IEEE Open Journal of
Intelligent Transportation
Systems

Received 22 November 2021; revised 19 April 2022; accepted 26 July 2022. Date of publication 31 August 2022; date of current version 12 September 2022.
Digital Object Identifier 10.1109/OJITS.2022.3197296

Driver Intention Recognition: State-of-the-Art Review

KOEN VELLENGA [1,2], H. JOE STEINHAUER [1], ALEXANDER KARLSSON [1], GÖRAN FALKMAN [1], ASLI RHODIN [2], AND ASHOK CHAITANYA KOPPISETTY [2]

[1] School of Informatics, University of Skövde, 541 28 Skövde, Sweden

[2] Department of Complete Vehicle Data Science, R&D, Volvo Car Corporation, 418 78 Gothenburg, Sweden

CORRESPONDING AUTHOR: K. VELLENGA (e-mail: koen.vellenga@volvocars.com)

This work was supported by the Intention Recognition in Real Time for Automotive 3D Situation Awareness (IRRA) Project (https://www.vinnova.se/p/intention-recognition-i-realtid-for-automotive-3d-situation-awareness-irra/).

ABSTRACT Every year worldwide more than one million people die and a further 50 million people are injured in traffic accidents. Therefore, the development of car safety features that actively support the driver in preventing accidents, is of utmost importance to reduce the number of injuries and fatalities. However, to establish this support it is necessary that the advanced driver assistance system (ADAS) understands the driver's intended behavior in advance. The growing variety of sensors available for vehicles together with improved computer vision techniques, hence led to increased research directed towards inferring the driver's intentions. This article reviews 64 driver intention recognition studies with regard to the maneuvers considered, the driving features included, the AI methods utilized, the achieved performance within the presented experiments, and the open challenges identified by the respected researchers. The article provides a high level analysis of the current technology readiness level of driver intention recognition technology to address the challenges to enable reliable driver intention recognition, such as the system architecture, implementation, and the purpose of the technology.

INDEX TERMS Driver intentions, intention recognition, driver behavior, driving maneuvers.

I. INTRODUCTION

THE NUMBER of annual traffic fatalities in the world rose from 1.15 million in the year 2000 to 1.35 million in 2016, and more than half of the victims were vulnerable road users (pedestrians, cyclists, and motorcyclists) [1]. In addition to the fatalities, an estimated number of 54 million injuries (e.g., limb fractures, traumatic brain injury, or amputations) were caused by road accidents worldwide in 2017 [2]. To realize a reduction of the number of traffic fatalities and injuries, car safety is named among one of the important factors [2], [3].

Car safety consists of passive, active, and pro-active safety. Passive safety [4], [5] aims to minimize the impact of a crash for the driver and the passengers (airbags, seat belts, or crumple zones). Active safety [6] aims to decrease human driving errors and avoid crashes from happening (e.g., an anti-lock brake system, or an autonomous emergency braking system). Given that traffic conflicts happen more often than actual collisions [7], pro-active safety quantifies the short-term risks of an accident and tries to influence the driver's behavior to avoid any conflicts (e.g., a front-collision warning system, or a driver monitoring system). The increasing number of sensors on a car enables the development of more advanced active safety features (e.g., driver intention recognition or driver behavior understanding). The sensor data can be divided into three categories: vehicle dynamics (e.g., velocity, yaw-rate, or steering wheel angle), driver state (e.g., head pose estimation, eye gaze direction, pupil size, or blink rate), or driving scene cues (e.g., localization of other road users, lane detection, or traffic sign detection).

For the driver intention recognition (DIR) research field to progress, it is fundamental to understand what is the current state-of-the-art (SOTA) of the field, and what gaps have to be addressed before the technology is ready to be integrated as part of the advanced driver assistance system (ADAS), or how it can support autonomous driving systems (ADS) in the future. From an industrial perspective, it is interesting to understand the needs (e.g., minimal computational capabilities, required software packages, or sensors) of the current

The review of this article was arranged by Associate Editor Chi-Hua Chen.

SOTA solutions, and whether the current hardware in the car is sufficient to fulfill the regulatory and legal requirements for an ADAS and ADS. Earlier DIR literature reviews from Doshi and Trivedi (2011) [8], and Lefèvre *et al.* (2014) [9] did not cover any deep learning methods. A more recent survey from Xing *et al.* (2019) [10] handpicked 21 lane change intention articles and presented an overview. Thus, this paper covers the need for a complete overview and analysis of the current DIR research field and formalizes the gaps that must be overcome to integrate the technology into a car. We focus on DIR methods that aim to recognize the intentions of the driver of the car. We performed a systematic literature review that analyzes studies from 2008 – 2020 based on experimental design, data collection, features, sampling distribution, applied methods, performance, and the listed open challenges. Due to the contextual nature of the DIR studies, no comparison can be made based on the reported performance. The distribution of the reported performance is visualized. Accordingly, a high level estimation of the technology readiness level is presented.

