Transportation Research Part C: Emerging Technologies, Volume 115, June 2020, Article number 102615
DOI:10.1016/j.trc.2020.102615

An Ensemble Deep Learning Approach for Driver Lane Change Intention Inference

Yang Xing [a], Chen Lv [a], Huaji Wang [b], Dongpu Cao [c], Efstathios Velenis [d]

[a] School of Mechanical and Aerospace Engineering, Nanyang Technological University, 639798, Singapore.
[b] AVL Powertrain UK Ltd, Coventry, CV4 7EZ, UK.
[c] Department of Mechanical and Mechatronics Engineering with the University of Waterloo, 200 University Avenue West Waterloo, ON, N2L3G1, Canada.
[d] Advanced Engineering Centre, Cranfield University, Bedford, MK43 0AL, UK.

***Abstract*—With the rapid development of intelligent vehicles, drivers are increasingly likely to share their control authorities with the intelligent control unit. For building an efficient Advanced Driver Assistance Systems (ADAS) and shared-control systems, the vehicle needs to understand the drivers' intent and their activities to generate assistant and collaborative control strategies. In this study, a driver intention inference system that focuses on the highway lane change maneuvers is proposed. First, a high-level driver intention mechanism and framework are introduced. Then, a vision-based intention inference system is proposed, which captures the multi-modal signals based on multiple low-cost cameras and the VBOX vehicle data acquisition system. A novel ensemble bi-directional recurrent neural network (RNN) model with Long Short-Term Memory (LSTM) units is proposed to deal with the time-series driving sequence and the temporal behavioral patterns. Naturalistic highway driving data that consists of lane-keeping, left and right lane change maneuvers are collected and used for model construction and evaluation. Furthermore, the driver's pre-maneuver activities are statistically analyzed. It is found that for situation-aware, drivers usually check the mirrors for more than six seconds before they initiate the lane change maneuver, and the time interval between steering the handwheel and crossing the lane is about 2 s on average. Finally, hypothesis testing is conducted to show the significant improvement of the proposed algorithm over existing ones. With five-fold cross-validation, the EBiLSTM model achieves an average accuracy of 96.1% for the intention that is inferred 0.5 s before the maneuver starts.**

***Index Terms*—Driver intention, ADAS, RNN, LSTM, intelligent vehicle.**

I. INTRODUCTION

A. *Motivation*

Millions of injuries and deaths are caused by traffic accidents each year worldwide. Most of the traffic accidents are caused by human drivers' issues, such as cognitive overload, judgment mistakes, and operation errors [1]-[3]. Drivers influence the other road entities and the surrounding traffic context by controlling their vehicles according to their driving willingness and traffic laws. The Traffic-Driver-Vehicle (TDV) loop can be defined as a dynamic social system, with each part being capable of influencing and interacting with the other dynamically [4]. Since drivers are positioned in the center of the TDV loop, a proper understanding of driver's intention and related behaviors can help effectively reduce the number of traffic accidents and increase driving safety [5] [6]. In the past two decades, a series of ADAS products has been implemented on commercialized vehicles. Currently, most of the successful ADAS products, such as the lane departure warning system (LDW) [7] [8], adaptive cruise control (ACC) [9], and side warning assist system (SWA), [10] are mainly designed to provide additional context information to assist the drivers. Although these products can be viewed as active safety systems, they still interact with the human drivers passively since most of these systems are unable to monitor and understand driver behaviors in real-time. The dynamic interaction and mutual understanding between the driver and the intelligent units are expected as one of the primary objectives for the development of the next-generation ADAS [11] [12].

In this study, a driver reasoning system is proposed towards real-time lane change intention inference (LCII). The reasons for understanding and inferring driver intentions are multifold [13]. First, the primary motivation is to improve the driving safety of human drivers. One of the significant tasks of driver intention inference (DII) is to anticipate the driver's behavior before the maneuver happens or at an early stage after the maneuvers are initiated. LCII enables ADAS or intelligent vehicles to focus on the potential traffic region as early as possible so that dangerous situations can be detected in an early stage. With precise inference of driver intention, ADAS can generate proper notification and assistance to the driver [14], or alter the possible dangers based on the drivers' future trajectories and interested regions [15]. As shown in [16], turn signals are only used in 66% of lane changes, and less than 50% of the turn indicators are activated in the initial phase of the lane change maneuvers. Therefore, with the prediction of the intention, the LCII-enabled vehicle will interact with human drivers more efficiently, as the system can understand the drivers'

*Corresponding author at Department of Mechanical and Mechatronics Engineering with the University of Waterloo, 200 University Avenue West Waterloo, ON, N2L3G1, Canada.
E-mail address: dongpu.cao@uwaterloo.ca (DP. Cao)

Published by Elsevier. This is the Author Accepted Manuscript issued with: Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.trc.2020.102615. Please refer to any applicable publisher terms of use.

plans and will not distract the drivers with false alarms even if no turn signal is activated. Also, driver intention prediction has been analyzed from the perspectives of law, regulations, and policies. For example, in [17], the intention to commit violations enables the prediction of traffic laws violation and is essential to the driver's assistance. Hence, it is shown that intention inference and behavior anticipating are efficient ways to improve the functional design of ADAS.

Second, as most of the ADAS and high-level automation systems tend to share the control authorities with the drivers, the LCII system can decrease the conflicts between the human driver and the automation functionalities. The mutual understanding between the human driver and the automation is essential for the construction of the intelligent shared control strategies. For example, Pentland and Liu described the human as a device that has a large number of internal mental states and control behaviors [18]. If a machine can anticipate human behavior, it can better serve humans' needs. In [19], the driver's intention on path planning is recognized based on the lateral offset and lateral velocity, and it is integrated into the shared obstacle-avoidance model predictive controller. However, current driver behavior enabled shared control systems mainly rely on the driver control commands, such as the steering angle and vehicle velocity. These systems cannot provide an early intention prediction and are hard to fully exploit the cooperation potential of the integrated shared control strategies [19][20]. The intelligent vehicles are believed to have great potential in comprehensive sensing and perception of the surrounding context. While understanding driver intention will further boost the automation system to generate smart sensing and precise control assistance to the driver, which will significantly enhance the public acceptance of intelligent vehicles.

Last, understanding the driver's intention mechanism is expected to contribute to a more naturalistic decision-making system for autonomous vehicles. It can be used to design the human-like decision-making and behavior generation algorithms [21]. People may argue that future autonomous cars will not maintain the driver in the control loop. However, this does not mean that human-like decision-making and control are not necessary for the vehicles. Learning how human drivers generate intentions and make decisions are long-term tasks for autonomous vehicles, as the rule-based methods cannot adequately meet the requirements of the various and uncertain situations in the real world. The data-driven intention and decision-making models are expected to be more efficient for the development of vehicle automation in the future [22].

Although the technology of driver intention inference is believed to be promising, there are still not many commercialized products that have been deployed in the automotive industry. The development of the driver intention inference systems is still challenging and needs further analysis and improvement in terms of both accuracy and robustness. In this study, a driver LCII system is developed based on an ensemble RNN structure to improve the accuracy of driver intention recognition. As a popular model enhancement approach, the ensemble learning method has achieved a vast amount of successes in the classification, regression, and data mining [23][24]. The reasons for applying ensemble learning into driver intention inference is multifold. First, most of the existing machine learning-based LCII algorithms are sensitive to the training data, and they can generate different sets of weights and performance if the training data changes. While the ensemble learning method can increase the model robustness by introducing various light networks [25]. Second, inspired by the Bootstrap aggregating (bagging) technique, several light LSTM networks can be developed based on the different sub-sampling datasets to improve the intention prediction accuracy and model generalization performance [26]. Last, it is a common challenge that massive naturalistic driver behaviors and intention data are hard to be collected and labeled. By sampling, augmenting the raw dataset, and assembling several light LSTM networks, a more precise intention inference model can be achieved with limited raw datasets. Hence, in this study, an ensemble learning-based RNN model with LSTM cells (EBiRNN) is developed to model the temporal process of driver intention inference. Based on the proposed EBiRNN model, the driver behaviors for preparing a lane change maneuver are statistically analyzed based on naturalistic highway driving data. The real-time performance of the LCII system is also illustrated, indicating the efficiency of the LCII system. Conclusions made in this work are expected to benefit the design of efficient LCII systems in the future.

B. Literature Review

Driver intention has been widely studied in the past two decades. The very beginning research can be found in [27], where only the vehicle dynamic data, such as the steering angle, steering velocity, and vehicle velocity, were used. An average detection rate of 88.3% was achieved after the lane change maneuver has been initiated for 0.5 seconds. After that, Oliver and Pentland proposed another intention prediction model based on the Coupled Hidden Markov model (CHMM) [28]. The CHMM model gave a high prediction accuracy for a start and stop intention one second prior, while it only achieved 29.4% and 6.3% detection rate for the left and right lane change maneuver, respectively. After these early studies, it was found that a precise intention inference system should rely on a holistic approach, which needs to fuse the multi-modal data within the TDV loop [29] [30]. To recognize the lane change intention, signals within the TDV loop that come from different sensors are needed. For the traffic context, the most widely used features are lane markings, surrounding vehicle positions, digital maps, and Global Positioning System (GPS), etc. [31]-[33]. Important driver behavioral features consist of head motion, eye gaze, body gestures, and even the Electroencephalography (EEG), etc. [34]-[36]. The vehicle status signals usually can be collected from the in-vehicle Controller Area Network (CAN) bus/Ethernet, which generally contains vehicle speed, acceleration, steering wheel angle and velocity, turn signal, and pedal, etc. [13] [37] [38].

In [39], Moreo *et at.* introduced an interactive-multiple-models (IMM) approach to predict the lane change maneuver on the highway. The GPS/IMU sensors were used to collect naturalistic driving data. The GPS can provide the location and time information in harsh weather conditions, and it is more robust than the camera and Light Detection and Ranging (Lidar) devices in such situations. Salvucci *et al.* introduced a four-step LCII system based on the framework of the human cognitive process, which

contains data collection, model simulation, action tracking, and thought inference [38] [40]. Schmidt *et al.* proposed a lane change intention recognition method based on an explicit mathematical steering wheel angle model [41]. In [42], driver intention was detected by observing the easily accessible vehicle and traffic signals, which focused on the analysis of the impact of environmental indicators on the prediction of lane change intention. Henning, *et al*. proposed a lane change recognition system which focused on the analysis of the impact of environmental indicators on the prediction of lane change intention [43]. The vehicle dynamic signals are the most straightforward signals that can be easily collected. However, the vehicle data is hard to reflect the lane change intention before the driver taking any control actions. Therefore, the vehicle status information can only be used to recognize the on-going intention instead of predicting the intention before any actions are taken.

On the other hand, driver behavioral features, such as the head pose and eye gaze signals, can give earlier clues about the driver's intention. Many studies have evaluated their influence on the intention prediction problem [44]-[47]. Zhou *et al.* evaluated how the driver behaviors were affected by the cognitive distraction during the lane change preparation process through the analysis of driver eye movement [48]. They concluded that a secondary task could affect the accuracy of intention inference. Li *et al*. proposed an integrated intention inference algorithm based on HMM and Bayesian Filtering (BF) techniques [49]. A preliminary output from the HMM was further filtered using the BF method to make the final decision. In [50], the authors proposed a lane change detection method based on the object-oriented Bayesian networks (OOBN). The system was designed according to the modularity and reusability of the Bayesian network, which makes the system easier to be extended considering different requirements.

