to improve traffic safety and the travel comfort of passengers. While Internet-of-Vehicles (IoVs) technology has focused on the operational efficiency of the central control system [1], [2], [3], vehicle-to-vehicle (V2V) communication technology has concentrated on the driving strategy of individual AVs [4], [5], [6], [7], [8]. V2V devices, alongside existing advanced

The associate editor coordinating the review of this manuscript and approving it for publication was Huan Zhou [ID].

driver assistant systems (ADAS), identify different sensor measurements for the same target to ensure reliable connectivity [9]. They acquire and assess driving information for the decision-making process. However, V2V technology still has some limitations in sharing the driving intentions of surrounding vehicles unless they have V2V devices attached [10], [11]. This indicates that there might be a probability that AVs lack traffic information about surrounding vehicles during the transitional period of mixed traffic. Therefore, the technology must be developed that AVs to appropriately assess and respond within the insufficient information.

Moreover, AVs should be harmonized with surrounding human-driven vehicles (HVs) during the transitional period, where mixed traffic is expected to continue for more than

30 years. Unnatural driving behaviors of AVs can impact the movements of their surrounding HVs, consequently leading to an increase in traffic congestion and accident rates. Recent AV-related accident reports have also pointed out the potential for human drivers to misjudge unfamiliar driving behaviors exhibited by AVs [12]. Therefore, AV driving strategies should be developed to resemble HV driving behaviors for better traffic flow in the near future [13].

To address this challenge, an understanding of lane-changing (LC) behaviors must precede. LCs are usually performed by drivers seeking improved driving conditions, such as higher speeds, lower density, or to follow their specific routes. It is a complicated behavior since both longitudinal and lateral movements occur simultaneously. A simple LC by a vehicle can impact the movements of its surrounding vehicles on the driving and target lanes. If these maneuvers are recognized as unnatural behavior by surrounding vehicles and consequently disturb their driving patterns, traffic flow problems may occur, such as traffic breakdowns, oscillations, or queue propagation upstream [14], [15].

Accordingly, numerous studies have employed analytical approaches to investigate the backgrounds of how HVs implement LC maneuvers. For example, Park et al. [16] used a logistic regression model and found that LCs occur frequently during faster travel or when securing larger spacing. Similarly, Toledo et al. [17] suggested that the target lane is chosen based on higher average speed and lower density using a logistic regression model. Meanwhile, Lee et al. [18] developed an exponential probability model and found that LCs are more likely to occur when there are larger values of relative velocity and lead gap. Ahmed [19] developed a forced merging model using the maximum likelihood estimation method and presented that the merging process involves drivers' decisions regarding whether they intend to merge into an adjacent gap. Additionally, Wan et al. [20] investigated the longitudinal and lateral movements of vehicles during merging to gain a better understanding of the interactions between vehicles during LC processes. They developed a series of acceleration-deceleration models for the merging vehicle and validated them with US 101 observation data using a genetic algorithm. Kita [21] analyzed the interaction between merging and pass-through vehicles and modeled the cooperative behavior of pass-through vehicles. Likewise, studies with analytical approach have been conducted including cumulative prospect theory-based model [22] and utility theory-based models [23], [24], [25]. However, these analytical approaches have limitations in distinguishing between LC and lane-keeping (LK) situations due to driver heterogeneity, resulting in low accuracy in the classification results.

Subsequently, data-driven methods have gained attention as alternatives. Machine-learning or deep-learning algorithms, such as Convolutional Neural Networks (CNN) and Long Short-term Memory (LSTM), have been recently adopted to develop LC behavior models (for example, Lee et al. [26]; Xie et al. [27], respectively). Even though there have been numerous studies to investigate LCs, most of them have not demonstrated interpretability due to their black box nature [28].

Meanwhile, support vector machine (SVM) can enhance model performance while maintaining the interpretability of driving behaviors [13]. Mandalia and Salvucci [29] utilized SVM for LC detection and tested five input sets, each containing four to seven features (e.g., heading, lane position directly ahead of the vehicle, acceleration, and steering angle). Wang et al. [30] classified driving styles into aggressive and normal categories using semi-supervised SVM, revealing the nonlinear relationship between vehicle speed and throttle opening for each driving style. Ramyar et al. [31] used one-class SVM to distinguish normal and abnormal LC instances. They found that the most dangerous behaviors occur at the beginning and final stages of the LC process. Kumar et al. [32] applied SVM and Bayesian filtering to detect LC intention. Liu et al. [33] presented an LC decision-making model that outperformed rule-based LC models, addressing the multi-parameter and nonlinear aspects of the LC process.

However, some major challenges remain unsolved, although the abovementioned studies have significantly achieved improvements in LC classification performance. First, there exists a gap between the analytical and data-driven methods. They have strength at solid academic formulations and causal relationships within informal data, respectively. Finding connections between these two methodologies, which are somewhat in opposition, can contribute for developing LC model. Second, there is no standardized approach for finding optimal V2V features to describe the LC process, leading to the selection of different variables in previous studies. And most of the selected V2V features are highly detailed and rare data, which makes further study challenging. These features are not representative of a human perspective to fully understand HV driving behaviors. Third, the previous models mainly focused on the movements of subject vehicle itself, not considering its surrounding HVs (e.g., lane position 30m directly ahead of the vehicle and throttle openings). Driving conditions should be assessed more delicate to develop the V2V communication-based AV technology.

Indeed, AVs should behave similar to HVs during the transitional period for traffic safety and operations. Within the constraints that AVs have insufficient information for driving intentions of surrounding vehicles, the V2V devices should be fully utilized to relibly identify movements of surroundings and properly assess driving condition based on optimal features. Moreover, LC models need to be improved by developing analytical formulations from data-driven methodologies for interpretability.

Therefore, this study aims to develop an interpretable LC implementation model for the naturalistic driving behaviors of AVs based on V2V communication. The framework of this study is shown in Fig. 1. First, a systematic process for
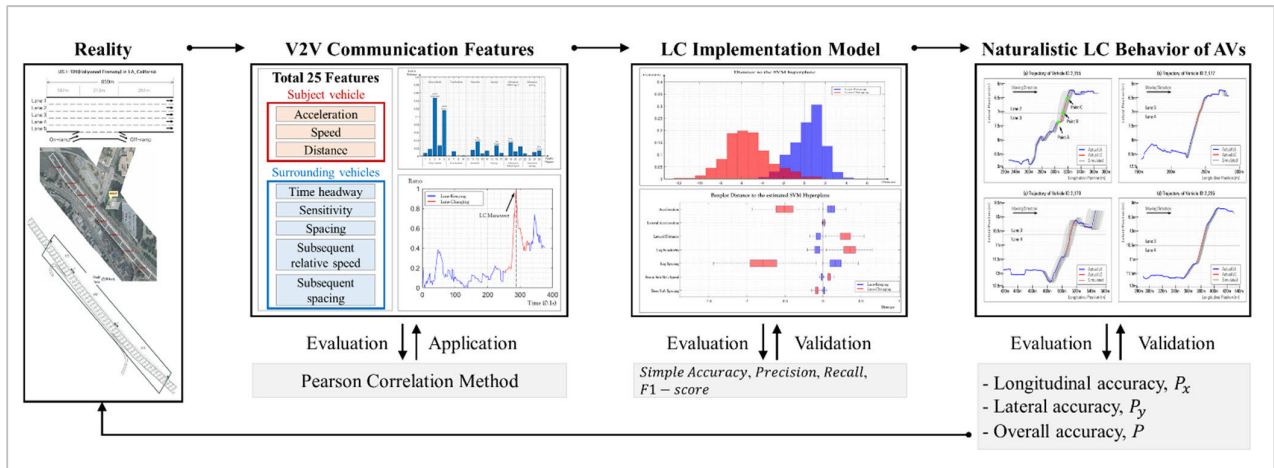
FIGURE 1. Framework of the study.

selecting optimal features for V2V communication is suggested. Based on the optimal V2V features, an LC classification model based on SVM is proposed to accurately identify the conditions of V2V features for each of LC and LK situations. Finally, the LC implementation model for AVs is proposed and then evaluated through simulations. The performance has shown higher accuracy in driving trajectories compared to existing models, indicating that our proposed model has achieved in designing LC behaviors of AVs to be harmonized with traffic flows. The main contributions are summarized as follows:

- This study contributes to the development of a naturalistic LC behavior model for AVs that harmonizes with surrounding HVs. The HVs should not misjudge AVs' unfamiliar driving behaviors.
- A standardized systematic process is established to find the optimal V2V features for describing the LC process. Seven features out of 25 have been selected that are most relevant to LCs with minumum redundancy.
- The proposed LC model provides interpretability using the SVM. The driving strategy is designed for AVs to target the V2V values in each LC and LK maneuver. This strategy informs that how far the implementations have gone through out of the entire LC process.
- The proposed model outperforms other existing models with high prediction accuracy. Particularly, errors in the lateral movements have significantly improved.

The remaining paper is organized as follows. Section II presents the systematic process of feature selection and introduces the LC implementation model using SVM. In Section III, a description of the field trajectory data NGSIM obtained from Federal Highway Administration (FHWA) is provided for the model's application. Section IV presents the evaluation of the model and provides detailed estimation results. Section V focuses on simulating strategies for naturalistic LC maneuvers in human-like AVs. Finally, Section VI concludes the paper with remarks and discussion.
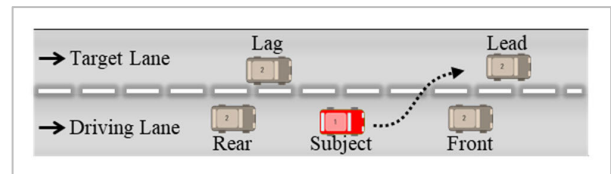

FIGURE 2. Subject vehicle and surrounding vehicles in the LC situation.