The article is organized as follows: Section II covers the concept of intentions, applied methods, and technology readiness levels. Section III describes the research strategy and eligibility criteria for this literature review. Section IV covers the evaluation of the experimental set-up, data collection, sampling and distribution, applied methodologies, open challenges, and a high level technology readiness level estimation. In Section V the identified gaps of current DIR approaches are discussed. Section VI concludes the key findings and suggests steps forward based on this review.

II. BACKGROUND

First, the intention recognition terminology is introduced, and its differences compared to the terms actions, activities, goals, and plans are highlighted with a driving scenario example. The term agent is in principle anything that can perceive the environment through sensors [11], but in this paper we limited the scope to human road users and refer to the driver of the considered car as the agent. A high-level introduction of the methods applied in single-agent driver intention recognition studies is presented, followed by an outline of evaluation metrics to understand how the performance of the methods can be assessed. Lastly, technology readiness levels are explained and their usage is discussed briefly.

A. INTENTION RECOGNITION

We use the term intention recognition as the identification of what an observed agent is aspiring to do in the immediate future (for example, described by [12], [13]). Where the immediate future refers to the upcoming seconds depending on the maneuver type and driving scene. The recognition of the observed agent's intention allows one to interact with that agent and to proactively adjust one's behavior to avoid accidents or solve problems that require cooperation.

For humans, intention recognition comes naturally and has been investigated in the literature for a long time, for example by Johansson [14], Blakemore and Decety [15], Gallese *et al.* [16], Gallese and Goldman [17], and Heider and Simmel [18]. Following the Theory of Mind [19], intentions, together with beliefs and desires, are mental states [20], [21], [22]. Here, beliefs denote everything a person knows about the world, desires describe the person's wish to achieve some goal, and intentions describe that the person is wanting to act towards achieving the goal. Given that we are about to enter a hybrid era of both autonomous and human driven cars on the road [23], intention recognition will be an essential element for ADAS and ADS. The prediction of the driver's intentions and the prediction of the potential actions of other traffic participants enable the driver or the car to anticipate proactively. This ensures safe and comfortable driving. Thus, research on intention recognition is a part of most autonomous driving and ADAS projects.

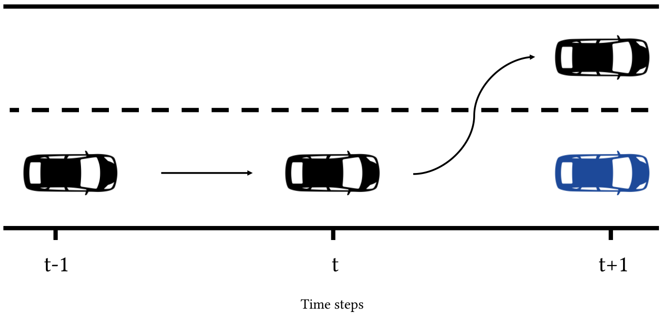

FIGURE 1. Visualization of a highway overtaking example which requires a lane change maneuver. When the ADAS or ADS infers the intention of the driver to overtake a car at time step t-1 it allows the system to support or intervene with the driver at time step t when the driver will make the first move (action). For example, if a car is approaching very fast on the left lane, the lane change maneuver could lead to a dangerous situation. Hence, the intention recognition performed by the ADAS allows for anticipation of the scenario and supports the driver to execute the intended maneuver safely.

Instead of the term intention recognition, several other terms can be found in the literature that are often used interchangeably. For example, goal recognition, action recognition, and plan recognition, as well as activity recognition and behavior recognition. Even though the differences between those terms might not appear important from a technical perspective, we would like to illustrate them with an example (see Figure 1 for a visualization):

Suppose a car driver (our agent), who travels on a highway, approaches another car from behind and which the driver wishes to overtake. In this example, the driver's (desired) goal is to be in front of the other car. When one decides to pursue this goal, an intention is formed to overtake the car ahead. However, it is not required to act immediately on this intention. The driver may stay behind the car and wait for a good moment to change lanes and overtake. The moment one starts to overtake the car, a sequence of actions will be executed that takes one in front of the other car. This sequence of actions is denoted as the plan.

Usually, when using the terms actions and plans, one is concerned with the concrete actions in an exact sequence, for example looking over the shoulder, looking in the rear mirror, switching on the turn signals, turning the steering wheel to

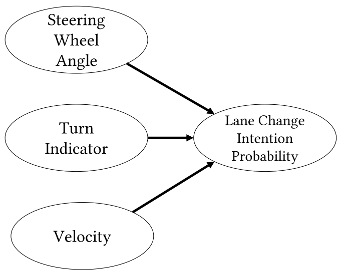

FIGURE 2. A Bayesian Network that shows the relation of the variables on the lane change intention probability.

the left, etc. When one is only concerned with the fact that the driver is overtaking another car, it is called an activity or behavior. For example, while overtaking another car actions such as a series of head movements, switching on the turn signals, and turning the steering wheel must be executed. When one observes a driver (or agent) doing those actions (in some order), one can conclude that the driver is overtaking.