Moreover, it was found that driver behavior information gives a more significant contribution to the intention inference, comparing with the traffic context and vehicle dynamics in [30] [51]. The head poses signals show significant advantages in the early recognition of the lane change intent. With various combinations of the input data, eye gaze signals were found not as informative as the head pose and did not significantly improve the precision of the intention inference system. In [52], a discriminative relevance vector machine (RVM) classifier was used to predict driver lane change intent with a sliding time window. The multimodal signals from ACC, SWA, LDW, and head motion were fused. It was found that the LDW system was more useful to predict the intent between 0 and 1.5 seconds before the lane change occurs, while the head motion is more informative between two and three seconds before the lane change happens. Recently, Jain *et al.* [15] [53] used the Autoregressive Input-Output HMM (AIOHMM) and RNN-LSTM algorithms to predict the intention and future maneuvers. The algorithms took the inside and outside video streams, vehicle dynamics, GPS, and street maps as input signals to anticipate the lane change, turn, and normal driving maneuvers. The authors concluded that the RNN model leads to the best detection precision of 88.2% and can anticipate the lane change maneuvers 3.5 seconds earlier. However, the strategical plan of the route and the utilization of the digital map in that study could reduce the system reliability in the unseen streets.

C. Contribution

Although some successful cases on the lane change intention inference have been reported in the past, few works have been devoted to analyzing the lane change intention on both the cognitive level and the activity level. In this study, we analyze the cognitive process of intention generation to build a unified intention inference framework. An overall analysis of driver intention and its framework are proposed. The frame is designed according to the driver cognitive process and considering behaviors in different time-scales. Second, a light ensemble RNN structure is developed for the intention inference task to improve the accuracy and robustness of the RNN model. The LCII system is designed with a low-cost vision system, which is scalable and easy to be implemented. By integrating the inside and outside driving context into the EBiLSTM model, the temporal LCII model achieves a state-of-the-art recognition accuracy for the lane change intention, compared with other baselines models. Finally, the naturalistic temporal properties of the lane change process on highways are analyzed to better understand the lane change process of the human driver, and the real-time intention inference results are illustrated and discussed.

D. Paper Organization

The remainder of this paper is organized as follows. Section II introduces a general framework for driver intention inference and the problem formulation for LCII. In section III, the experiment setup and the feature extraction for the inside and outside driving context are proposed. The ensemble deep learning scheme based on the RNN model is described in Section IV. In Section V, the statistical analysis for the temporal characteristic of lane change maneuver and the intention inference results are proposed. The discussion and future works are included in Section VI. Finally, this study is concluded in Section VII.

II. The Framework of Driver Intention Inference

In this section, A driver LCII framework is introduced. The framework describes the development procedure of the lane change intention and the relationship and connection of different driver behaviors. This framework can also be extended to describe intention inference tasks in many other application domains.

A. Intention Inference Framework

Michon pointed out that the cognitive structure of human behavior is a four-level hierarchical architecture, which contains the road user, transportation consumer, social agent, and psycho-biological organism [54]. The road user level is directly connected to the driver's intent and behaviors, and it can be further divided into three levels: strategy, tactical, and operational (also known as control) level according to the time constant property. The strategical level defines the high-level plans for each trip, such as the route, destination, and risk assessment, etc. The time constant of this level is much longer than that in the rest two levels, which can

last for minutes or hours. The tactical tasks are the driving maneuvers that a driver can take during normal driving, such as lane change, lane keeping, deceleration, and acceleration. The time constant of these maneuvers is usually at around several seconds. Finally, the control level describes the human control actions on the vehicles, and it stands for the willingness to remain safe and comfortable during driving. The time constant at this level is generally in milliseconds [37]. A general lane change intention framework is shown in Fig. 1, which indicates the relationship between the tactical level and the control level [13]. As discussed in [13], human intention can be generated according to some stimuli. For example, the lane change decision is usually made due to an uncomfortable driving context. Once the intention occurs, it can be reflected by different driver dynamics, such as the variation of EEG signals, the head, and eye motion, etc. These behavioral signals play a crucial role in the estimation of the intention. After the drivers have decided to execute the intention, they will control the vehicle through the steering wheel and the acceleration/brake pedal. Finally, the vehicle responds to these control actions with the variation of vehicle dynamics.

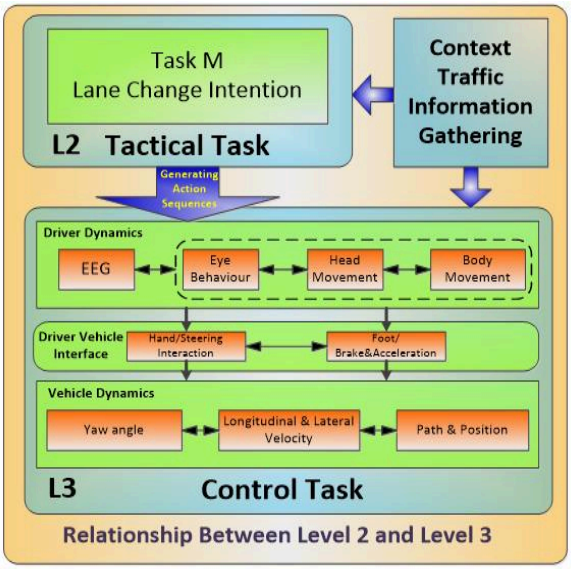

Fig. 1. The LCII framework and its relationship with the control actions in the third level.

It can be found that before the lane change maneuver is initiated, the driver must execute a series of checking behaviors according to the traffic context and his/her situation awareness. Therefore, the traffic context and driver behaviors are the two critical clues for predicting the lane change intention. The drivers will only change the vehicle control strategy after they have decided to execute their lane change intention. Hence, the vehicle dynamic signals have a limited contribution to the prediction of the intention. However, these signals are still useful for the recognition of the driving intention at an early stage.

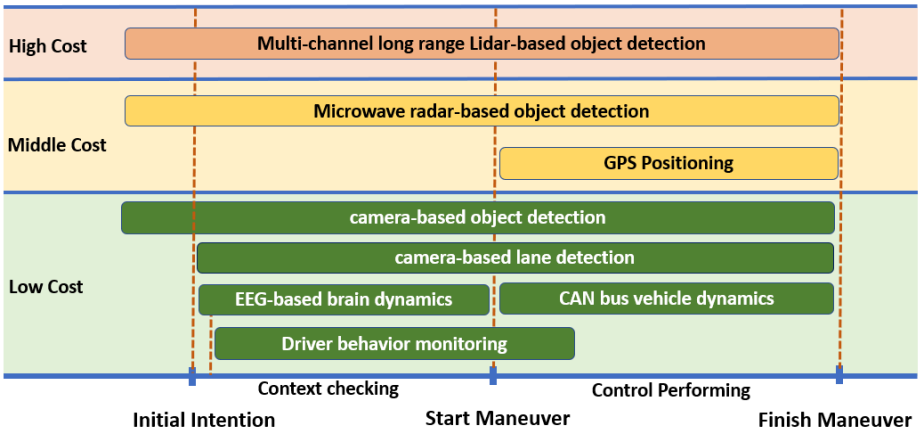

Fig. 2. Exemplar sensor systems for driver intention inference.

Fig.2 illustrates some typical sensory systems that can be used for driver intention inference considering the critical conditions of lane change maneuver. Nowadays, the Lidar/Radar devices are widely accepted in the precise positioning and velocity detection of surrounding vehicles. The Lidar/Radar detection results also can be used for efficient driving intention inference of the surrounding vehicles. Moreover, based on Fig. 1, driver intention can be stimulated by outer traffics, which can be further classified into destination-oriented or surrounding stimuli-oriented ones. Therefore, the Lidar/Radar based object detection and motion predication systems will benefit the intention analysis of the ego-driver during the whole process of one tactical maneuver (i.e.,

intention generation, maneuver execution, and maneuver finish). Based on the cognitive process of driver intention illustrated in Fig. 1, it can be found that driver behavior and physical dynamics, such as EEG signals, is believed to be more straightforward than the behavior monitoring, making it closer to the real intention generation moment. However, the real-time noise of the EEG signals is one of the significant drawbacks that worsen the acceptance of the device in driver monitoring. The camera-based systems have a lower cost than the Lidar and Radar-based ones in road perception, which makes them more accessible for the real-world low-cost application. Moreover, the camera devices are convenient to be implemented inside the vehicle cabin for driver monitoring. The driver behavior monitoring systems (e.g., head tracking, gaze tracking, breath, heart rate, etc.) can be used as the primary systems to recognize the driving intention before the maneuvers happen. Once the maneuver is initialized, the driver usually maintains their hands on the steering wheel and slightly change the vehicle control input. Hence, at this moment, the global GPS positioning, vehicle dynamics, and local lane position detection become essential cues. One of the objectives of this study is to analyze the driver's intention based on the features that can bring early cues. Therefore, the driver's head pose, eye gaze, vehicle velocity, and lane marking styles are jointly used in this study for precise intention inference.

B. Problem Formulation

Based on the intention inference framework shown in Fig. 1, this study expects to solve the following problems. First, to construct an efficient LCII system based on the fusion of multi-modal sensors. The primary objective of this study is to infer the lane change intention before the driver initiates the maneuver. Thus, the steering wheel and pedal signals will not be considered as the primary input for the LCII system. Secondly, since the driver's intention is not an instant detection task, the inference model should be able to process the temporal information and mining the temporal dependency between the features. Therefore, the LSTM based bi-directional RNN model is adopted in this study. To further improve the intention prediction accuracy, an ensemble RNN structure is developed. Next, it is essential to understand the naturalistic driver behaviors and analyze the statistic roles within the lane change maneuvers. The statistic results for the mirror checking behavior before the lane change maneuver and the lane change duration are analyzed. In this study, only the intended lane change maneuvers are included, while the unintended and aborted lane changes will not be considered.

III. METHODOLOGIES

This section describes the experimental setup and data processing for lane change intention inference. In Section III. A, the experiment design, data collection, and feature processing framework are introduced. Section III B and Section III. C describes the feature extraction for the traffic context and driver behavior information, respectively.

A. Experiment Setup and Data Processing

In this study, naturalistic highway driving data are collected for experiment analysis and model evaluation. The vehicle testbed is a commercialized sport utility vehicle (SUV), which is equipped with a multi-modal sensor system. The sensory platform includes three low-cost CMOS cameras and a Velocity BOX (VBOX) datalogger. The VBOX datalogger is a high-performance GPS receiver that can capture a series of vehicle dynamic information such as vehicle velocity, acceleration, heading, and positions, etc. The three cameras are all mounted inside the vehicle cabin. One is mounted in front of the driver to capture driver head motion and eye gaze information. The outside traffic context monitoring camera is mounted in the middle on top of the front window. The third one is mounted on the sunroof to monitor driver hand motion and the usage of the turn signal. It is only used for labeling driver maneuver, and no data stream are used to model construction. A diagram of the system architecture is shown in Fig. 3.

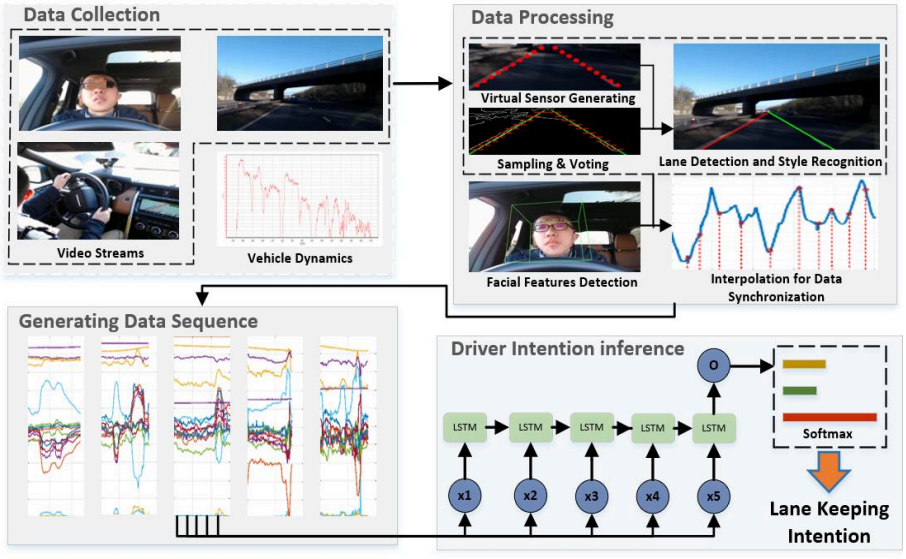

Fig. 3. Experimental platform and system architecture of the ane change intention inference system.