II. DEVELOPING THE LC IMPLEMENTATION MODEL

The LC decision-making process is conducted by the driver's judgement regarding the driving conditions of the subject vehicle and its surrounding vehicles. As shown in the problem statement of Fig. 2, a total of four surrounding vehicles are considered. Two front and rear vehicles in each of the driving and target lanes are included in this study.

A. PROBLEM STATEMENT

The LC implementation process comprises three steps: anticipation, execution, and relaxation. In the LC anticipation step, the subject vehicle provides preliminary motion signals before executing the LC maneuver, allowing its surrounding vehicles to recognize its intention. In the LC execution step, the subject vehicle visibly moves laterally from the driving lane to the target lane. In the LC relaxation step, it adjusts its driving behavior to the LK situation after completing the LC. Due to driver heterogeneity, it is not easy to clearly separate these three LC implementation steps with distinct boundaries. In this study, we assume that the LC implementation consists of the LC execution step only, which has a duration of 0.5 seconds before and after the moment the driving lane was changed. Therefore, the entire LC process is defined as consecutive steps of LK – LC – LK in this study. To secure the entire driving, the duration time is set as 15 seconds.

Meanwhile, there are two different types of LC: discretionary lane-changing (DLC) and mandatory lane-changing (MLC). While DLCs commonly occur on the mainline as vehicles pass through, MLCs are related to the behaviors of merging or diverging vehicles [19]. This study focuses on

DLCs due to the fact that MLC vehicles often forcefully change lanes and cause unreasonable traffic flow situations accordingly. Among DLCs, only the vehicles that changed to the left is selected. DLCs to the right are filtered since those executed on the application site might be part of MLC procedure to diverge downstream in following their routes.

B. V2V COMMUNICATION FEATURES FOR LC MANEUVER

To investigate level of significance for V2V communication features in the LC situations, a total of 25 features are considered in this study. These features are a combination of some from previous studies and others that have been newly processed, providing the model with both reliability and redundancy. The set of features is expressed in (1):

$$X_n(t) = \left\{ X_n^0(t), X_n^d(t), X_n^h(t), X_n^s(t), X_n^{\bar{v}}(t), X_n^{\bar{d}}(t) \right\} \tag{1}$$

where,

$X_n(t)$:   a set of V2V communication features of subject vehicle n at time t,

$X_n^0(t)$:   a set of driving information of subject vehicle n at time t,

$X_n^d(t)$:   a set of spacing of surrounding vehicles at time t,

$X_n^h(t)$:   a set of time headway of surrounding vehicles at time t,

$X_n^s(t)$:   a set of sensitivity to surrounding vehicles at time t, and

$X_n^{\bar{v}}(t)$:   a set of subsequent relative speed of surrounding vehicles at time t.

An input set, denoted as $X_n(t)$ at time t, consists of six subsets: one for the subject vehicle n, and five for the surrounding vehicles. The subset related to the subject vehicle includes driving information on longitudinal and lateral movements. The remaining five subsets related to the surrounding vehicles focus on longitudinal movement. The candidate V2V communication features for the target vehicles are summarized in Table 1.

TABLE 1. Candidate V2V communication features.

V2V Communication Features	Sub	Lead	Front	Lag	Rear
		Target Vehicles			
Longitudinal acceleration	✓	-	-	-	-
Longitudinal speed	✓	-	-	-	-
Lateral acceleration	✓	-	-	-	-
Lateral speed	✓	-	-	-	-
Lateral distance	✓	-	-	-	-
Time headway	-	✓	✓	✓	✓
Sensitivity	-	✓	✓	✓	✓
Spacing	-	✓	✓	✓	✓
Subsequent relative speed	-	✓	✓	✓	✓
Subsequent spacing	-	✓	✓	✓	✓

1) DRIVING INFORMATION OF SUBJECT VEHICLE

A set of variables for the subject vehicle ID n at time t, denoted as $X_n^0(t)$, consists of five V2V communication features, as shown in (2). The longitudinal information includes acceleration, $a_n(t)$, and speed, $v_n(t)$, while the lateral information includes acceleration, $\bar{a}_n(t)$, speed, $\bar{v}_n(t)$, and distance from the center of the lane, $\bar{d}_n(t)$. The units for acceleration, speed, and distance are set as [m/s^2], [m/s], [-], respectively. In this study, the distance is calculated as a ratio ranging from 0 to 1, taking into account the different lane widths. If the vehicle is on the center, the value is set to zero, and if the vehicle is on the lane marking, it is set to one.

$$X_n^0(t) = \{a_n(t), v_n(t), \bar{a}_n(t), \bar{v}_n(t), \bar{d}_n(t)\} \tag{2}$$

where,

$a_n(t)$:   longitudinal acceleration of vehicle ID n at time t,

$v_n(t)$:   longitudinal speed of vehicle ID n at time t,

$\bar{a}_n(t)$:   lateral acceleration of vehicle ID n at time t,

$\bar{v}_n(t)$:   lateral speed of vehicle ID n at time t, and

$\bar{d}_n(t)$:   lateral distance from the center of lane of vehicle ID n at time t.

2) TIME HEADWAY

A set of variables for the time headway between the respective surrounding vehicles and the subject vehicle ID n at time t, denoted as $X_n^h(t)$, is shown in (3). The time headway at time t, $h_{n,veh}(t)$, is calculated by dividing the spacing by actual speed of the following vehicle at time t, as shown in (4). The subject vehicle becomes the following vehicle when considering the lead or front vehicles ($veh = \{1, 2\}$). On the other hand, the surrounding vehicle becomes the following vehicle when considering the lag or rear vehicles ($veh = \{3, 4\}$). The unit of the time headway is set as [s].

$$X_n^h(t) = \{h_{n,1}(t), h_{n,2}(t), h_{n,3}(t), h_{n,4}(t)\} \tag{3}$$

where,

$$h_{n,veh}(t) = \begin{cases} d_{n,veh}(t) \div v_n(t), & veh = \{1, 2\} \\ d_{n,veh}(t) \div v_{veh}(t), & veh = \{3, 4\} \end{cases}, \tag{4}$$

$v_n(t)$: speed of vehicle ID n at time t, and

     $v_{veh}(t)$: speed of surrounding vehicle veh at time t.

3) SENSITIVITY

The sensitivity is directly related to drivers' reaction time. It has been found to be proportional to the relative speed but inversely proportional to the spacing [34]. It has been observed that when the relative speed is small and the spacing is sufficient, drivers seldom react to the surrounding vehicle or change their speed. A set of variables for the sensitivity between the respective surrounding vehicles and the subject vehicle ID n at time t, denoted as $X_n^s(t)$, is shown in (5). The sensitivity at time t, $s_{n,veh}(t)$, is obtained by dividing the relative speed, $\Delta v_{n,veh}(t)$, by the spacing, $d_{n,veh}(t)$, as shown in (6). The unit is set as [1/s]. Here, the relative speed of the surrounding vehicle can be calculated using (7).

$$X_n^s(t) = \{s_{n,1}(t), s_{n,2}(t), s_{n,3}(t), s_{n,4}(t)\} \tag{5}$$

where,

$$s_{n,veh}(t) = \Delta v_{n,veh}(t) \div d_{n,veh}(t), \, veh = \{1, 2, 3, 4\}, \tag{6}$$

$$\Delta v_{n,veh}(t) = \left| v_{n,veh}(t) - v_n(t) \right|, \, veh = \{1, 2, 3, 4\}, \text{ and} \tag{7}$$

$d_{n,veh}(t)$: spacing between surrounding vehicle *veh* and subject vehicle n at time t.

4) SPACING

A set of variables for the spacing between the respective surrounding vehicles and the subject vehicle ID n at time t, denoted as $X_n^d(t)$, is shown in (8). The spacing at time t, $d_{n,veh}(t)$, is calculated as the difference between the position values of the subject vehicle n and the surrounding vehicle $veh(=1,2,3,4)$, as shown in (9). The unit is set as [m].

$$X_n^d(t) = \{d_{n,1}(t), d_{n,2}(t), d_{n,3}(t), d_{n,4}(t)\} \tag{8}$$

where,

$$d_{n,veh}(t) = \left| p_{n,veh}(t) - p_n(t) \right|, \text{ and} \tag{9}$$

$p_{n,veh}(t)$: position of surrounding vehicle *veh* of vehicle ID n at time t.

5) SUBSEQUENT RELATIVE SPEED

The subsequent relative speed is defined as the estimated relative speed that will occur in the following time step. A set of variables obtained for the respective surrounding vehicles and the subject vehicle ID n at time t, denoted as $X_n^{\bar{v}}(t)$, is shown in (10). It is calculated as the difference between the subsequent speed values of the subject vehicle n and the surrounding vehicle *veh* (= 1, 2, 3, 4), as shown in (11). The unit is set as [m/s]. It is assumed that all driving information remains constant for a unit time Δt. The subsequent speed of the surrounding and the subject vehicles can be obtained using (12) and (13), respectively.

$$X_n^{\bar{v}}(t) = \{\bar{v}_{n,1}(t), \bar{v}_{n,2}(t), \bar{v}_{n,3}(t), \bar{v}_{n,4}(t)\} \tag{10}$$

where,

$$\bar{v}_{n,veh}(t) = \left| v_{n,veh}(t + \Delta t) - v_n(t + \Delta t) \right|, \tag{11}$$

$$v_{n,veh}(t + \Delta t) = v_{n,veh}(t) + a_{n,veh}(t) \times \Delta t, \tag{12}$$

$$v_n(t + \Delta t) = v_n(t) + a_n(t) \times \Delta t, \text{ and} \tag{13}$$

$a_{n,veh}(t)$: acceleration of surrounding vehicle *veh* of vehicle ID n at time t.

6) SUBSEQUENT SPACING

The subsequent spacing is defined as the estimated relative speed that will occur in the following time step. A set of variables obtained for the respective surrounding vehicles and the subject vehicle ID n at time t, denoted as $X_n^{\bar{d}}(t)$, is shown in (14). It is calculated as the difference between the subsequent position values of the subject vehicle n and

the surrounding vehicle *veh* (= 1, 2, 3, 4), as shown in (15). The unit is defined as [m]. The subsequent positions of the surrounding and the subject vehicles can be obtained using (16) and (17), respectively.

$$X_n^{\bar{d}}(t) = \{\bar{d}_{n,1}(t), \bar{d}_{n,2}(t), \dots, \bar{d}_{n,6}(t)\} \tag{14}$$

where,

$$\bar{d}_{n,veh}(t) = \left| p_{n,veh}(t + 2\Delta t) - p_n(t + 2\Delta t) \right|, \tag{15}$$

$$p_{n,veh}(t + 2\Delta t) = p_{n,veh}(t) + 2v_{n,veh}(t)\Delta t + a_{n,veh}(t)\Delta t^2, \text{ and} \tag{16}$$

$$p_n(t + 2\Delta t) = p_n(t) + v_n(t) \times 2\Delta t + a_n(t) \times \Delta t^2. \tag{17}$$

C. FEATURE SELECTION USING THE MINIMUM REDUNDANCY MAXIMUM RELEVANCE (mRMR) ALGORITHM

Selecting input features with a systematic process is a critical issue in developing the LC implementation model. The classification performance could be degraded if the selected features are numerous and highly correlated. Moreover, irrelevant input features can increase computational costs and lead to overfitting.

In this study, the minimum redundancy maximum relevance (mRMR) algorithm is applied to address these concerns. The mRMR algorithm aims to reduce training times, enhance generalization by reducing overfitting, and simplify the model for easier interpretation. It identifies a subset of features that are relevant to the target but not redundant with each other. In other words, the algorithm penalizes a feature's relevancy based on its redundancy in the presence of other selected features. By assigning feature weights, the algorithm identifies the significance of features related to LC maneuvers. The algorithm helps in selecting relevant and non-redundant features for the SVM model, contributing to improved classification performance and interpretability. The specific process of the mRMR algorithm is summarized in Table 2.

The first step is to set a set of m features $F = \{x_1, x_2, \dots, x_m\}$, and initialize an empty subset S that will contain the selected features. Let the number of selected features be k. Define $y_j (j = 1, 2)$ as the class labels for LC and LK, respectively. The second step is to calculate the relevance $R(x_i, y_j)$ between each feature $x_i (i = 1, \dots, m)$ and the class label y_j using a correlation coefficient or mutual information. In this study, mutual information is used to measure relevance. The third step is to find the first relevant feature that records the highest measure. Include this feature in the subset S and remove it from set F. The fourth step is to repeat the process for k times. The fifth step is to find the weight vector, which represents the importance of each feature and indicates its correlation with LC maneuvers.

TABLE 2. Pseudo algorithm of feature selection.

Step 1: Initial settings
$F = \{x_1, x_2, \ldots, x_m\}$: a set of m features.
$S = \emptyset$: a set of selected features.
$R = \emptyset$: a set of relevance index of the selected features.
k : the number of selected features.

for *iteration* = 1 : m features
  │   set m_F be the number of features in F
  │
  │   *Step 2*: compute relevance and redundancy of m_F features
  │   **for** $i = 1 : m_F$ features
  │ │   relevance, $I(x_i, y_j) = \sum_{i,j} P(X = x_i, Y = y_j) log P(X = x_i, Y = y_j)/P(X = x_i)P(Y = y_j)$
  │ │   redundancy, $W(x_i) = \frac{1}{|F|}\sum_{l \in F} P(X = x_i, Y = x_l) log P(X = x_i, Y = x_l)/P(X = x_i)P(Y = x_l)$
  │   **end for**
  │
  │   *Step 3*: find the next most important feature
  │   **if** *iteration* == 1
  │ │ $x_S = argmax_{x_i=1,\ldots,m_F} I(x_i, y_j)$
  │   **else**
  │ │ $x_S = argmax_{x_i=1,\ldots,m_F}\{I(x_i, y_j) \mid W(x_i) = 0\}$
  │ │   if all $W(x_i) \neq 0$, then $x_S = argmax_{x_i=1,\ldots,m_F} I(x_i, y_j)/W(x_i)$
  │   **end if**
  │
  │   *Step 4*: update sets F and S
  │ $F \leftarrow F - x_S$
  │ $S \leftarrow S + x_S$
  │ $R \leftarrow R + I(x_i, y_j)$
end for

Step 5: repeat process

D. DISTINGUISHING LC AND LK MANEUVERS USING SVM

SVM is an efficient method for estimating boundaries between different classes. In this study, the LC implementation model is proposed as a binary classification problem, distinguishing between LC and LK (hereafter, labeled as 1 and -1, respectively). The model separates the two classes in a multi-dimensional feature space, with the selected features being the ones that determine the LC. Using the selected features, $x_i'(i = 1, 2, \ldots, n)$, the point $\boldsymbol{x}' = (x_1', x_2', \ldots, x_n')$ represents either the LC or LK class. The basic elements for SVM, which determines the boundary, are expressed as follows.

$$\boldsymbol{\omega}^T \boldsymbol{x} + b = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + b = 0, \quad (18)$$

$$D(\boldsymbol{x}') = \frac{|\boldsymbol{\omega}^T \boldsymbol{x}' + b|}{||\boldsymbol{w}||^2}$$
$$= \frac{|w_1 x_1' + w_2 x_2' + \ldots + w_n x_n' + b|}{\sqrt{w_1^2 + w_2^2 + \ldots + w_n^2}}, \text{ and} \quad (19)$$

$$\rho = \frac{2}{||\boldsymbol{\omega}||^2}. \quad (20)$$

Consider a boundary, or hyperplane, represented as a linear classifier in (18). $\boldsymbol{\omega}(= (w_1, w_2, \ldots, w_n))$ is the normal vector (gradient) of the hyperplane, and b represents a bias. The hyperplane separates the n-dimensional space into two regions for LC and LK situations. Equation (19) calculates the shortest distance D from point $\boldsymbol{x}'$ to the hyperplane, known as the "classification margin." The points closest to the hyperplane, called "support vectors," are critical in estimating the LC classification boundary. Equation (20) introduces ρ, which represents the margin of the hyperplane and scales the distance between the support vectors. Maximizing the margin is necessary to achieve clear classification between LC and LK situations. Thus, the objective function and constraints for SVM are expressed as follows.

$$\text{Minimize F} = (||\boldsymbol{\omega}||^2 + C \sum_i^N \xi_i) \quad (21)$$

Subject to,

$$y_i \left(\boldsymbol{\omega}^T \boldsymbol{x_i} + b\right) \geq 1 - \xi_i, \text{ for i} = 1, \ldots, N, \text{ and} \quad (22)$$

$$\xi_i = \max\left\{0, 1 - y_i\left(\boldsymbol{\omega}^T \boldsymbol{x_i} + b\right)\right\} \geq 0,$$
$$\text{for i} = 1, 2, \ldots, N. \quad (23)$$

The objective function that optimizes the boundary is formulated in (21), and it consists of two components. The first component minimizes the inverse number of the margin, thereby maximizing the margin. The second component minimizes the penalty for misclassified samples. The penalty is calculated multiplying the cost, C, by the margin of misclassified LC maneuver sample i, ξ_i. Typically, the cost is set to one. Equation (22) presents a constraint on N samples. The term $y_i = \{-1, 1\}$ indicates the LC decision of sample i. It is set as 1 when the point of sample i is above the hyperplane and classified as LC; otherwise, it is set as -1. Similarly, $(\boldsymbol{\omega}^T \boldsymbol{x_i} + b)$ is larger than 1 if the point of sample i is above the boundary, less than -1 otherwise. The multiplication of these two terms always yields a value greater than 1 for proper classification. However, if a sample is misclassified, margin ξ_i is subtracted on the right-hand side to satisfy the constraint. In (23), the margin ξ_i represents a loss function, calculated as the maximum value between zero and the term $(1 - y_i(\boldsymbol{\omega}^T \boldsymbol{x_i} + b))$. This term is below zero for properly classified samples and positive for misclassified ones. Thus, the margin is calculated only for misclassified samples, with the margin for properly classified samples set to zero.