All three cameras are synchronized with the time step of 25 fps. The video streams are recorded at the resolution of 640×480. Two VBOX antennas are fixed on top of the vehicle roof ceiling to measure the vehicle dynamic states, including velocity, acceleration, and heading. The sampling frequency for the VBOX data logger is 20 HZ. All the signals are collected using a laptop with an Intel Core i7 2.5 GHz CPU. Three adult drivers with varying ages and experiences participated in the data collection process. They were asked to drive as usual without being told the real objectives of the experiment. Each driver drove the vehicle on the highway for about one hour, and a total of 150 miles of naturalistic data were collected.

Based on the on-board sensory platform shown in Fig. 3, the proposed system for data collection and feature processing can be further illustrated in Fig. 4. The comprehensive feature extraction module is responsible for the feature extraction of driver behaviors and road context. Specifically, for the driver's behavioral data, the head poses and eye gaze dynamics will be extracted. Regarding the road context, the LCII system needs to be aware of the proper road and lane mark style (dashed or solid lane marking) to make the prediction. Hence, the lane styles and positions will be detected to guide the intention inference system of making reasonable predictions. These features will be concatenated together and fed into the LCII model part, which processes the time-series feature set based on the ensemble RNN models. Finally, the predicted driver intention, along with the traffic context, can be integrated into the risk assessment and intelligent assistance modules for intelligent vehicles. In the following sections, the processing for the driver-oriented and road-oriented features will be discussed in detail.

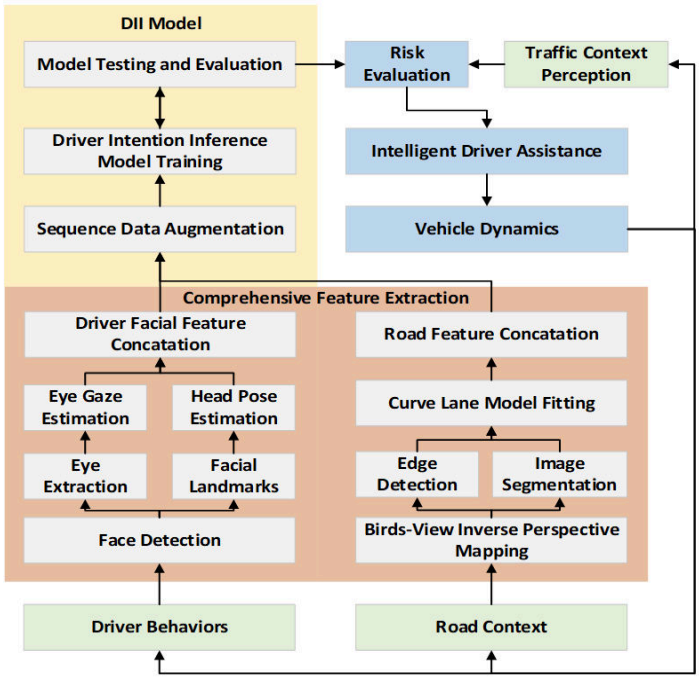

Fig. 4. Data collection and feature processing procedure for the proposed ensemble-based LCII system.

B. Features of the Traffic Context and Vehicle Dynamics

The road-facing camera captures the front traffic context, and it mainly focuses on lane detection. The lane detection system is constructed with an integrated edge detector, Hough Transform, and polynomial lane function fitting modules. This method has been successfully used for lane detection in many existing studies [7]. The camera is calibrated to make an inverse perspective mapping and detect the distance between the virtual vehicle central line as well as the adjacent lanes. The detected lane positions are tracked using a Kalman filter. Then, a lane style detector based on the lane sampling and voting (LSV) scheme is developed [7]. Three different lane styles, namely, the solid, double solid, and dashed lane, are recognized. First, an odd number of sampling points are generated in the detected lane position. Then, the lane style for the left and right lanes can be recognized by extending the sampling points to short sampling segments. The scanning is proposed on the edge image given by the Sobel edge detector, and the lane style is determined by counting the rising edge along each sampling line. The voting results of the sampling lines determine the final lane style. The system is further developed into a lane departure warning system. An entire lane change process is visualized in Fig. 5. In the first image of Fig. 5, the yellow lines represent the dashed line styles, which means the driver is allowed to make a lane change maneuver to the corresponding side. When the vehicle is approaching the lane, the algorithm can detect a lane approaching maneuver and generate lane change warning signal based on the estimation of the distance between the virtual vehicle central line and the adjacent lanes.

Two vehicular signals, namely the vehicle speed (V), and heading angle (H) are collected using the VOBX. The total feature vectors for the outside traffic context and vehicular dynamics can be formed as a four-dimensional vector.

$$O_t = [L_r \; L_l \; V \; H] \tag{1}$$

where L_r and L_l are the lane style for right and left lanes, and $L \in [-1, 0, 1]$ represents the three different lane styles.

Fig. 5. An illustration of a lane change process which consists of lane-keeping, lane changing, lane crossing, and recovering stages.

C. *Driver Behavioral Features*

The inside features for driver head motion and eye gaze are detected with the Openface system [55]. The driver head position and facial landmarks are identified using the Conditional Local Neural Fields (CLNF) approach [56]. The CLNF estimates the 3D head pose angles by projecting the 3D representation of the facial marks to the image plane using orthographic camera projection. The pupil locations for both eyes are detected with the deformable shape registration approach. The gaze directions are estimated according to the pupil locations and the 3D eyeball center [57]. The inside feature vector at each time constant can be formed as follows.

$$I_t = [G_r\ G_l\ G_a\ H_t\ H_r] \tag{2}$$

where G_a, is the 2D gaze angle in x and y coordinate, G_r, G_l, H_t, and H_r are the 3D gaze direction for each eye, head pose translation vector, and head pose direction vector, respectively. This gives a 14-dimensional vector of driver behavior features. A comparison of the head yaw angles for different maneuvers is illustrated in Fig. 6. From the time-series yaw angles, it can be found that the head rotation can indicate different dynamics for different context checking.

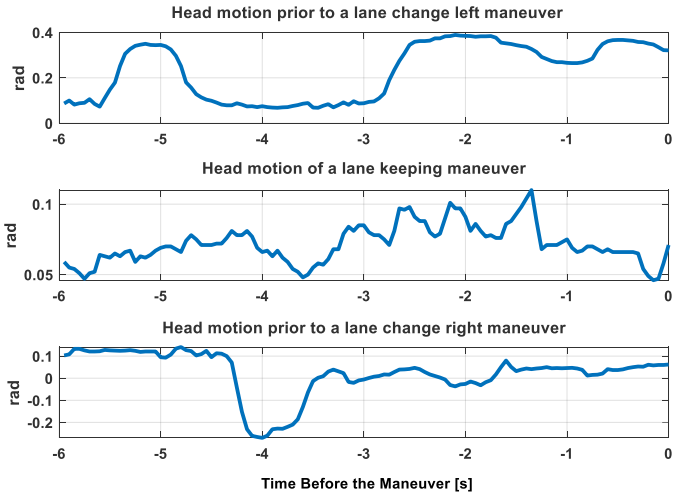

Fig. 6. An example of the dynamics of the head yaw angle for the three different maneuvers. The times-series data shows the variation of the yaw angle occurs six seconds before the maneuver happens. Time 0 represents the initial time of the lane change maneuver.

IV. ALGORITHM

In this section, the recurrent neural network, LSTM cell, and ensemble RNN model are introduced. The RNN model is used to learn the temporal dependency between the input data, and the LSTM efficiently improves the performance of the RNN model by capturing the long-term context dependencies.

A. *Recurrent Neural Network and Bi-directional Structure*

Since driver intention inference is not an instant detection task, historical behavioral data needs to be taken into consideration. The recurrent neural network is applied in this study to process the sequential inputs. RNN allows exhibiting the dynamic temporal behavior of a sequence by forming a directed connection between previous states and the current states [58]-[60]. Based on the vanilla RNN structure, an advanced version of the RNN, i.e., the bi-directional RNN (BRNN), is adopted [61]. To overcome the shortcomings of vanilla RNN, which is only able to use the backward information to make a prediction, the bi-directional RNN enables the process of the whole temporal sequence with both forward and backward information. The bi-directional RNN increases the amount of input information that is available on the network. In terms of intention inference, given a fixed period of

sequence, the driver's intention within this sequence is expected to be better identified by considering the bi-directional connection between the current and the previous contexts. A single layer BRNN structure is shown in the right part of Fig. 7.

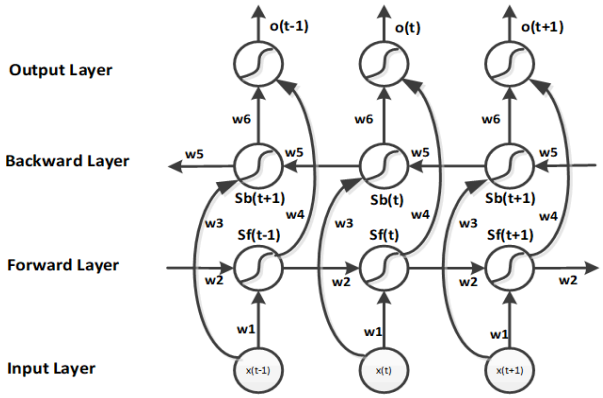

Fig. 7. A simplified single layer Bi-directional Recurrent Neural Network structure.

The BRNN can be expressed as:

$$s_{ft} = f(\boldsymbol{W}_1 x_t + \boldsymbol{W}_2 s_{ft-1} + b_x) \tag{3}$$

$$s_{bt} = f(\boldsymbol{W}_3 x_t + \boldsymbol{W}_5 s_{bt-1} + b_x) \tag{4}$$

$$o_t = f(\boldsymbol{W}_4 s_{ft} + \boldsymbol{W}_6 s_{bt} + b_o) \tag{5}$$

where f is the activation function of the states, which is normally selected as $tanh$ and $sigmoid$ function, X, S_f, S_b and O represent the input, forward hidden states, backward hidden states, and the output of the RNN, respectively. $\boldsymbol{W}$ are the weight matrix for the input, hidden states, and the output. b_x and b_o are the input and output bias vector, respectively.