Solving the objective function generally requires significant computational power. Therefore, the dual problem is applied to transform it into a simple form [35]. After algebraic calculation, the transformed objective function and constraints are obtained as follows.

$$\max_{\alpha} \sum_i^N \alpha_i - \frac{1}{2}\sum_i^N \sum_j^N \alpha_i \alpha_j y_i y_j K(x_i x_j) \quad (24)$$

subject to,

$$\sum_j^N y_j \alpha_j = 0, \quad (25)$$

$$0 \leq \alpha_j \leq C, \tag{26}$$

$$K\left(x_i x_j\right) = \exp\left(-\frac{\left|\left|x_i - x_j\right|\right|_2^2}{2\sigma^2}\right), \text{ and} \tag{27}$$

$$\boldsymbol{\omega} = \sum_i^N \alpha_i y_i \boldsymbol{x}_i. \tag{28}$$

The transformed objective function in (24) aims to find a matrix of Lagrangian multipliers $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ that maximizes the function. Equations (25) and (26) represent stationary and feasibility conditions of the Karush-Kuhn-Tucker method, respectively. In this study, the Gaussian kernel function K is applied, as shown in (27). Consequently, the gradient $\boldsymbol{\omega}$ of the optimal hyperplane is obtained in (28) by multiplying the estimated $\boldsymbol{\alpha}$, with the LC decision class and driving conditions.

III. DATA DESCRIPTION

The proposed LC implementation model was evaluated using field data from Next Generation Simulation (NGSIM), an open dataset provided by the Federal Highway Administration (FHWA). NGSIM data consist of detailed vehicle trajectory data recorded by digital cameras every tenth of a second at the freeway weaving section of US 101 in California. The recording spans a 45-minute period during the peak morning hour, from 07:50 AM to 08:35 AM, showing traffic congestion with an average travel speed of around 30 km/h. The surveillance site has a geometric length of 650 m, comprising an upstream section of 187 m, a weaving section of 213 m, and a downstream section of 251 m. The freeway consists of 5 main lanes and 1 auxiliary lane, as depicted in Fig. 3.
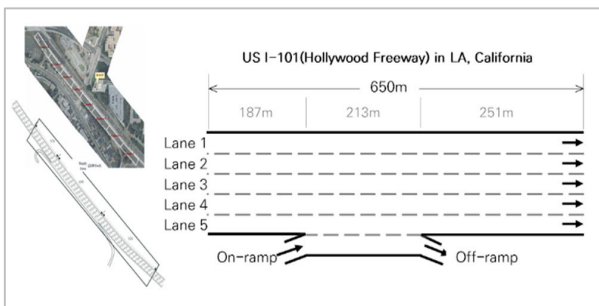

FIGURE 3. Geometric configuration of US101 freeway weaving section.

IV. ESTIMATION RESULTS

In this section, we present the estimation results of the proposed model in three steps: (i) the mRMR algorithm for selecting optimal V2V communication features; (ii) the SVM model to distinguish the characteristics between the LC and LK situations; and (iii) performance of the proposed model.

A. ESTIMATED RESULTS OF FEATURE SELECTION

1) CORRELATION BETWEEN V2V FEATURES

Even though the mRMR algorithm is efficient for feature selection, it has limitations when considering correlations

among the selected features. In this study, the authors adopt the Pearson correlation method to estimate the level of correlation between the 25 features, thereby complementing the mRMR algorithm.

Fig. 4 shows the estimated correlation values, represented by gradient colors. The correlations between subsequent relative speed features (#18 to #21) and sensitivity features (#10 to #13) are all estimated to be over 0.5, indicating a strong correlation between them. For features #3 and #4 (lateral acceleration and lateral speed, respectively), the correlation value is estimated to be near -1, suggesting a high negative correlation. Conversely, the correlations with other features are close to zero or exhibit low correlation.
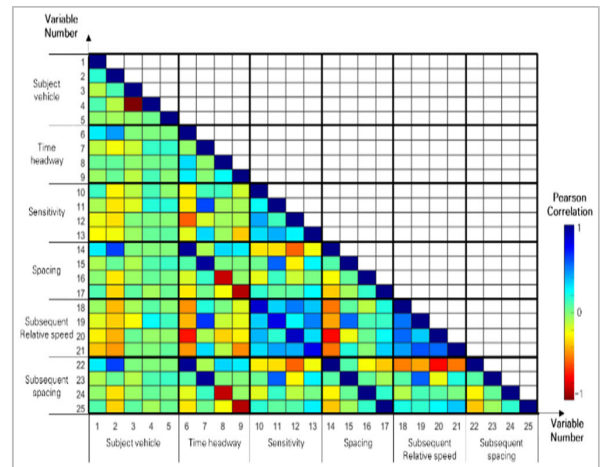

FIGURE 4. Pearson correlations between V2V communication features.

The results of the Pearson correlation analysis are then incorporated into the feature selection process to minimize overlapping effects. Generally, a correlation with an absolute value over 0.5 is considered strongly correlated. Since the mRMR algorithm ranks the importance of features, we select the features with high ranks and low correlation in the end.

2) ESTIMATED WEIGHTS OF V2V COMMUNICATION FEATURES USING MRMR ALGORITHM

The estimated weights of V2V communication features using the mRMR algorithm are shown in Fig. 5 and Table 3. Among the five features related to the subject vehicle, lateral acceleration (#3) and lateral distance (#5) are estimated to have weights of 0.1466 and 0.1153, respectively. This indicates that they are the most significant features for distinguishing LC and LK maneuvers. Regarding time headway, the feature of the front vehicle is ranked twelfth and has the most influence on LC implementation among the surrounding vehicles, while the features of other vehicles rank below twentieth. Due to the homogeneity of travel speed under traffic congestion, time headway seems to have little impact compared to other feature subsets. For sensitivity, the weights of the front and lag vehicles are estimated to be ninth and third, respectively. As the value of sensitivity increases with shorter spacing and larger speed differences, these results
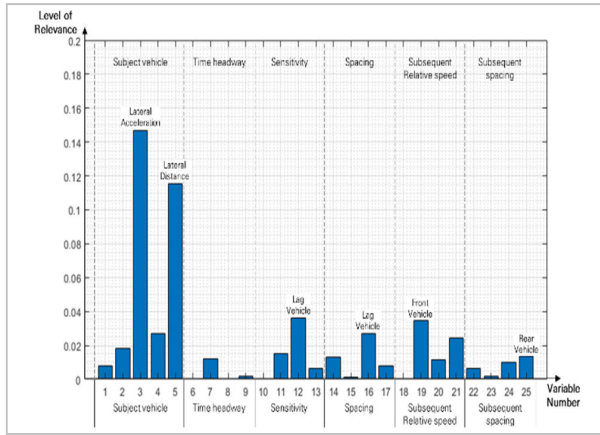
FIGURE 5. V2V communication features with the estimated weights.

TABLE 3. V2V communication features with estimated weights.

V2V communication features	Target vehicle	Estimated weight	Rank	Selected
Longitudinal acceleration	1. Subject vehicle	0.0079	15	✓
Longitudinal speed	2. Subject vehicle	0.0185	8	-
Lateral acceleration	3. Subject vehicle	0.1466	1	✓
Lateral speed	4. Subject vehicle	0.0270	6	-
Lateral distance	5. Subject vehicle	0.1153	2	✓
Time headway	6. Lead vehicle	3.1e-14	25	-
	7. Front vehicle	0.0122	12	-
	8. Lag vehicle	3.5e-14	24	-
	9. Rear vehicle	0.0019	20	-
Sensitivity	10. Lead vehicle	6.1e-14	23	-
	11. Front vehicle	0.0152	9	-
	12. Lag vehicle	0.0364	3	✓
	13. Rear vehicle	0.0063	18	-
Spacing	14. Lead vehicle	0.0134	11	-
	15. Front vehicle	0.0014	21	-
	16. Lag vehicle	0.0271	5	✓
	17. Rear vehicle	0.0078	16	-
Subsequent relative speed	18. Lead vehicle	8.9e-14	22	-
	19. Front vehicle	0.0345	4	✓
	20. Lag vehicle	0.0115	13	-
	21. Rear vehicle	0.0242	7	-
Subsequent spacing	22. Lead vehicle	0.0064	17	-
	23. Front vehicle	0.0020	19	-
	24. Lag vehicle	0.0101	14	-
	25. Rear vehicle	0.0134	10	✓

speed, the front vehicle is regarded as the most significant, as DLC is implemented for higher travel speeds. As for subsequent spacing, the rear vehicle is estimated to have the most impact.

The feature selection process is conducted based on three conditions: (i) features are selected in the order of ranks; (ii) a feature excluded if it is strongly correlated with at least one of the previously selected features; (iii) the process stops when the cumulative sum of weights of the selected features exceeds 50% of the total sum.

As a result, seven V2V communication features with ranks 1, 2, 3, 4, 5, 10, and 15 are selected in this study. Features with ranks ranging from 6 to 9 and 11 to 14 are excluded as they exhibit strong correlations with the initially selected features during the process.

The results of the feature selection can be interpreted with respect to the target vehicles. First, for the subject vehicle, three features are selected: longitudinal acceleration, lateral acceleration, and lateral distance (#1, #3, and #5, respectively), while speed-related features are not chosen. This suggests that acceleration plays a better role in distinguishing between LC and LK maneuvers compared to speed. The apparent representation of lateral movement through lateral distance from the center of the lane differentiates LC from LK situations. Second, for the front vehicle, subsequent relative speed (#19) is selected. Vehicles change lanes to achieve better driving conditions, such as higher speed or lower density. This indicates that the subject vehicle accelerates to move faster than the front vehicle after implementing the LC. Third, for the lag vehicle, sensitivity (#12) and spacing (#16) are selected. It appears that the subject vehicle is aware of the driving conditions of the lag vehicle, which can significantly influence the LC decision-making process. The low Pearson correlation value suggests that both the speed difference and spacing of the lag vehicle undergo significant changes between LC and LK situations. Lastly, for the rear vehicle, subsequent spacing is selected. The driving conditions of the rear vehicle change as the subject vehicle accelerates after executing the LC, similar to the case of the front vehicle. Interestingly, features related to the lead vehicle are not selected. This indicates that the driving conditions of the lead vehicle do not exhibit significant differences between LC and LK situations. In other words, there are situations where vehicles do not change lanes even though the conditions of the lead vehicle are suitable for LC execution.

Fig. 6 shows the changes in the selected feature values over time, accompanied by the trajectory of vehicle ID 1,147. The blue lines represent feature values in the LK situation, while the red lines indicate feature values in the LC situation. Note that the LC situation consists of anticipation, execution, and relaxation steps. In Fig. 6(a), prior to the LC implementation, the value of lateral distance from the center of the lane remains below 0.4. As the vehicle approaches a lane line and executes the LC maneuver around the time frame 290, the value increases to 1.0. After the execution, the value in the changed lane decreases below 1.0 as the
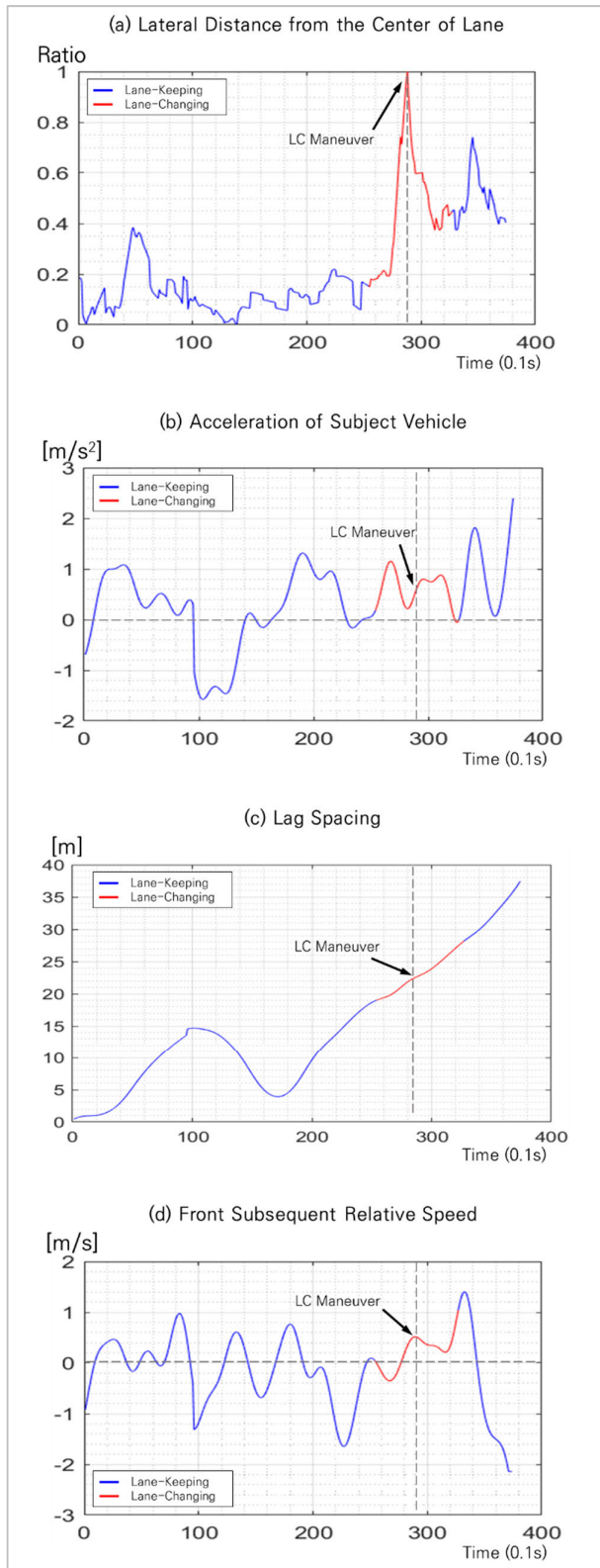
suggest that the subject vehicle drives close to the front or lag vehicles with a larger speed difference compared to the lead or rear vehicles. For spacing, the subject vehicle considers the lag vehicle to be the most influential vehicle in the LC decision-making process. In terms of subsequent relative

FIGURE 6. Values of the selected features along the trajectory of vehicle ID 1,147.

vehicle adjusts its position within the lane. Furthermore, as shown in Fig. 6(b), the acceleration value of the subject vehicle exhibits random fluctuations during the LK situation.

However, throughout the entire LC process, the vehicle accelerates, and this acceleration continues even after the maneuver. Fig. 6(c) demonstrates that the value of lag spacing steadily increases before, during, and after the LC implementation. It is noteworthy that the lag vehicle is still tracked even though it is positioned behind the subject vehicle after the LC maneuver. The increase in value indicates that the subject vehicle speeds up after executing the LC, as DLC is implemented to create a better driving environment. Similarly, Fig. 6(d) shows the patterns in the value of front subsequent relative speed, which aligns with the values of lag spacing. The front vehicle is still tracked even though it being in the lead position after the LC maneuver. The subject vehicle accelerates after the execution, causing the value of relative speed decrease as time progresses.

B. ESTIMATED RESULTS OF SVM MODEL
The SVM model is developed based on the selected features. The hyperplane of the model is estimated as a binary linear classifier that distinguishes the characteristics between the LC and LK situations. Since the SVM hyperplane separates the two classes in a multi-dimensional feature space, the distance of each LC and LK data from the hyperplane is calculated using (19). Note that all feature data is normalized.

Fig. 7 displays the probability distributions of the distances to the SVM hyperplane for LC and LK situations. In the case of LC situations, the distance range is estimated to be from -12 to 2. The average and standard deviation of the distance distribution in LC situations are estimated to be -6 and 2, respectively. Most of the LC data is classified properly as LC, as they are located on the negative side of the boundary. Shape of the distribution of LC data is wide and symmetrical, indicating a predictable pattern in the LC maneuver.
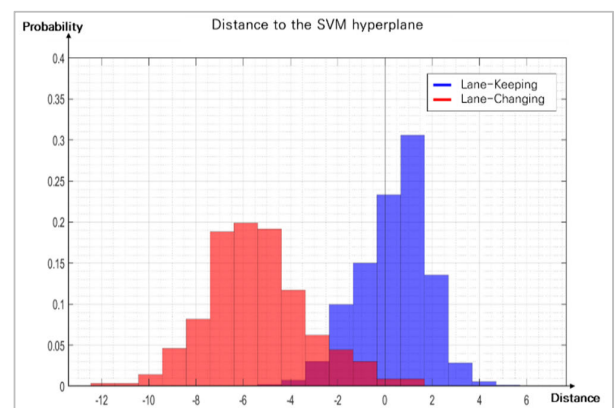

FIGURE 7. Estimation result of LC prediction model using SVM.

On the other hand, the distances of LK situations are estimated to range from -4 to 4, indicating that data on the negative side of the boundary are incorrectly classified as LC situations. Additionally, the average and standard deviation of the distance distribution in LK situations are estimated to

be 1 and 0.5, respectively. Shape of the distribution of LK data is narrow and asymmetrical compared to that of LC data.

Furthermore, there is an ambiguous area where both LC and LK data coexist within a range from −4 to 2. This area represents driver heterogeneity, where the LC decision-making process differs among drivers in the same driving conditions. As the LC relaxation and anticipation steps are not clearly defined and difficult to determine, this area may indicate a transitional phase between LC and LK situations. Moreover, there is an imbalance in the size of the data, with the number of LC samples significantly smaller than that of LK samples. Once a vehicle changes lanes during the observation, the duration of the LC process is shorter than that of LK situations.

C. PERFORMANCE EVALUATION OF THE MODEL

The performance of the model is measured for the positive and negative classes. In this study, the positive class represents the rare event of an LC maneuver. True Negative (TN) refers to the number of actual LK samples that are correctly predicted as LK. True Positive (TP) is the number of actual LC samples that are correctly predicted as LC. False Positive (FP) is the number of LK samples that are incorrectly predicted as LC. False Negative (FN) is the number of LC samples that are incorrectly predicted as LK. The measures can be calculated as follows:

$$\text{Simple Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FP} + \text{FN}}, \quad (29)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (30)$$

$$\text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}, \text{ and} \quad (31)$$

$$\text{F1-score} = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (32)$$

Simple accuracy, represented by (29), calculates the ratio of the number of true samples out of the total. Precision, represented by (30), is the ratio of the number of TP samples out of the predicted positives. It indicates the confidence level of positive sample detection. Recall, represented by (31), is the ratio of the number of TP samples out of the actual positives. It reflects the detection power of positive samples. The F1-score in (32) represents the harmonic mean of precision and recall.