Although the number of training parameters is reduced with the parameter sharing scheme, RNN or BRNN still suffers another severe problem, namely, the gradient vanishing or exploding [62], which cause the vanilla RNN structure has limited ability to capture the long-term dependency. Hence, Hochreiter and Schmidhuber developed the LSTM cell to overcome the gradient vanishing and exploding drawbacks of RNN [63]. The LSTM-RNN solves the long-term dependency problem by introducing three extra gates, known as the input gate, forget gate, and output gate. The central idea behind LSTM is that the gates in the LSTM cell cooperate to control how much information should be remained and forgotten. A good illustration of the LSTM gated cell framework can be found in [64]. The LSTM-RNN still follows the chain-like structure. The difference is LSTM-RNN replace the hidden unit (usually a $sigmoid$ or $tanh$ activation function) with an LSTM cell. The LSTM cell has the following mathematic representation. Firstly, the forget gate controls what information to throw away. Then, the input gate chooses information to be updated and stored. The output gate controls the candidate layer output.

$$f_t = \sigma\left(\boldsymbol{U}_f x_t + \boldsymbol{W}_f s_{t-1} + b_f\right) \tag{6}$$

$$i_t = \sigma(\boldsymbol{U}_i x_t + \boldsymbol{W}_i s_{t-1} + b_i) \tag{7}$$

$$o_t = \sigma(\boldsymbol{U}_o x_t + \boldsymbol{W}_o s_{t-1} + b_o) \tag{8}$$

where σ in the above equations represents the $sigmoid$ function. $*$ is the element-wise production, x_t and s_{t-1} are the current input and previous LSTM cell output. f, i, o are the forget gate, input gate, and output gate, respectively. U, W, b are the corresponding weight matrix. A value $\widetilde{c_t}$ in the candidate cell state can be represented as:

$$\widetilde{c_t} = tanh(\boldsymbol{U}_c x_t + \boldsymbol{W}_c s_{t-1} + b_c) \tag{9}$$

The c_t is the key parameter of LSTM unit, which determines the cell output and information processing, and it is the combination of previous c_{t-1} and the current candidate states.

$$C_t = f_t * c_{t-1} + i_t * \widetilde{c_t} \tag{10}$$

Finally, the output of the layer is the products of the cell state C_t and the candidate output from the output gate.

$$s_t = o_t * tanh(C_t) \tag{11}$$

The driver intention inference (DII) model takes the driver behavioral feature and lane style features as the input, and outputs the predicted maneuver based on cross-entropy loss function. The general time-series DII model can be represented as:

$$y_t = DII(X)$$

where y_t is the output of DII model. X is the sequence input of the ensemble-LSTM based DII model at each moment, and it can be further extended as $X = \{x_{t-p}, x_{t-p+1}, \cdots, x_t\}$. p is the previous data threshold that determines how much past data is used. Based on the statistical analysis of the lane change dynamics in the next section, the prediction horizon is selected as six seconds. Hence, the previous data threshold p can be calculated as $p = predict\ horizon \times sampling\ rate$, the sampling rate is 25, and p is 150.

The input x_t for the LSTM-RNN at each time step is the concatenation version of the inside and outside feature vector $x_t = [I_t, O_t]$ as given in the last section, which is an 18-dimensional feature vector. Hence, each training sample for the LSTM-RNN is an 18-D temporal sequence, and the total dataset can be formed as$\{(x_{t-p}, x_{t-p+1}, \cdots, x_t)_j, y_j\}_{j=1}^{N}$, where N is the total number of the training sample. Finally, the input sequence to the model is formed into the format of 18×150. y_j is the scalar intention label of each sequence. In this study, the target values are manually labeled according to the video streams, which has three candidate values, as $\{lane\ change\ left,\ lane\ change\ right,\ lane\ keeping\}$. The overall ensemble-DII model architecture is shown in Fig. 8. The raw data sequence is augmented into a series of sub-sequence to train different light LSTM networks. Then, the results voting module takes the outputs from different sub-models and vote for the majority label as the final inferred maneuver. The ensemble process is given in the next part in detail.

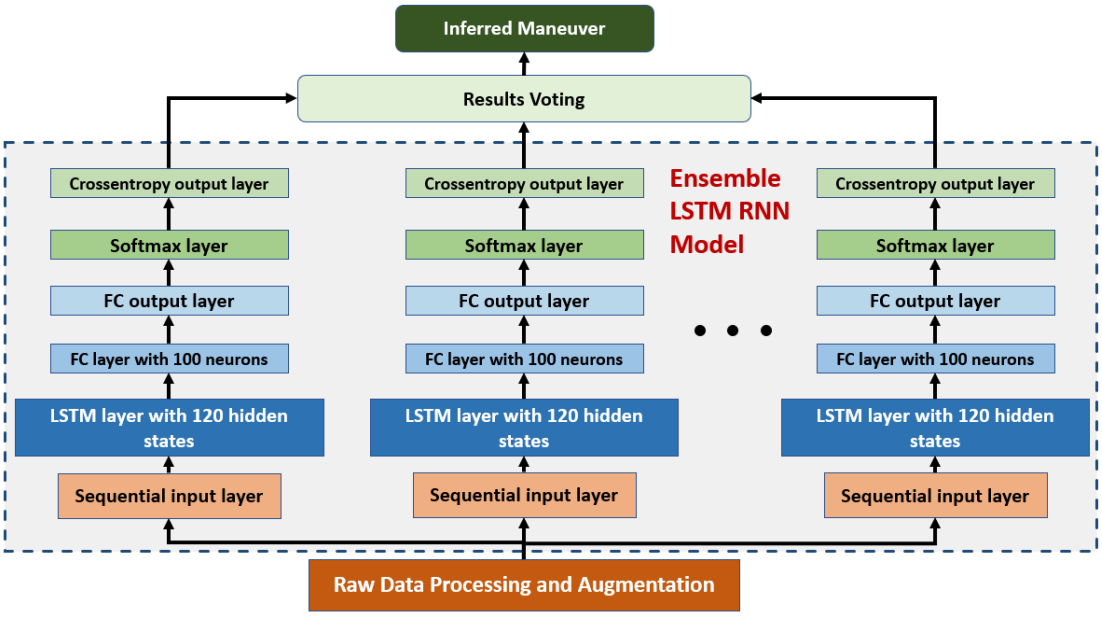

Fig. 8. The architecture of the proposed ensemble LSTM-RNN model.

B. Ensemble LSTM-RNN Structure

A light ensemble method structure is applied in this study to improve the intention inference accuracy of the LSTM-RNN model. As discussed in [65], a single model can be weak to make predictions, while assembling these weak models can build a strong predictor. The ensemble method is adopted for the intention inference tasks so that the integration of multiple LSTM-RNN models can guarantee the inference accuracy. The diversity of each single RNN is one of the most important factors that need to be considered to construct the ensemble LSTM-RNN model. Hence, in this study, the diversity for each RNN is controlled based on an augmented random data sampling method. The data augmentation scheme is designed with two objectives. First, to increase the data volume to avoid overfitting during the LSTM-RNN model training process. Second, based on the data augmentation, the different sub-training datasets can be generated for multiple model training, and the sub-RNN models can maintain its diversity and compensate with each other. The data augmentation can be represented as:

$$X_{new} = g(X, \theta, n),\ \theta \in [0.65, 0.9] \tag{12}$$

where the X is the original training dataset, $g(\cdot)$ is the sampling function which randomly samples sub-dataset according to the random number θ. θ is a randomly generated parameter in each iteration. Its value is between $[0.65, 0.9]$, which follows the uniform distribution. n is the total number of the augmentation steps, which determines how many sub-datasets should be generated. In this study, n is set as three so that the augmentation process will be applied for three times. The final training dataset is the combination of the original dataset and the augmented one. Hence, the new training dataset $X_{new}$is four times larger than the original one. For each augmentation iteration (iteration < 4), a random sampling rate θ is generated. Then, one augmentation dataset at each iteration can be obtained by randomly sampling the raw data sequence according to the sampling rate θ.

The random sample control threshold [0.65,09] is manually and empirically selected. This threshold is used to control the diversity of the dataset. For each sequence, a sample rate within this range is first generated to sample the original data. The primary motivation of using the sample control threshold is to create a different dataset for each sub-LSTM model to increase the model diversity while maintaining the temporal patterns as many as possible. The mechanism behind the ensemble method is similar to that of the ensemble tree model given by Breiman [65]. In [65], the random forest can generate out-of-bag (OOB) samples based on the Bootstrap aggregating (Bagging) method, which creates new samples for the trees by sampling from the dataset uniformly and with replacement. Hence, the sampled dataset can own about 63.2% unique examples from the original

dataset, while the rest one-third can be used as OOB data to evaluate the model. In this study, the random sample control threshold is used to increase the data diversity and model performance. The results indicate that the ensembled LSTM network can contribute to a more precise estimation than the BiLSTM network. The sampling threshold selected may not be the optimal one. Although the value selection for the threshold is not included in this study, it is worthwhile exploring in the future.

The data augmentation and model ensemble process are shown in Fig. 9. The raw dataset is first split into training and testing datasets. The data augmentation is only proposed to the training set, and testing data is only used for model testing. Each of the single LSTM can generate its prediction of the labels for the driver on-going intention, which can be one of the three studied maneuvers. Then, the final prediction for the driver intention task should be determined based on the majority conclusion from the multiple RNN models. One primary concern for the ensemble RNN framework is its real-time computational efficiency. Unlike the decision trees, which have a light volume and can make predictions very fast, the LSTM-RNN is hard to be ensembled into a large-scale forest structure. Therefore, the total number of models is much smaller than the classical random forest structure. Based on the real-time evaluation of the ensemble RNN structure, it can be found that a light ensemble scheme that combines three different RNN models is accurate enough for the lane change intention inference. It takes about 20ms for each RNN model to make a prediction, and it costs 60 ms in total for the three RNN models. This is acceptable for the intention inference since the timescale for the lane change intention is about a few seconds. We also tested the system with nine RNN models, which can provide a more accurate result than the three models case. However, it takes about 200 ms to make each prediction and does not show significant improvement. Therefore, in this study, the three different RNN models are used to process the intention inference task.

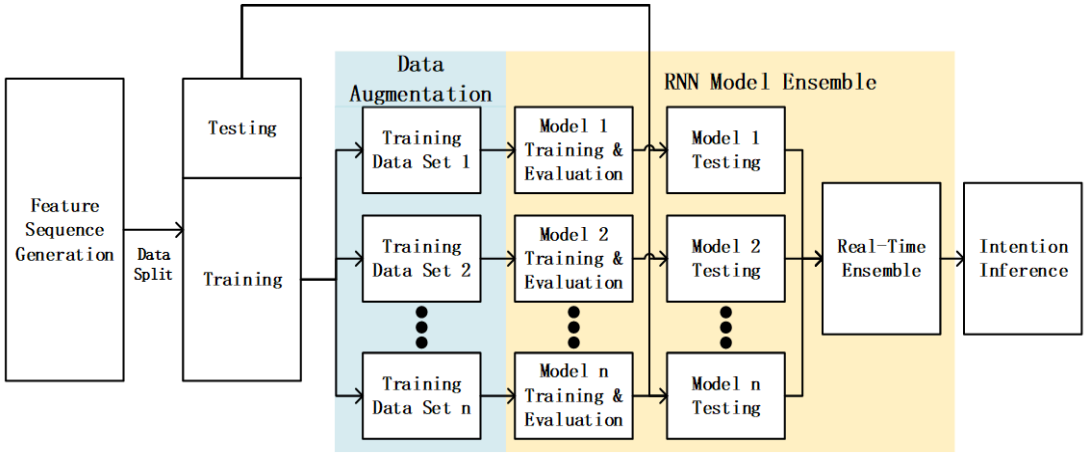

Fig. 9. The block diagram of the ensemble LSTM-RNN model construction process.

C. Model construction

There are 135 lane change maneuvers detected in the experiment video, with 65 left lane change cases and 70 right ones. Then, 66 regular lane-keeping driving sequences are randomly picked from the raw data. The sequential data are labeled based on the final maneuver of the video sequences. The naturalistic driving dataset is randomly split into the training and testing data. 80% of the data is used for model training, and the rest 20% is used for model testing. The random splitting ensures that the data from the three different maneuvers can be equally split into the training and testing groups so that the proportion of each maneuver is similar to each other in both the training and testing datasets. Each data sequence is an 18×150 tensor that represents an 18-dimensional input feature with 150-time steps (six seconds). The LCII model is trained with the Adam optimizer [66] with an initial learning rate of 0.001 and a gradient decay factor of 0.9. The maximum epoch is 150, and the mini-batch size is 32. The input weights are initialized using the Glorot initializer [67], and the recurrent cell weights are initialized using the orthogonal initializer [68]. The LSTM-RNN model is trained with the MATLAB Deep Learning Toolbox. The LSTM layer has 120 hidden units, followed by two FC layers with 100 neurons and three neurons for the output layer, respectively.