To evaluate the proposed model, the vehicle trajectory dataset was divided into train, valid, and test sets. The measures, summarized in Table 4, show high performance in distinguishing between LC and LK moments. The train dataset consists of 6,191 situations, including 563 LC moments and 5,628 LK moments. The specific results for TP, FN, FP, and TN are estimated as 467, 96, 35 and 5,593, respectively. The measures for simple accuracy, precision, recall, and F1-score are calculated as 0.9788, 0.9303, 0.8295, and 0.8770, respectively. On the other hand, the test dataset contains 2,641 situations, including 176 LC moments and 2,465 LK moments. The specific results for TP, FN, FP, and

TN are estimated as 140, 36, 13 and 2,452, respectively. The measures for simple accuracy, precision, recall, and F1-score are calculated as 0.9814, 0.9150, 0.7955, and 0.8511, respectively. High performance of all datasets verifies and validates the proposed model. These results indicate that the proposed model with the optimal V2V features can identify how the V2V features exhibit in different values between the LC and LK maneuvers.

TABLE 4. LC estimation results.

Data-set	Number of Samples				Simple accuracy	Pre-cision	Recall	F1-score
	TP	FN	FP	TN				
Train	467	96	35	5,593	0.9788	0.9303	0.8295	0.8770
Valid	113	8	6	2,803	0.9952	0.9495	0.9338	0.9416
Test	140	36	13	2,452	0.9814	0.9150	0.7955	0.8511

The estimation results provide several implications regarding LC maneuvers. First, the high performance of the model indicates that most LC and LK moments can be distinguished using only seven V2V communication features. Since these features are easily obtainable and interpretable, the model is applicable for developing a naturalistic driving behavior model. Second, the LC behavior is visible and intuitive through the feature values, allowing surrounding vehicles to recognize the intention of the subject vehicle and cooperative accordingly. This is supported by the finding that the most significant features are the lateral acceleration and lateral distance of the subject vehicle. Third, the relationship between the selected features should be carefully examined in the LC decision-making process. The presence of an ambiguous area indicates driver heterogeneity, suggesting a transitional phase between LC and LK maneuvers. Fourth, despite the high model performance, it is important to note that the proposed model needs to be developed with more details in terms of safety. FN samples indicate misjudgments of LC maneuvers as LK maneuvers, implying that lag vehicles may fail to recognize the LC maneuvers of subject vehicles, potentially leading to traffic accidents.

V. STRATEGY FOR NATURALISTIC LC BEHAVIOR OF AUTONOMOUS VEHICLES

It would take more than 30 years to fully replace all vehicles with AVs. Therefore, it is crucial to develop the naturalistic driving behavior of AVs to resemble that of human drivers. This will help prevent confusion among human drivers and enable them to make proper driving judgments. Abnormal driving behaviors exhibited by AVs, such as abrupt LCs without properly assessing the surrounding traffic conditions or failure to yield based on misinterpretation of the traffic environment, can lead to increased rates of traffic accidents and congestion when interacting with human drivers. Therefore, this section proposes and evaluates a strategy for naturalistic LC behavior of AVs using the proposed model.

A. DISTINGUISHED V2V COMMUNICATION FEATURES

To investigate the characteristics of naturalistic LC behavior, the post-processing stage involves removing the inappropriately classified LC and LK data. Fig. 8 and Table 5 show statistics on the distance to the SVM hyperplane of the processed data, displayed in boxplots according to each feature. Note that these values are expressed using the normalized z-score since the distance is calculated using the vector. Interestingly, it is observed that, generally, during LK situations, the values of acceleration tend to be higher than those observed during LC situations. Additionally, the absolute values or range widths of the LK data tend to be smaller compared to the LC data across all selected features.
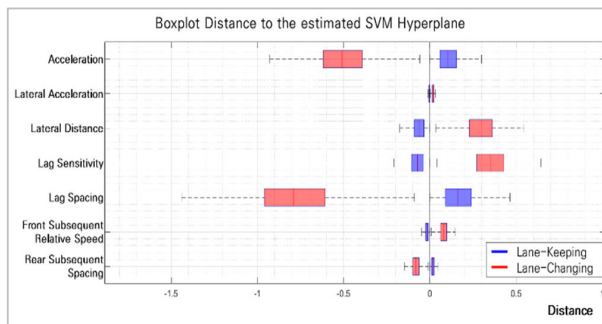

FIGURE 8. Distinguished V2V communication features.

TABLE 5. Distinguished V2V communication features.

V2V communication features	LK situation		LC situation	
	Average (Std.)	Range	Average (Std.)	Range
1. Longitudinal acceleration	0.114 (0.074)	[0, 0.612]	- 0.501 (0.185)	[-1.134, 0]
2. Lateral acceleration	- 0.004 (0.003)	[-0.022, 0]	0.018 (0.007)	[0, 0.040]
3. Lateral distance	- 0.067 (0.043)	[-0.358, 0]	0.293 (0.108)	[0, 0.663]
4. Lag Sensitivity	- 0.079 (0.051)	[-0.424, 0]	0.348 (0.128)	[0, 0.786]
5. Lag Spacing	0.177 (0.115)	[0, 0.949]	- 0.774 (0.286)	[-1.759, 0]
6. Front subsequent relative speed	- 0.018 (0.012)	[-0.097, 0]	0.079 (0.029)	[0, 0.179]
7. Rear subsequent spacing	0.018 (0.012)	[0, 0.098]	- 0.081 (0.030)	[-0.182, 0]

B. STRATEGY FOR NATURALISTIC LC MANEUVERS

Based on the estimated results, we propose a naturalistic AV driving strategy for implementing LC maneuvers. AVs follow to target the estimated average values of V2V features in each of the LK and LC situations. A total of 100 simulations with variations were conducted in this study. Variations in the values were randomly sampled within one standard deviation range of the average from the normal distribution for each

feature at every time step. Then, the optimal longitudinal and lateral accelerations at each time step were determined by averaging the obtained values during the simulation.

1) SIMULATION RESULTS OF TRAJECTORIES

Examples of the HV and AV trajectories are depicted in Fig. 9. The x- and y- axes represent longitudinal and lateral positions, respectively. The trajectory lines directing the east show the movements of vehicles. The blue and red lines indicate the LK and LC maneuvers of HV trajectories, respectively. The gray lines indicate simulated AV trajectories. The results represent that AVs have shown human-like driving behaviors during LC implementation. The simulation results are descibed below.

As shown in Fig. 9(a), the actual trajectory of the HV displays a DLC maneuver, transitioning from lane 3 to lane 2. During the LK situation before the LC, the vehicle shifts its lateral position from approximately 9 m to 7.5 m, and its longitudinal position from around 220 m to 305 m. Then, it enters into the LC situation with the anticipation step (Point A). The vehicle gradually approaches the lane marking and informs an LC intention to its surrounding vehicles. Then, it crosses over the marking at the longitudinal position 310 m (Point B), which indicates the LC execution step. After the execution, the vehicle is in the LC relaxation step and it aims to align its lateral position of around 6.5 m (Point C). As the LC situation is over, the vehicle returns back to the LK situation while maintaining its lateral position. For the case of simulation, AV trajectories show similarities with the HV trajectory in the longitudinal and lateral movements. For another example, as shown in Fig. 9(b), the simulation results for AVs demonstrate small variations in trajectories compared to other samples. This is due to smaller spacings with their surrounding vehicles, not allowing the subject vehicle to have many options for relaxed drivings. On the other hand, as shown in Fig. 9(c), the results for AVs demonstrate greater variations due to larger spacings with their surrounding vehicles. Nevertheless, all these AVs have shown enough spacings with their surrounding vehicles. In the case of Fig. 9(d), as the actual HV trajectory shows the most smooth and stable movements, the simulated AV trajectories also display their small variations during both LK and LC situations accordingly.

Notably, no collisions have occurred in any of the simulation results. In terms of the qualitative aspect, it can be concluded that the proposed model has developed the AV driving strategy to resemble HV driving behaviors for harmonizing with the traffic flow. This V2V-based AVs assess its given traffic condition carefully and performs human-like driving behavior. Then human drivers may not misjudge AVs' unfamiliar behaviors and thereby reduce both traffic accident rates and level of congestion.