The model learning process considering training loss and training accuracy is shown in Fig. 10. The upper two graphs show the model learning performance of the BiRNN model with data augmentation. While the bottom two represent the training process of the BiRNN without using data augmentation. As the maximum epoch is selected 150 and the mini-batch size is 32, the overall iteration of the data augmentation enabled BiRNN is 3000, which is four times longer than that for the original data based BiRNN model. Based on the model learning results shown in Fig. 10, it is found that the learning process of the BiRNN model can converge within about 60 epochs with both augmented and original datasets using the Adam optimizer.

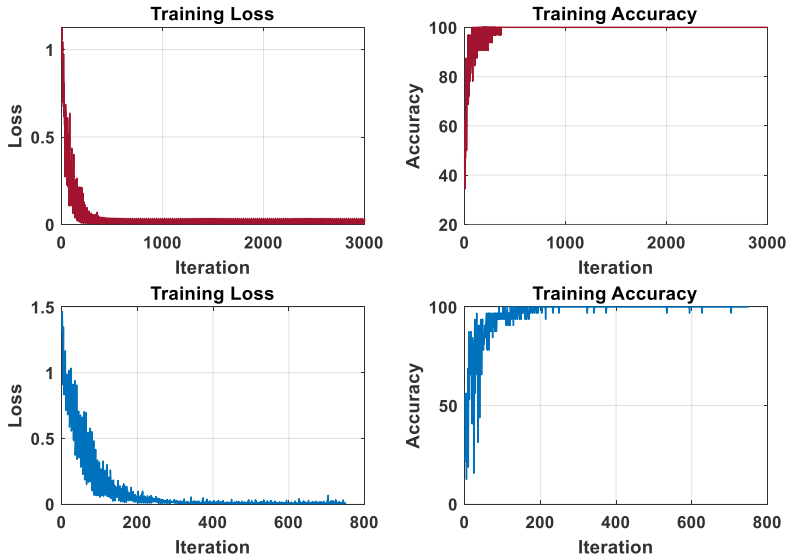

Fig. 10. Illustration of the learning process of the Bi-LSTM-RNN model. The upper two graphs show the training loss and training accuracy of the BiRNN model with the augmented dataset (shown in red line), while the bottom two show the training loss and training accuracy of the BiRNN model without using data augmentation.

V. Experiment Results

In this section, the statistical analysis of the lane change maneuver is conducted based on the critical driving conditions of the lane change process. Then, the LCII results of the EBiRNN model with different prediction horizons are compared over various algorithms.

A. Driver Lane Change Maneuver Analysis

As shown in the left subplot of Fig. 11, for each lane change maneuver, four critical moments during the lane change process, namely the intention occurs point (denoted as T1), the maneuver start point (T2), the lane crossing point (T3), and the maneuver finishing point (T4), need to be identified. At T1, the driver generates a lane change intention according to the traffic context stimuli. Most of the time, this specific initial moment of the intent is undetectable. Hence, T1 is roughly replaced by the time of the first mirror-checking behavior in the following analysis. At T2, the driver uses the turn signals to indicate their lane change intention and then turns the steering wheel. Finally, T3 and T4 are the moments that the vehicle crosses the lane and finishes the lane change maneuver, respectively. In this study, the statistical analysis of the gap between T1 and T2, T2, and T3 for all the recorded lane change maneuvers are investigated. The time interval between T1 and T2 (denoted as T1-T2) measures the time cost for the lane change preparation, and the interval between T2 and T3 (indicated as T2-T3) measures how long it takes to cross the lane after the driver starts the maneuver.

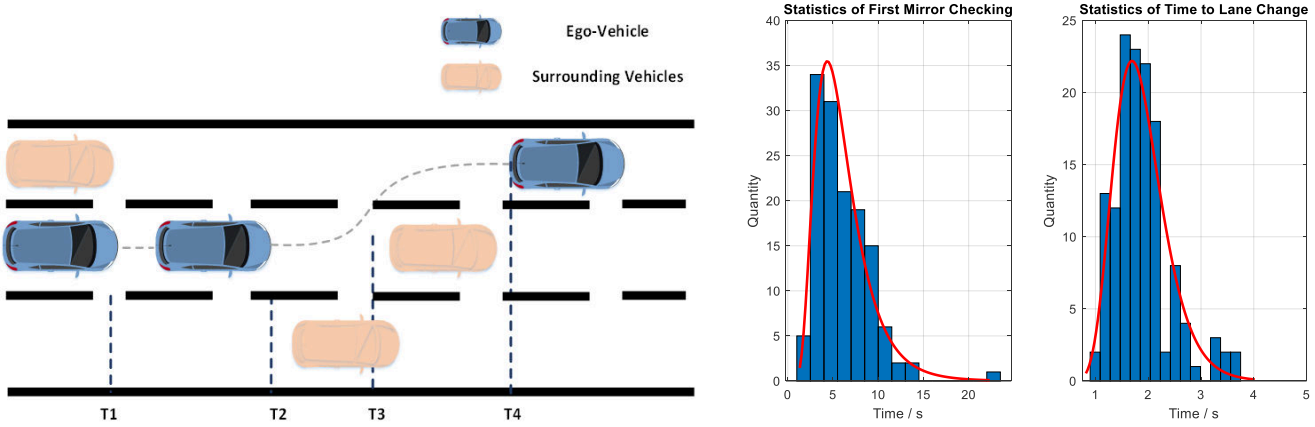

Fig. 11. Illustration of the lane change maneuver progress and the statistical analysis for driver behaviors before the lane change maneuver.

TABLE I
DESCRIPTIVE STATISTICS OF THE LANE CHANGE PREPARATION AND EXECUTION

Statistics	T1-T2	T2-T3
Mean	6.085 s	1.881 s
SD	3.033 s	0.530 s
Variance	9.200 s^2	0.281 s^2
Median	5.437 s	1.805 s
Mode	3.060 s	1.890 s
Maximum	22.423 s	3.680 s
Minimum	1.843 s	0.930 s

*SD is short for standard deviation

In the right subplot of Fig.11, the time distributions of T1-T2 and T2-T3 are fitted by the lognormal distribution function. According to the statistical results shown in Table 1, the drivers tend to perform their first mirror checking behavior about six seconds earlier before they start the lane change maneuver. The average time cost between initiating the lane change and crossing the lane is about 1.88 seconds. Therefore, to predict the lane change intention before the driver turns on the turn signal or turn the steering wheel, a six-seconds window is large enough to cover the featured behaviors. After the driver starts the lane change maneuver, the LCII system still has nearly two seconds to recognize an intended lane change. From the dataset, it is found that mirror checking behaviors can occur during the lane-keeping maneuver. However, the duration of these behaviors is much shorter than that during the lane change preparation process. Meanwhile, drivers tend to perform single side mirror checking multiple times during the lane change preparation stage instead of performing both sides checking behaviors as they do during the normal lane-keeping process. The difference between the mirror checking behaviors will be further analyzed in future work.

B. Lane Change Intention Inference

In this study, we examine the performance of the intention inference model from two aspects: the detection accuracy and the prediction horizon. The prediction horizon measures how early the LCII can make a precise prediction. Since all the drivers are asked to drive as usual, some of the lane change maneuvers may not be indicated by the turn signal. In such a case, the starting point of the maneuver is marked as the first moment of operating the steering wheel. The BiRNN-5 and EBiRNN-5 models are evaluated with 5-fold cross-validation. Besides, the BiRNN-20 and EBiRNN-20 are trained and tested for 20 times with different datasets to avoid bias and used for a reasonable T-Test. The EBiRNN-5 and EBiRNN-20 represent the overall results that are obtained based on the five-fold cross-validation and 20 times testing with different testing data, respectively. The structures of these models are the same.

In Table II, Table III and Table IV, the performance of the proposed EBiRNN models concerning different prediction horizon are compared with some existing machine learning methods, including the support vector machine (SVM), hidden Markov model (HMM), feedforward neural network (FFNN, with 50 hidden neurons), Naïve Bayesian and Random Forest (RF, with 50 decision trees). The inferences are made 0.5 s, 2 s, and 3.5 s before the maneuver starts, respectively, and all the baseline models are evaluated with a 5-fold cross-validation method. It is shown that the proposed EBiRNN model can achieve the most accurate result among all the methods with different prediction horizons. Among the baseline methods, The FFNN model shows significant advantages on the intention inference task over other conventional methods, and it even achieves better results than the vanilla LSTM approach in some cases. Therefore, the EBiRNN and the FFNN will be further evaluated and compared in the next section. It is shown that the statistic results with 20 times testing (EBiRNN-20) are slightly lower than that of the five-fold cross-validation. We think the results of EBiRNN-20 are closer to the real model performance. Therefore, the hypothesis testing will be proposed based on the EBiRNN-20 results.

For quantitative measurement, the precision, recall, F1 scores, and the general average precision are used to measure the model performance on each maneuver (right lane change, left lane change, and lane-keeping). Specifically, four classification results are defined and statistically calculated for each maneuver, which is the true positive (T_p, the model correct detects this maneuver), true negative (T_N, the mode correct detects the other maneuvers), false positive (F_p, the model detects the other maneuvers as target one), and false negative (F_N, the mode detects the target maneuver as other maneuvers).

Accordingly, the Precision (P_r) is calculated as:

$$\mathrm{Pr} = \frac{T_P}{T_P + F_P} \tag{13}$$

The Recall (R_e) is calculated as:

$$\mathrm{Re} = \frac{T_p}{T_p + F_n} \tag{14}$$

The F1-score considers both the P_r and R_e, and it is the harmonic mean of these two values.

$$F1 = 2 \times \frac{Pr \times Re}{Pr + Re} \tag{15}$$

Lastly, the general average precision is calculated as:

$$G_{Ave} = \frac{Total\ number\ of\ correct\ prediction}{Total\ number\ of\ samples} \tag{16}$$

TABLE II
RESULTS OF METHOD COMPARISON WITH THE PREDICTION MADE AT 0.5S BEFORE THE MANEUVER STARTS.

Algorithms	Left Lane Change			Right Lane Change			Lane Keeping			General Ave (%)
	Pr (%)	*Re* (%)	*F1 Score*	*Pr* (%)	*Re* (%)	*F1 Score*	*Pr* (%)	*Re* (%)	*F1 Score*	
SVM	**96.8±7.1**	85.2±4.7	90.4±3.3	80.0±9.3	**93.8±6.6**	85.8±3.3	95.8±4.2	**97.6±2.4**	**96.5±1.1**	91.5±1.8
FFNN	91.3±9.6	97.0±4.2	93.8±5.9	**98.1±2.5**	90.9±4.9	**94.4±3.2**	91.0±4.1	93.1±8.2	91.9±5.5	93.4±4.6
Naïve Bayes	91.6±2.6	93.0±2.5	92.3±1.4	93.0±3.3	85.2±3.7	88.9±2.3	89.5±4.2	95.9±2.8	92.5±3.1	91.3±1.9
HMM	83.6±2.9	96.2±3.9	89.4±0.9	73.4±2.8	94.5±3.8	82.9±2.2	97.3±5.9	53.3±6.2	68.6±3.9	81.6±1.3
RF	95.0±3.6	92.1±2.6	93.5±2.1	92.9±5.4	92.0 ±5.7	92.3±2.9	91.9±3.5	95.7±4.4	93.6±2.1	93.1±2.1
LSTM-RNN	91.9±2.7	94.4±3.6	93.5±1.4	95.2±3.4	92.5±3.5	93.7±1.7	90.4±4.6	90.1±5.7	90.1±2.8	92.5±1.5
Bi-RNN-5	96.7±1.9	94.4±4.6	95.5±3.1	93.5±2.2	95.0±3.5	94.2±2.4	90.6±7.6	91.9±2.8	91.1±4.4	93.8±2.7
EBiRNN-5	93.7±2.6	**98.1±2.6**	**95.9±2.6**	95.3±8.3	93.1±5.1	94.1±5.7	**98.0±4.6**	95.0±6.1	96.3±3.1	**96.1±3.4**
Bi-RNN-20	95.3±4.8	92.1±5.3	93.5±3.2	91.5±5.9	91.9±7.2	91.4±4.4	92.2±6.6	96.2±4.6	93.9±3.6	93.1±2.8
EBiRNN-20	95.6±5.7	95.1±6.4	95.1±4.5	95.6±5.2	92.5±6.3	93.7±3.2	94.9±5.4	97.5±4.3	96.0±3.3	95.1±2.2

TABLE III
RESULTS COMPARISON WITH THE PREDICTION MADE AT 2S BEFORE THE MANEUVER.