2) PERFORMANCE EVALUATION

To evaluate performance of the proposed strategy, we employ measurements that evaluate accuracy of simulated trajectories over time frames. These measures calculate the discrepancy
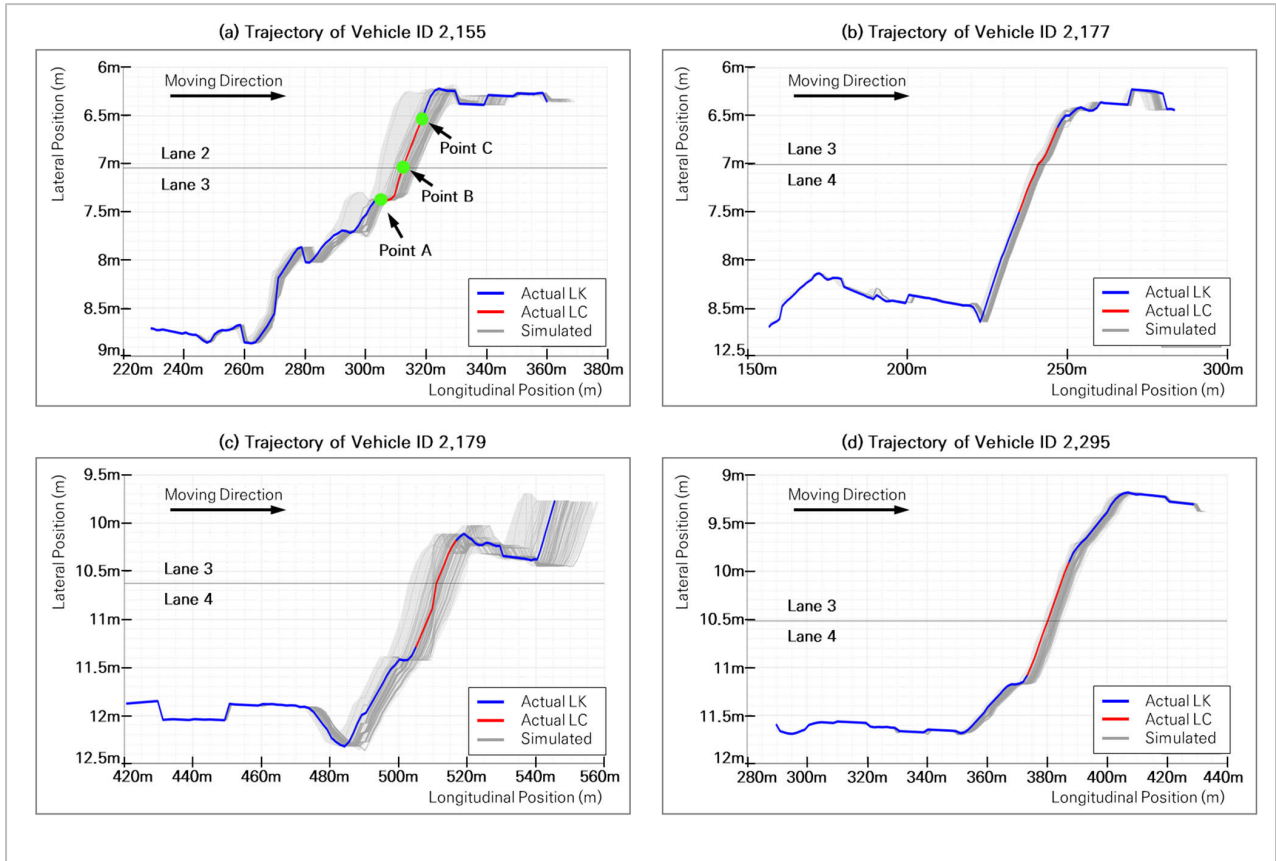
FIGURE 9. Comparison between the actual and simulated trajectories of LC vehicles.

in trajectories between the actual HV and simulated AV data.

$$\text{RMSE} = \begin{cases} \sqrt{\dfrac{1}{t_{\max}} \sum_{t=1}^{t_{\max}} \left(p_n(t) - \hat{p}_n(t)\right)^2} \\ \sqrt{\dfrac{1}{t_{\max}} \sum_{t=1}^{t_{\max}} \left(y_n(t) - \hat{y}_n(t)\right)^2} \end{cases}, \quad (33)$$

and

$$\text{FDE} = \begin{cases} \text{abs}\left(p_n(t_{\max}) - \hat{p}_n(t_{\max})\right) \\ \text{abs}\left(y_n(t_{\max}) - \hat{y}_n(t_{\max})\right) \end{cases} \quad (34)$$

where,

$p_n(t)$: longitudinal position of actual HV ID n at time t,

$\hat{p}_n(t)$: longitudinal position of simulated AV ID n at time t,

$y_n(t)$: local lateral position of actual HV ID n at time t,

$\hat{y}_n(t)$: local lateral position of simulated AV ID n at time t,

t : the frame number, and

$t_{\max}$: the maximum value of frame number.

RMSE (Root Mean Squared Error), represented by (33), calculates the mean value of the longitudinal and lateral displacement error. FDE (Final Displacement Error), represented by (34), is the value of the final displacement error between the actual and simulated trajectories. Since there

have been a total of 100 simulations conducted, the performance measures obtained from individual simulations have been averaged.

We have conducted a comparative analysis of the trajectory prediction results between the proposed model and other models. The other models are adopted from the study conducted by Ren et al. [36]. They proposed a state-of-the-art LSTM model that predicts trajectories of LC vehicles. They have demonstrated the model's performance by comparing with a classical LSTM model. They set the maximum duration for the prediction as 5 seconds. While our model has been simulated with the duration of about 15 seconds, we have selected same duration of LC situations to compare the results.

As summarized in Table 6, the RMSE of the longitudinal and lateral positions are calculated as 7.8826 and 0.4650 in the classical model, 4.5821 and 0.2987 in the model of Ren et al., and 4.1443 and 0.0314 in our proposed model. The overall RMSEs have been calculated as 7.8963, 4.5918, and 4.1444, respectively. The overall error has been reduced by 47.5%. For the FDE, the errors of the longitudinal and lateral positions are calculated as 10.6916 and 0.6602 in the classical model, 6.6774 and 0.2720 in the model of Ren et al., and 6.0688 and 0.0304 in our proposed model. The overall FDEs have been calculated as 10.7120, 6.6829, and 6.0689, respectively. The overall error has been reduced by 43.3%.

TABLE 6. Simulation results.

Model	RMSE (m)		FDE (m)	
	Longitudinal / Lateral	Overall	Longitudinal / Lateral	Overall
Classical LSTM	7.8826 / 0.4650	7.8963	10.6916 / 0.6602	10.7120
Ren *et al.* [36]	4.5821 / 0.2987	4.5918	6.6774 / 0.2720	6.6829
Proposed	4.1443 / 0.0314	4.1444	6.0688 / 0.0304	6.0689
Error Reduction	-	47.5%	-	43.3 %

These results demonstrate that our proposed model has reduced the prediction error. Notably, AV driving behaviors in the lateral direction have significantly improved. It can be interpreted that the simulated AVs harmonize with their surrounding vehicles while changing lanes. The proposed model is capable of effectively simulating naturalistic LC behaviors based on V2V communication. Moreover, as the model based on SVM has achieved interpretability, it can be said that AV driving behaviors can be adjusted anytime for better traffic operations and safety.

VI. CONCLUSION AND DISCUSSION

Given that the transition period of mixed traffic is expected to continue for more than 30 years, it is crucial to evolve AV technology to resemble human driving. This will help prevent confusion among drivers and enable them to make proper driving judgments. Abnormal driving behaviors exhibited by AVs, such as abrupt LCs without properly assessing the surrounding traffic conditions or failing to yield based on misinterpretation of the traffic environment, can lead to increased rates of traffic accidents and congestion when interacting with human drivers.

The existing models have predicted AV trajectories either using a predicted permutation or a recognized permutation. As the prediction results largely depend on the train dataset, some critical bias and errors can occur if unexperienced dataset are input during simulations. To address this issue, this study proposes the intuitive and interpretable model. First, the standardized approach is established to select the optimal V2V features using the mRMR algorithm with Pearson correlation method. Then, the SVM is applied to intuitively and simply separate boundaries of the V2V features between LC and LK situations. Subsequently, the AV driving strategy is proposed as AVs to target the estimated average values of V2V features in each LK and LC situation. As a result, our proposed model achieved both interpretability and high model performance for LC implementation.

Performance measures for distinguishing the characteristics between the LC and LK situations have shown high accuracy when applied to the NGSIM dataset. Simple accuracy, precision, recall, and F1-score have been measured as 0.9814, 0.9150, 0.7955, and 0.8511, respectively.

Subsequently, the simulation results of the AV driving strategy have significantly reduced the overall prediction error by 43.3%, compared to the existing model. Notably, AV driving behaviors in the lateral direction have significantly improved. The results suggest that the proposed model effectively simulates naturalistic LC behaviors for AVs to harmonized with their surrounding vehicles based on the seven optimal V2V features.

There are several issues in this study that merit future investigation to improve AV technology. First of all, other V2V candidate features can be investigated further. More effective features can improve misjudgments between the LC and LK maneuvers at the moment. For example, it would reduce a probability that an AV executes an LC even though the gap between the lead and lag vehicles on the target lane is not acceptable. Additionally, other scenarios where surrounding vehicles are less than six should also be considered. Some other time periods where there are not much traffics should be applied as well. Furthermore, other driving situations, such as MLC situations, other geometric configurations, or interrupted flows, should also be further studied. Lastly, integrating other driving behaviors such as car-following maneuvers or cooperative driving could enhance the driving behavior model in a more natural manner.

Although the SVM model estimates the boundary conditions between the LC and LK situations, there exists other machine learning tools that can be applied for further models such as spatio-temporal algorithms [37]. Future research should focus on developing technological tools that provide clear interpretations of complex driving patterns in greater detail. Moreover, further studies with larger datasets are expected to refine the proposed model for improved efficiency.

REFERENCES

[1] K. Jiang, C. Sun, H. Zhou, X. Li, M. Dong, and V. C. M. Leung, "Intelligence-empowered mobile edge computing: Framework, issues, implementation, and outlook," *IEEE Netw.*, vol. 35, no. 5, pp. 74–82, Sep. 2021.

[2] Q. Luo, S. Hu, C. Li, G. Li, and W. Shi, "Resource scheduling in edge computing: A survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2131–2165, 4th Quart., 2021.

[3] H. Zhou, K. Jiang, S. He, G. Min, and J. Wu, "Distributed deep multi-agent reinforcement learning for cooperative edge caching in Internet-of-Vehicles," *IEEE Trans. Wireless Commun.*, early access, May 11, 2023, doi: 10.1109/TWC.2023.3272348.