Algorithms	Left Lane Change			Right Lane Change			Lane Keeping			General Ave (%)
	Pr (%)	*Re* (%)	*F1 Score*	*Pr* (%)	*Re* (%)	*F1 Score*	*Pr* (%)	*Re* (%)	*F1 Score*	
SVM	90.3±5.6	**96.0±5.5**	92.9±3.6	94.6±4.2	89.1±2.0	91.7±2.0	89.4±3.3	92.0±5.9	90.7±3.6	91.8±2.3
FFNN	**97.1±2.6**	88.5±7.3	92.4±3.5	92.3±6.2	93.9±9.2	92.8±5.4	88.2±9.5	93.9±7.7	90.6±5.8	92.1±4.5
Naïve Bayes	91.7±1.1	79.2±1.0	84.3±5.7	81.2±1.3	**95.9±1.2**	87.0±6.0	95.9±4.6	90.4±6.9	92.8±1.5	88.2±3.7
HMM	72.8±8.6	85.7±9.5	78.6±8.1	71.8±9.5	91.8±6.7	80.4±7.3	91.4±13.5	46.7±9.2	65.4±10.2	75.0±7.6
RF	92.6±5.3	91.7±1.7	92.1±2.1	87.6±5.9	95.8±4.4	91.3±2.0	95.4±4.5	85.9±15.0	90.0 ±10.6	91.8±2.8
LSTM-RNN	92.9±5.0	88.9±4.8	90.8±2.8	88.9±6.8	91.5±5.4	89.9±3.5	**92.7±4.5**	92.7±2.9	92.7±3.5	91.1±2.7
Bi-RNN-5	95.6±2.6	90.7±7.3	93.0±4.9	91.0±2.8	90.6±5.0	90.7±3.0	90.7±4.8	96.1±3.9	93.2±2.6	92.5±2.5
EBiRNN-5	95.2±3.7	95.4±7.7	**95.0±3.6**	**95.5±0.7**	91.2±5.2	**93.3±2.9**	91.9±5.0	**96.3±3.4**	**94.0±3.7**	**94.5±0.7**
Bi-RNN-20	92.4±9.1	92.5±8.4	91.8±5.0	91.9±8.8	91.6±6.2	91.3±4.5	93.3±6.5	91.4±12.4	91.5±6.6	91.8±3.7
EBiRNN-20	95.9±4.4	92.3±8.8	93.7±4.6	93.2±7.2	93.9±6.8	93.2±4.4	94.7±6.4	95.8±6.5	95.1±5.1	94.4±2.6

TABLE IV
RESULTS COMPARISON WITH THE PREDICTION MADE AT 3.5S BEFORE THE MANEUVER.

Algorithms	Left Lane Change			Right Lane Change			Lane Keeping			General Ave (%)
	Pr (%)	*Re* (%)	*F1 Score*	*Pr* (%)	*Re* (%)	*F1 Score*	*Pr* (%)	*Re* (%)	*F1 Score*	
SVM	86.2±5.7	**91.8±4.0**	88.8±3.5	75.5±1.3	76.2±3.7	75.2±7.5	**90.3±6.3**	82.6±8.5	**86.0±5.3**	83.6±4.2
FFNN	87.4±1.0	90.8±7.8	88.9±8.4	**91.6±2.7**	85.7±5.9	88.4±2.3	83.8±6.6	86.2±8.5	84.8±6.3	87.2±4.8
Naïve Bayes	85.0±4.9	76.8±8.2	80.6±6.1	77.5±7.0	82.9±6.5	79.7±2.9	82.9±1.1	**86.7±6.1**	84.6±8.2	81.9±5.8
HMM	67.4±5.5	81.9±10.9	73.7±6.1	70.9±7.2	86.4±5.6	77.7±4.8	72.0±8.6	39.0±9.2	50.3±9.1	69.4±5.3
RF	**95.5±4.5**	88.0±7.4	91.4±3.4	83.6±3.4	87.4±5.6	85.4±6.5	80.5±10.0	85.2±6.6	82.6±6.9	86.6±4.5
LSTM-RNN	88.5±13.2	92.6±6.0	89.9±7.2	82.9±8.6	**90.5±8.2**	86.1±5.7	87.0±13.6	73.9±14.2	78.4±8.0	85.6±5.7
Bi-RNN-5	90.8±8.1	91.2±6.0	90.7±4.0	88.0±5.6	92.0±6.6	**89.6±1.8**	85.9±6.7	79.8±13.2	82.1±7.1	87.8±3.2
EBiRNN-5	93.7±6.6	90.6±5.7	**91.9±3.8**	82.8±6.8	89.1±8.8	85.7±7.1	86.6±5.4	82.7±1.2	84.4±8.3	**89.5±4.7**
Bi-RNN-20	91.0±8.9	85.3±9.0	87.7±6.8	82.3±9.3	88.4±11.1	84.6±6.8	84.4±8.3	81.6±10.8	82.3±6.0	85.2±5.1
EBiRNN-20	93.8±5.9	90.1±6.3	91.6±2.9	88.9±9.9	84.1±9.1	86.2±8.5	79.3±10.3	87.2±8.2	82.8±8.3	87.3±5.4

To analyze the difference between the proposed method and the existing ones, a two-sample one-tailed T-Test analysis between EBiRNN and BiRNN is proposed. The null hypothesis (H_0) and the alternative hypothesis (H_1) are first defined for the T-Test as follows.

$$[H_0]: \textit{The EBiRNN method has similar performance with BiRNN on the LCII task}$$
$$[H_1]: \textit{The EBiRNN method performs better than the BiRNN on the LCII task}$$

The two-sample T-Test is calculated as follows with an alpha level equal to 0.05.

$$t = \frac{\overline{u_1} - \overline{u_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{17}$$

where $\overline{u_1}$ and $\overline{u_2}$ are the mean performance of the EBiRNN and BiRNN, s_1 and s_2 are the standard deviation, and n_1, n_2 are the sample sizes, which are determined as 20.

Based on the two-sample T-Test, the analysis of the two algorithms with different prediction horizon is shown in Table 5. The *F1* score for each maneuver and the overall average prediction results are compared. As shown in Table 5, the H_0 means no significant improvement existing for EBiRNN compared with the BiRNN, and the null hypothesis cannot be rejected, while the index H_1 means that EBiRNN achieves considerable improvement with an alpha level equal to 0.05.

TABLE V
STATISTICAL HYPOTHESIS TEST RESULT (TWO-SAMPLE T-TEST) FOR EBIRNN AND BIRNN MODELS

Prediction Horizon	Lane Change Left (*F1*)	Lane Change Right (*F1*)	Lane Keeping (*F1*)	General Ave (%)
0.5 s	0.100 (H0)	0.038 (H1)	0.035 (H1)	0.010 (H1)
2.0 s	0.100 (H0)	0.097 (H0)	0.035 (H1)	0.008 (H1)
3.5 s	0.001 (H1)	0.256 (H0)	0.419 (H0)	0.111(H0)

Based on Table 5, it is shown that when the intention prediction is made 0.5 s before the maneuver starts, the *p_value* of the general performance of EBiRNN is 0.01, which is smaller than the alpha level. The null hypothesis can be rejected, and the EBiRNN shows a significant advantage over the BiRNN model at the 0.05 significance level. Similarly, the null hypothesis can be rejected when the intention prediction is made two seconds before the maneuver. However, when the intention is inferred as 3.5 s before the maneuver, the null hypothesis cannot be rejected, indicating there is no significant difference between the two models at the alpha significance level. This may because there is no sufficient evidence and cues to be used to precisely infer driver intent when the prediction is made too early. Hence, none of the models can generate a confident and accurate performance. Regarding the F1-scores for the three maneuvers, the significant advantage of EBiRNN only occurs when the prediction is made 0.5 s prior to the maneuver, and the lane-keeping intention generates the most accurate results. In sum, the proposed EBiRNN approach is shown to be more precise than the conventional BiRNN method when the intention is inferred with the prediction horizon being less than two seconds.

Fig.12 illustrates the confusion matrix of the intention inference results using EBiLSTM-RNN and FFNN for one test. The general inference accuracy of the EBiRNN for the three maneuvers is 96.7%. Unlike [54], which trains the non-temporal discriminative model by concatenating the feature vector at each step and form a large feature set, the FFNN is trained with a much smaller feature vector in this study. For each training sample, the 18-D temporal sequence data is transformed into a 72×1 feature vector, consisting of the mean, standard deviation, maximum value, minimum value of the 18 signal channels. The 72-dimensional feature vector is much smaller compared to the 3840-dimensional vector reported in [54]. The intention prediction results using the FFNN is shown in the right subplot of Fig. 12. The general performance of the FFNN model is 93.4%. Although the detection rate is lower than the RNN models, the FFNN method still shows an acceptable ability on the intention prediction task and achieves better results compared with other discriminative models.

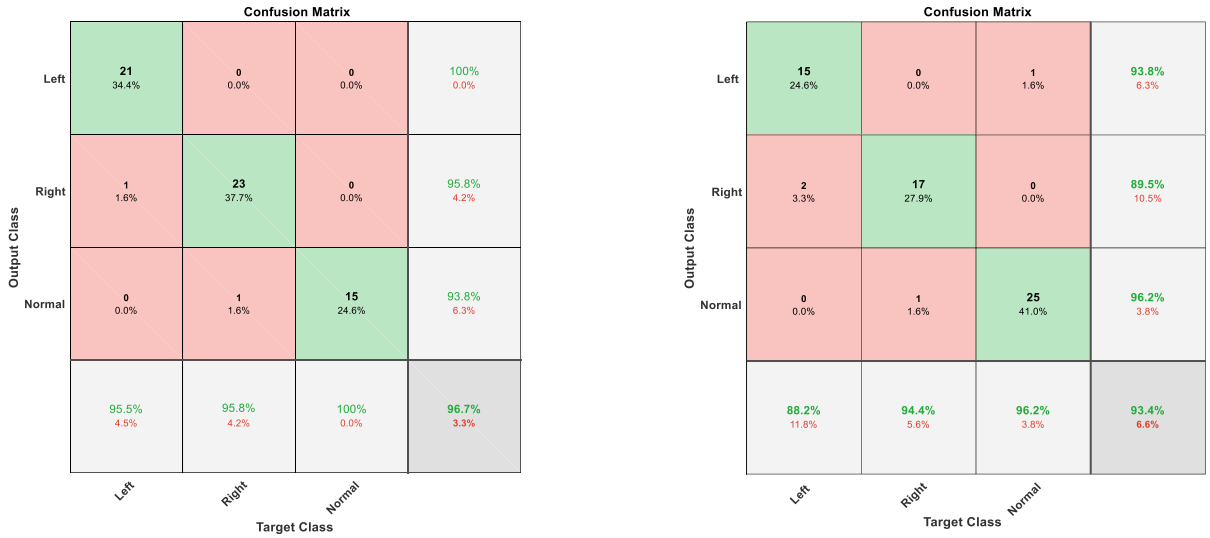

Fig. 12. An illustration of the confusion matrix for lane change intention inference using EBiLSTM-RNN and FFNN with the prediction made 0.5s before the maneuver starts.

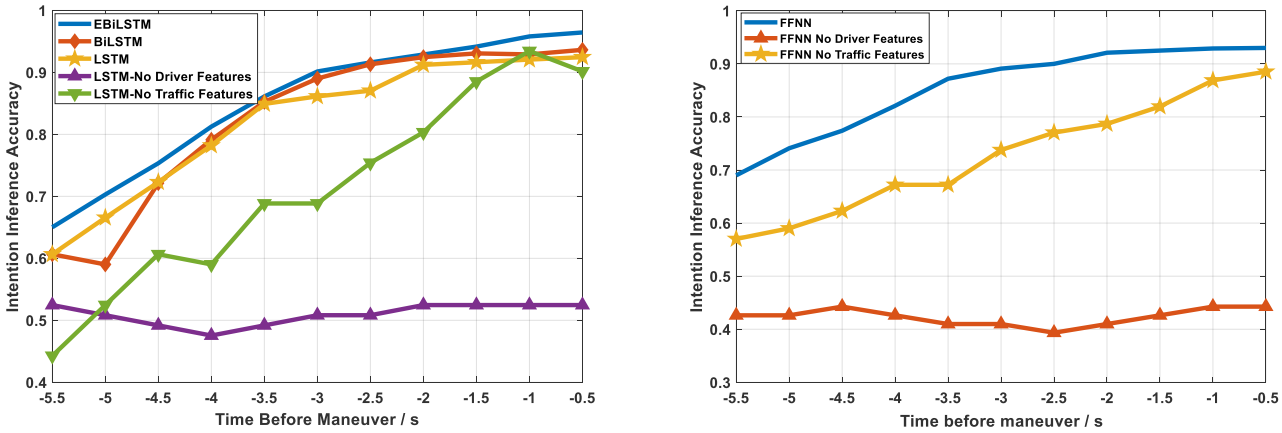

Fig. 13. Prediction performance vs. time-to-maneuver for lane change with EBiLSTM-RNN models and FFNN model.

The model prediction performance for the RNN model and FFNN model are further evaluated using the sliding prediction horizon method. Specifically, the testing sequence is shifted back for every 0.5 seconds. Moreover, the impact of different features is also evaluated for different models, as shown in Fig. 13. The EBiRNN and FFNN models are trained and evaluated based on

different feature sequences. The “no driver” lines for each model indicates that the model is trained without using the driver-related features (head pose and eye gaze features). In contrast, the “no traffic” lines represent that the models are trained without using the lane marking features. As shown in the left subplot of Fig. 13, the RNN models (both EBiRNN and BiRNN) are more accurate than the FFNN model when inferring the intention 2.5 seconds earlier. While the FFNN model gives a slightly better detection result when the intention is inferred more than four seconds. One explanation for the above results could be the FC layers and structure of the FC layer may dominate the classification. This can be a good motivation to exploit the impact of the structure selection of the FC layers on the model robustness and precision in the future. Another explanation is, as the testing sequence move earlier, more irrelevant information is involved in the temporal sequence, which would confuse the RNN model on the intention inference task. However, since the FFNN uses the statistic features of the sequence data, such as maximum and standard derivate, the irrelevant information may have less influence on the FFNN models. And the FFNN may still be able to capture the significant features within the temporal sequence. Therefore, in the future, more efficient time-series features can be exploited to increase the model performance.

Moreover, it is shown that the driver-related features play a critical role in the early stage of the driver intention estimation procedure. The models that are trained with driver features can also achieve a reasonable intention inference rate. In contrast, the models that are trained without driver features show significant low intention estimation accuracy when making predictions before the maneuvers. Based on the results from the two different methods, the models with only traffic context features show consistently low detection accuracy of the lane change intention. While the model based on the driver's behavioral features achieves slightly lower accuracy than the model with a full feature vector. This means the driver’s checking behaviors are more important than the traffic and vehicle features in terms of the early inference of the lane change intention.

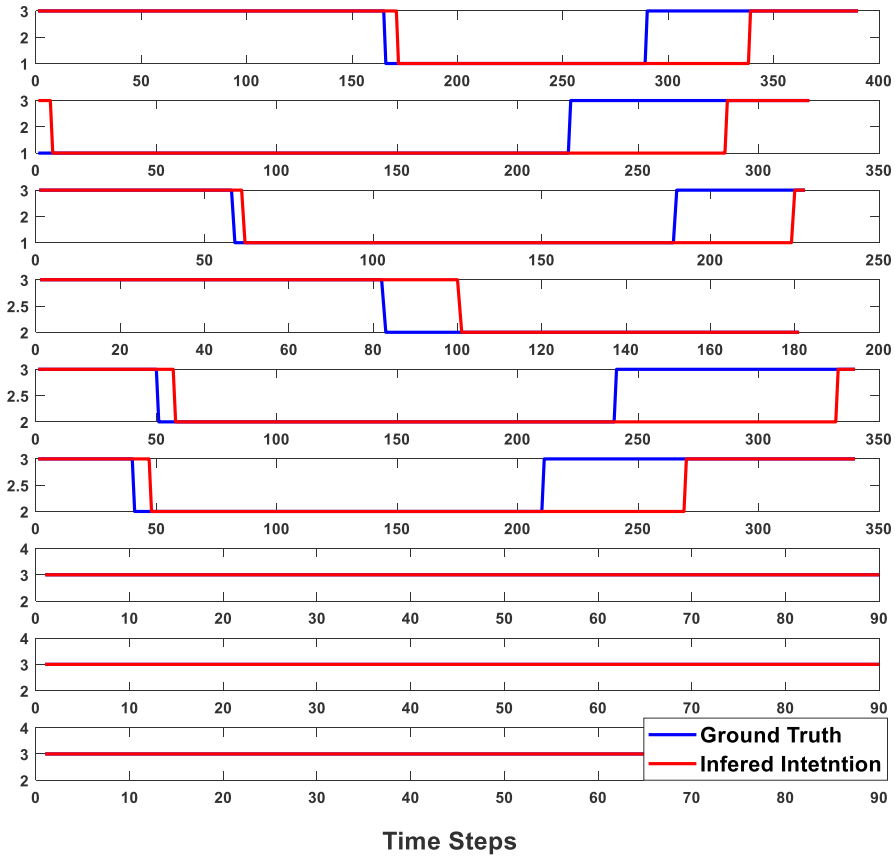

Fig. 14. Real-Time lane change intention inference results considering the three different maneuvers.

Fig. 14 illustrates the real-time intention inference results with the EBiLSTM model. The testing dataset is expanded about five seconds before and after a lane change maneuver, to simulate a whole inference cycle for the lane change intention. The typical lane-keeping intent is labeled with number three, while state number 2 and 1 refers to the right lane change and left lane change, respectively. The first three rows are the prediction results of left lane change scenarios, and the three rows in the middle are the right lane change scenarios. The bottom three rows indicate the lane-keeping process. Here the lane change initiation point is manually selected as the time at the first mirror checking behavior before a lane change maneuver. And the finishing point is determined as the moment when the vehicle has just crossed the lane. As shown in Fig. 14, the EBiLSTM-based LCII system can efficiently detect the lane change intention in a very early stage after the driver generates the mirror checking behaviors. As the

RNN models take the sequence data as inputs, the lane change intention signal can still be made even after the vehicle has crossed the lane. However, this should not be a significant concern in real-world implementation since the LCII system can be integrated into the LDW system. Thus, the lane departure can be recognized, and the intention state can be recovered.

VI. Discussion and Future Works

In this study, a DII system that focuses on the lane change maneuver is proposed based on the naturalistic highway driving data. It has been found that it is possible to infer the human driver's intention before the driver initiates the lane change maneuver. However, there are still some works that are expected to be investigated before the DII system can be implemented in the real world.

First, the intention inference system in this study assumes that human drivers always concentrate on driving tasks during the experiments. The experiment duration for each participant is about one hour, which can make sure the drivers are not overloaded. However, for real-world application purposes, the DII system should not work as an isolated function. The system needs to cooperate with other driver assistance systems to make a holistic prediction. It is found that when the drivers are overloaded, their behaviors change significantly, and the driver workload is highly related to the intention inference system [69]. Therefore, an integrated system that combines the driver workload estimation and intention inference system is expected for real-time driver assistance.

Second, as vehicle intelligence increases significantly nowadays, vehicle automation is more likely to take or share the control authority from or with the driver. What happens if drivers are performing the secondary tasks, and how to infer their intentions and situation awareness correctly? The DII system must comprehensively recognize the driver's status before estimating the intention, and determine whether the driver is focusing on the driving task or not at first. By considering driver distraction and intention has a whole, the control conflicts between the driver and the vehicle can be minimized. Meanwhile, for those partially automated vehicles, the estimation of secondary tasks will help to determine the drivers' intention and understand whether they are capable of taking over the control in some emergency tasks [70] [71].

Finally, more challenging work is to learn a more comprehensive understanding of the cognitive intention generation process according to the traffic context and human behaviors. By analyzing the driver intention concerning the current traffic context, a human-like intention generation scheme can be developed for automated vehicles, considering the traffic roles, ride comfort, and even ethical issues. This knowledge will benefit the design of intelligent decision-making system for autonomous vehicles.

VII. Conclusion

In this study, a driver LCII system is proposed based on the ensemble LSTM-RNN structure. The experiment data is collected in real-world highway scenarios. A general framework of intention inference process is introduced. According to the frame, the traffic context is viewed as the stimuli for driver intention, while the driver behaviors and the vehicle dynamics are seen as the corresponding responses. Based on the four critical moments of the lane change process, statistical analysis of the driver behaviors are proposed. It is found that the driver tends to perform the first mirror checking action at six seconds on average before the lane change maneuver happens, and it takes about two seconds for the vehicle to cross the lane. Multiple baseline methods are evaluated and compared with the EBiRNN model to show the efficiency of the proposed ensemble method. The results suggest that the EBiLSTM model can achieve an average intention inference accuracy of 96.1% with a 0.5 s prediction horizon. Future works will focus on the comprehensive analysis of the driver's intention towards a more intelligent driver-vehicle collaboration and the construction of a holistic driver reasoning system based on multi-modal driver status.

ACKNOWLEDGEMENT

This work was supported in part by the SUG-NAP Grant (No. M4082268.050) of Nanyang Technological University, Singapore, and the Intra-Create Seed Collaboration Grant Project (NRF2019-ITS005-0011), Singapore.

References

[1] Birrell, Stewart A., et al. "How driver behavior and parking alignment affects inductive charging systems for electric vehicles." *Transportation Research Part C: Emerging Technologies* 58 (2015): 721-731.

[2] Bellis, Elizabeth, and Jim Page. National motor vehicle crash causation survey (NMVCCS) SAS analytical user's manual. No. HS-811 053. 2008.

[3] Ohn-Bar, Eshed, and Mohan Manubhai Trivedi. "Looking at humans in the age of self-driving and highly automated vehicles." *IEEE Transactions on* Intelligent Vehicles 1.1 (2016): 90-104.

[4] Pursula, Matti. "Simulation of traffic systems-an overview." *Journal of geographic information and decision analysis* 3.1 (1999): 1-8.

[5] Xing, Yang, et al. "Driver activity recognition for intelligent vehicles: A deep learning approach." *IEEE Transactions on Vehicular Technology* 68.6 (2019): 5379-5390.

[6] Lv, Chen, et al. "Characterization of driver neuromuscular dynamics for human–automation collaboration design of automated vehicles." *IEEE/ASME Transactions on Mechatronics* 23.6 (2018): 2558-2567.

[7] Xing, Yang, et al. "Dynamic integration and online evaluation of vision-based lane detection algorithms." *IET Intelligent Transport Systems* 13.1 (2018): 55-62.