[4] S. M. Mohtavipour, T. Z. Ehsan, H. J. Abeshoori, and M. Mollajafari, "Smooth longitudinal driving strategy with adjustable nonlinear reference model for autonomous vehicles," *Int. J. Dyn. Control*, vol. 11, no. 5, pp. 2320–2334, Oct. 2023, doi: 10.1007/s40435-023-01142-4.

[5] L. Li, H. Liang, J. Wang, J. Yang, and Y. Li, "Online routing for autonomous vehicle cruise systems with fuel constraints," *J. Intell. Robotic Syst.*, vol. 104, no. 4, pp. 1–16, Apr. 2022.

[6] S. Zhao, X. Wang, H. Chen, and Y. Wang, "Cooperative path following control of fixed-wing unmanned aerial vehicles with collision avoidance," *J. Intell. Robotic Syst.*, vol. 100, nos. 3–4, pp. 1569–1581, Dec. 2020.

[7] A.-T. Nguyen, C. Sentouh, and J.-C. Popieul, "Fuzzy steering control for autonomous vehicles under actuator saturation: Design and experiments," *J. Franklin Inst.*, vol. 355, no. 18, pp. 9374–9395, Dec. 2018.

[8] K. Osman, J. Ghommam, and M. Saad, "Guidance based lane-changing control in high-speed vehicle for the overtaking maneuver," *J. Intell. Robotic Syst.*, vol. 98, nos. 3–4, pp. 643–665, Jun. 2020.

[9] M. Cantas, A. Chand, H. Zhang, G. Surnilla, and L. Guvenc, "Data association between perception and V2V communication sensors," *SAE Tech. Paper 2023-01-0856*, 2023, doi: 10.4271/2023-01-0856.

[10] S. Lefevre, C. Laugier, and J. Ibanez-Guzman, "Risk assessment at road intersections: Comparing intention and expectation," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2012, pp. 165–171.

[11] C. Sun, J. Leng, and B. Lu, "Interactive left-turning of autonomous vehicles at uncontrolled intersections," *IEEE Trans. Autom. Sci. Eng.*, early access, Dec. 20, 2022, doi: 10.1109/TASE.2022.3227964.

[12] W. Schwarting, A. Pierson, J. Alonso-Mora, S. Karaman, and D. Rus, "Social behavior for autonomous vehicles," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 50, pp. 24972–24978, 2019.

[13] F. S. de Sio, G. Mecacci, S. Calvert, D. Heikoop, M. Hagenzieker, and B. van Arem, "Realising meaningful human control over automated driving systems: A multidisciplinary approach," *Minds Mach.*, 2022, doi: 10.1007/s11023-022-09608-8.

[14] A. Kusuma, R. Liu, C. Choudhury, and F. Montgomery, "Analysis of the driving behaviour at weaving section using multiple traffic surveillance data," *Transp. Res. Proc.*, vol. 3, pp. 51–59, Jan. 2014, doi: 10.1016/j.trpro.2014.10.090.

[15] B. Son, T. Kim, H. J. Kim, and S. Lee, "Probabilistic model of traffic breakdown with random propagation of distribution for ITS application," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, vol. 3215, 2004, pp. 45–51.

[16] M. Park, K. Jang, J. Lee, and H. Yeo, "Logistic regression model for discretionary lane changing under congested traffic," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 4, pp. 333–344, Apr. 2015, doi: 10.1080/23249935.2014.994686.

[17] T. Toledo, C. F. Choudhury, and M. E. Ben-Akiva, "Lane-changing model with explicit target lane choice," *Transp. Res. Rec.*, vol. 1934, no. 1, pp. 157–165, 2005, doi: 10.3141/1934-17.

[18] J. Lee, M. Park, and H. Yeo, "A probability model for discretionary lane changes in highways," *KSCE J. Civil Eng.*, vol. 20, no. 7, pp. 2938–2946, Nov. 2016, doi: 10.1007/s12205-016-0382-z.

[19] K. I. Ahmed, "Modeling drivers' acceleration and lane changing behavior," Ph.D. thesis, Dept. Civil Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 1999.

[20] X. Wan, P. J. Jin, F. Yang, J. Zhang, and B. Ran, "Modeling vehicle interactions during merge in congested weaving section of freeway ramp," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2421, no. 1, pp. 82–92, Jan. 2014, doi: 10.3141/2421-10.

[21] H. Kita, "A merging-giveway interaction model of cars in a merging section: A game theoretic analysis," *Transp. Res. A, Policy Pract.*, vol. 33, nos. 3–4, pp. 305–312, 1999, doi: 10.1016/S0965-8564(98)00039-1.

[22] X. Long, L. Zhang, S. Liu, and J. Wang, "Research on decision-making behavior of discretionary lane-changing based on cumulative prospect theory," *J. Adv. Transp.*, vol. 2020, pp. 1–16, Jan. 2020, doi: 10.1155/2020/1291342.

[23] C. F. Choudhury and M. M. Islam, "Modelling acceleration decisions in traffic streams with weak lane discipline: A latent leader approach," *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 214–226, Jun. 2016.

[24] C. F. Choudhury, T. F. Email, M. E. Ben-Akiva, and A. Rao, "Modeling cooperative lane changing and forced merging behavior," in *Proc. 86th Annu. Meeting Transp. Res. Board*, 2006, pp. 1–16.

[25] A. Kusuma, R. Liu, C. Choudhury, and F. Montgomery, "Lane-changing characteristics at weaving section," in *Proc. Transp. Res. Board 94th Annu. Meeting*, Jan. 2015, pp. 49–55.

[26] S. Lee, D. Ngoduy, and M. Keyvan-Ekbatani, "Integrated deep learning and stochastic car-following model for traffic dynamics on multi-lane freeways," *Transp. Res. C, Emerg. Technol.*, vol. 106, pp. 360–377, Sep. 2019.

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

[28] L. Fridman, L. Ding, B. Jenik, and B. Reimer, "Arguing machines: Human supervision of black box AI systems that make life-critical decisions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–9.

[29] H. M. Mandalia and M. D. D. Salvucci, "Using support vector machines for lane-change detection," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 2005, vol. 49, no. 22, pp. 1965–1969.

[30] W. Wang, J. Xi, A. Chong, and L. Li, "Driving style classification using a semisupervised support vector machine," *IEEE Trans. Hum.-Mach. Syst.*, vol. 47, no. 5, pp. 650–660, Oct. 2017, doi: 10.1109/THMS.2017.2736948.

[31] S. Ramyar, A. Homaifar, A. Karimoddini, and E. Tunstel, "Identification of anomalies in lane change behavior using one-class SVM," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 004405–004410, doi: 10.1109/SMC.2016.7844924.

[32] P. Kumar, M. Perrollaz, S. Lefèvre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 797–802, doi: 10.1109/IVS.2013.6629564.

[33] Y. Liu, X. Wang, L. Li, S. Cheng, and Z. Chen, "A novel lane change decision-making model of autonomous vehicle based on support vector machine," *IEEE Access*, vol. 7, pp. 26543–26550, 2019, doi: 10.1109/ACCESS.2019.2900416.

[34] R. Jiang, M.-B. Hu, H. M. Zhang, Z.-Y. Gao, B. Jia, and Q.-S. Wu, "On some experimental features of car-following behavior and how to model them," *Transp. Res. B, Methodol.*, vol. 80, pp. 338–354, Oct. 2015, doi: 10.1016/j.trb.2015.08.003.

[35] S. Sra, S. J. Wright, and S. Nowozin, *Optimization for Machine Learning*. Cambridge, MA, USA: MIT Press, 2011.

[36] Y.-Y. Ren, L. Zhao, X.-L. Zheng, and X.-S. Li, "A method for predicting diverse lane-changing trajectories of surrounding vehicles based on early detection of lane change," *IEEE Access*, vol. 10, pp. 17451–17472, 2022.

[37] E. H. Lee, S. Y. Kho, D. K. Kim, and S. H. Cho, "Travel time prediction using gated recurrent unit and spatio-temporal algorithm," *Proc. Inst. Civil Eng.-Municipal Engineer*, vol. 174, no. 2, pp. 88–96, 2021, doi: 10.1680/jmuen.20.00004.

EUNTAK LEE received the B.S. degree in civil and environmental engineering and the Ph.D. degree in urban planning and engineering from Yonsei University, Seoul, South Korea, in 2018 and 2023, respectively. His current research interests include traffic engineering, traffic flow theory, intelligent transportation systems, machine learning, autonomous vehicles, and data analysis.

YOUNGJUN HAN received the B.S. and M.S. degrees in urban planning and engineering from Yonsei University, Seoul, South Korea, in 2002 and 2006, respectively, and the Ph.D. degree in civil and environmental engineering from the University of Wisconsin–Madison, in 2017. He is currently an Associate Research Fellow with The Seoul Institute.

JU-YEON LEE received the B.S. degree in civil and environmental engineering and the M.S. and Ph.D. degrees in urban planning and engineering from Yonsei University, Seoul, South Korea, in 2003, 2005, and 2010, respectively. She is currently a Researcher with the Korea Transport Institute.

BONGSOO SON received the B.S. degree in civil and environmental engineering from Yonsei University, Seoul, South Korea, in 1982, the M.S. degree from McMaster University, Hamilton, Canada, in 1990, and the Ph.D. degree from the University of Toronto, Toronto, Canada, in 1996. He is currently a Professor with Yonsei University.

• • •