[8] Chen, Yixin, Mohammed Quddus, and Xuesong Wang. "Impact of combined alignments on lane departure: a simulator study for mountainous freeways." *Transportation research part C: emerging technologies* 86 (2018): 346-359.

[9] Milanés, Vicente, et al. "Cooperative adaptive cruise control in real traffic situations." *IEEE Transactions on Intelligent Transportation Systems* 15.1 (2014): 296-305.

[10] Adell, Emeli, András Várhelyi, and Mario Dalla Fontana. "The effects of a driver assistance system for safe speed and safe distance–a real-life field study." *Transportation research part C: emerging technologies* 19.1 (2011): 145-155.

[11] Tran, Duy, et al. "A Human-Vehicle Collaborative Driving Framework for Driver Assistance." *IEEE Transactions on Intelligent Transportation Systems* 20.9 (2018): 3470-3485.

[12] Zhang, Pinchao, and Zhen Sean Qian. "Managing traffic with raffles." Transportation Research Part C: Emerging *Technologies* 107 (2019): 490-509.

[13] Xing, Yang, et al. "Driver lane change intention inference for intelligent vehicles: framework, survey, and challenges." *IEEE Transactions on Vehicular Technology* 68.5 (2019): 4377-4390.

[14] Doshi, Anup, and Mohan Manubhai Trivedi. "On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes." *IEEE Transactions on Intelligent Transportation Systems* 10.3 (2009): 453-462.

[15] Jain, Ashesh, et al. "Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture." *arXiv preprint arXiv:1601.00740* (2016).

[16] Berndt, Holger, Jorg Emmert, and Klaus Dietmayer. "Continuous driver intention recognition with hidden markov models." *Intelligent Transportation* Systems, 2008. ITSC 2008. 11th International IEEE Conference on. IEEE, 2008.

[17] Yagil, Dana. "Reasoned Action and Irrational Motives: A Prediction of Drivers' Intention to Violate Traffic Laws 1." *Journal of Applied Social Psychology* 31.4 (2001): 720-739.

[18] Pentland, Alex, and Andrew Liu. "Modeling and prediction of human behavior." *Neural computation* 11.1 (1999): 229-242.

[19] Li, Mingjun, et al. "Shared control driver assistance system based on driving intention and situation assessment." *IEEE Transactions on Industrial Informatics* 14.11 (2018): 4982-4994.

[20] Benloucif, Amir, et al. "Cooperative trajectory planning for haptic shared control between driver and automation in highway driving." *IEEE Transactions on Industrial Electronics* 66.12 (2019): 9846-9857.

[21] Kuefler, Alex, Jeremy Morton, Tim Wheeler, and Mykel Kochenderfer. "Imitating driver behavior with generative adversarial networks." In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 204-211. IEEE, 2017.

[22] Fridman, Lex. "Human-centered autonomous vehicle systems: Principles of effective shared autonomy." *arXiv preprint arXiv:*1810.01835 (2018).

[23] Krawczyk, Bartosz, et al. "Ensemble learning for data stream analysis: A survey." *Information Fusion* 37 (2017): 132-156.

[24] Liaw, Andy, and Matthew Wiener. "Classification and regression by random Forest." *R news* 2.3 (2002): 18-22.

[25] Izmailov, Pavel, Dmitrii Podoprikhin, Timur Garipov, Dmitry Vetrov, and Andrew Gordon Wilson. "Averaging weights leads to wider optima and better generalization." *In 34th Conference on Uncertainty in Artificial Intelligence 2018*, UAI 2018, pp. 876-885. Association For Uncertainty in Artificial Intelligence (AUAI), 2018.

[26] Webb, Geoffrey I., and Zijian Zheng. "Multi-strategy ensemble learning: Reducing error by combining ensemble learning techniques." *IEEE Transactions on Knowledge and Data Engineering* 16.8 (2004): 980-991.

[27] Liu, Andrew, and Alex Pentland. "Towards real-time recognition of driver intentions." *Intelligent Transportation System, 1997. ITSC'97., IEEE Conference on*. IEEE, 1997.

[28] Oliver, Nuria, and Alex P. Pentland. "Driver behavior recognition and prediction in a SmartCar." *PROC SPIE INT SOC OPT ENG*. Vol. 4023. 2000.

[29] Liebner, Martin, and Felix Klanner. "Driver intent inference and risk assessment." *Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort* (2014): 1-20.

[30] Beauchemin, Steven S., et al. "Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality." *IEEE Transactions on Instrumentation and Measurement* 61.2 (2012): 391-401.

[31] Gwon, Gi-Poong, et al. "Generation of a precise and efficient lane-level road map for intelligent vehicle systems." *IEEE Transactions on Vehicular Technology* 66.6 (2017): 4517-4533.

[32] Wang, Yafei, et al. "Vision-based lateral state estimation for integrated control of automated vehicles considering multirate and unevenly delayed measurements." *IEEE/ASME Transactions on Mechatronics* 23.6 (2018): 2619-2627.

[33] Pi, Xidong, Wei Ma, and Zhen Sean Qian. "A general formulation for multi-modal dynamic traffic assignment considering multi-class vehicles, public transit and parking." *Transportation Research Part C: Emerging Technologies* 104 (2019): 369-389.

[34] Gou, Chao, et al. "A joint cascaded framework for simultaneous eye detection and eye state estimation." *Pattern Recognition* 67 (2017): 23-31.

[35] Chen, Jixu, and Qiang Ji. "A probabilistic approach to online eye gaze tracking without explicit personal calibration." *IEEE Transactions on Image Processing* 24.3 (2015): 1076-1086.

[36] Haufe, Stefan, et al. "EEG potentials predict upcoming emergency brakings during simulated driving." *Journal of neural engineering* 8.5 (2011): 056001.

[37] Doshi, Anup, and Mohan M. Trivedi. "Tactical driver behavior prediction and intent inference: A review." *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*. IEEE, 2011.

[38] Salvucci, Dario D., et al. "Lane-change detection using a computational driver model." *Human factors* 49.3 (2007): 532-542.

[39] Toledo-Moreo, Rafael, and Miguel A. Zamora-Izquierdo. "IMM-based lane-change prediction in highways with low-cost GPS/INS." *IEEE Transactions on Intelligent Transportation Systems* 10.1 (2009): 180-185.

[40] Salvucci, Dario D., and Andrew Liu. "The time course of a lane change: Driver control and eye-movement behavior." *Transportation research part F: traffic psychology and behaviour* 5.2 (2002): 123-132.

[41] Schmidt, Kim, et al. "A mathematical model for predicting lane changes using the steering wheel angle." *Journal of safety research* 49 (2014): 85-e1.

[42] Driggs-Campbell, Katherine, and Ruzena Bajcsy. "Identifying modes of intent from driver behaviors in dynamic environments." *Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on*. IEEE, 2015.

[43] Henning, Matthias J., et al. "Modelling driver behaviour in order to infer the intention to change lanes." *Proceedings of European Conference on Human Centred Design for Intelligent Transport Systems*. Vol. 113. 2008.

[44] Doshi, Anup, and Mohan Manubhai Trivedi. "On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes." *IEEE Transactions on Intelligent Transportation Systems* 10.3 (2009): 453-462.

[45] Lethaus, Firas, et al. "A comparison of selected simple supervised learning algorithms to predict driver intent based on gaze data." *Neurocomputing* 121 (2013): 108-130.

[46] Murphy-Chutorian, Erik, and Mohan Manubhai Trivedi. "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness." *IEEE Transactions on intelligent transportation systems* 11.2 (2010): 300-311.

[47] Jang, Young-Min, Rammohan Mallipeddi, and Minho Lee. "Identification of human implicit visual search intention based on eye movement and pupillary analysis." *User Modeling and User-Adapted Interaction* 24.4 (2014): 315-344.

[48] Zhou, Huiping, Makoto Itoh, and Toshiyuki Inagaki. "Influence of cognitively distracting activity on driver's eye movement during preparation of changing lanes." *SICE Annual Conference, 2008*. IEEE, 2008.

[49] Li, Keqiang, et al. "Lane changing intention recognition based on speech recognition models." *Transportation research part C: emerging technologies* 69 (2016): 497-514.

[50] Kasper, Dietmar, et al. "Object-oriented Bayesian networks for detection of lane change maneuvers." *IEEE Intelligent Transportation Systems Magazine* 4.3 (2012): 19-31.

[51] McCall, Joel C., et al. "Lane change intent analysis using robust operators and sparse Bayesian learning." *IEEE Transactions on Intelligent Transportation Systems* 8.3 (2007): 431-440.

[52] Doshi, Anup, Brendan Morris, and Mohan Trivedi. "On-road prediction of driver's intent with multimodal sensory cues." *IEEE Pervasive Computing* 10.3 (2011): 22-34.

[53] Jain, Ashesh, Hema S. Koppula, Bharad Raghavan, Shane Soh, and Ashutosh Saxena. "Car that knows before you do: Anticipating maneuvers via learning temporal driving models." *In Proceedings of the IEEE International Conference on Computer Vision*, pp. 3182-3190. 2015.

[54] Michon, John A. "A critical view of driver behavior models: what do we know, what should we do?." *Human behavior and traffic safety*. Springer, Boston, MA, 1985. 485-524.

[55] Baltrušaitis, Tadas, Peter Robinson, and Louis-Philippe Morency. "Openface: an open source facial behavior analysis toolkit." *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 2016.

[56] Baltrusaitis, Tadas, Peter Robinson, and Louis-Philippe Morency. "Constrained local neural fields for robust facial landmark detection in the wild." *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*. IEEE, 2013.

[57] Wood, Erroll, et al. "Rendering of eyes for eye-shape registration and gaze estimation." *Proceedings of the IEEE International Conference on Computer Vision*. 2015.

[58] Zyner, Alex, Stewart Worrall, and Eduardo Nebot. "A recurrent neural network solution for predicting driver intention at unsignalized intersections." *IEEE Robotics and Automation Letters* 3.3 (2018): 1759-1764.

[59] Morton, Jeremy, Tim A. Wheeler, and Mykel J. Kochenderfer. "Analysis of recurrent neural networks for probabilistic modeling of driver behavior." *IEEE Transactions on Intelligent Transportation Systems* 18.5 (2017): 1289-1298.

[60] Xing, Yang, et al. "Energy oriented driving behavior analysis and personalized prediction of vehicle states with joint time series modeling." *Applied Energy* 261 (2020): 114471.

[61] Schuster, Mike, and Kuldip K. Paliwal. "Bidirectional recurrent neural networks." *IEEE transactions on Signal Processing* 45.11 (1997): 2673-2681.

[62] Bengio, Yoshua, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult." *IEEE transactions on neural networks* 5.2 (1994): 157-166.

[63] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." *Neural computation* 9.8 (1997): 1735-1780.

[64] Greff, Klaus, et al. "LSTM: A search space odyssey." *IEEE transactions on neural networks and learning systems* 28.10 (2017): 2222-2232.

[65] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.

[66] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv*:1412.6980 (2014).

[67] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249-256. 2010.

[68] Saxe, Andrew M., James L. McClelland, and Surya Ganguli. "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks." *arXiv preprint arXiv:1312.6120* (2013).

[69] Xing, Yang, et al. "Driver workload estimation using a novel hybrid method of error reduction ratio causality and support vector machine." *Measurement* 114 (2018): 390-397.

[70] Chen, Huei-Yen Winnie, et al. "Self-reported engagement in driver distraction: An application of the Theory of Planned Behaviour." *Transportation research part F: traffic psychology and behaviour* 38 (2016): 151-163.

[71] Eriksson, Alexander, and Neville A. Stanton. "Takeover time in highly automated vehicles: noncritical transitions to and from manual control." *Human factors* 59.4 (2017): 689-705.