It is important to be aware that only actions can be directly observed, all the others (intention, plan, goal, activity, or behavior) are subject to inference based on these observed actions and possible changes within the environment through someone's actions. Hence, action recognition always describes the first step for any of the denoted recognition tasks. Since the intention is to achieve the desired goal, intention recognition and goal recognition are closely related and often regarded as similar in the literature. However, the goal denotes the achieved state at the end of the executed sequence of actions, whereas the intention denotes the wish of pursuing the goal and thus the beginning of the maneuver even before the first action is taken. Therefore, intentions must be recognized before the first action whereas for pure goal recognition one has time after the first few actions have already taken place to identify the goal. Knowing the goal can be helpful to identify the intention of pursuing it. However, in traffic situations, traffic participants' goals are of secondary importance. For example, it does not matter why (for what goal) the driver wants to sheer out of his lane, what matters is that the ADAS can timely anticipate this action and smoothly support. Therefore, one needs to capture the intention of taking (the first) action which, in this example, is steering to the other lane, before the driver actually makes that move.

B. DRIVER INTENTION RECOGNITION METHODS

Approaches commonly used in DIR studies are outlined in this section. For every method, the intuition is described, followed by the limitations of the approach.

1) PROBABILISTIC GRAPHICAL MODELS

Predominantly three types of probabilistic graphical models (PGMs) are used in DIR studies: Bayesian networks (BN) [24], dynamic Bayesian networks (DBN) [25], [26], and hidden Markov models (HMM) [27].

PGMs use a graph-based representation to visualize the relationship between the modeled variables. In a BN, every node corresponds to a different variable. Directed links connect the nodes. When there is a link from node X to node Y, this means that X is the parent node of Y. A node can have multiple parents and multiple children, depending on the topology of the network. Every variable (node) in the network is described by a conditional probability distribution that represents the joint effect of the parent nodes on that variable (equation 1).

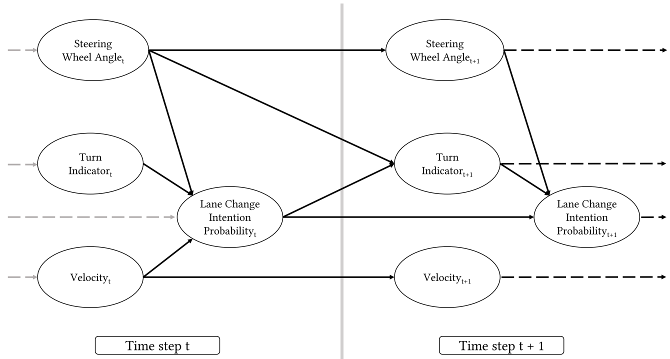

FIGURE 3. Two-time slice Bayesian network to illustrate how the variables influence each other over time.

$$P(X_1 = x_1, \ldots, X_n = x_n) \triangleq \prod_{i=1}^{n} p\big(x_i|x_{parents(i)}\big) \qquad (1)$$

where:

$P(X_n)$	= The joint distribution per node as a product of the conditional distributions
$p(x_i\|x_{parents(i)})$	= The conditional distribution of node x_i given the parent nodes

Considering the example introduced in Section II-A, it is possible to model the probability that the driver has an intention to change lanes. To keep the methodology examples clear, we assume that there are three independent variables that describe the vehicle dynamics (the steering wheel angle, state of the turn indicators, and velocity). The BN in Figure 2 is an example of how the variables relate to the identification (recognition) of the lane change intention of the driver.

A DBN extends a BN over time and is commonly visualized as a two-timeslice Bayesian network (2TBN) (refer to Figure 3 for an example). The variables in the network do not only depend on the state in the current time step but can also be affected by their state in the previous time step. For example, the lane change intention at time step t+1 depends on the turn indicator state at time step t+1, and on the lane change intention state at time step t.

A hidden Markov model (HMM) uses observed outputs to estimate an unobservable state. In our example, the driving intentions is the unobservable state that is based on the observable driving actions (Figure 4). HMMs rely on three assumptions [28]: the limited horizon assumption presumes that the previous state holds enough information to predict the next state (also called the Markov assumption),

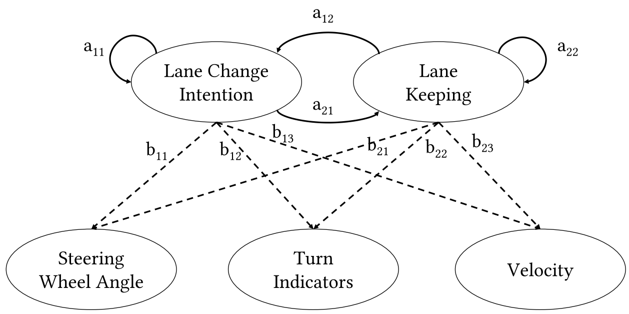

FIGURE 4. Hidden Markov Model that shows the relations between the driving intention states (lane change and lane keeping) and the observed variables.

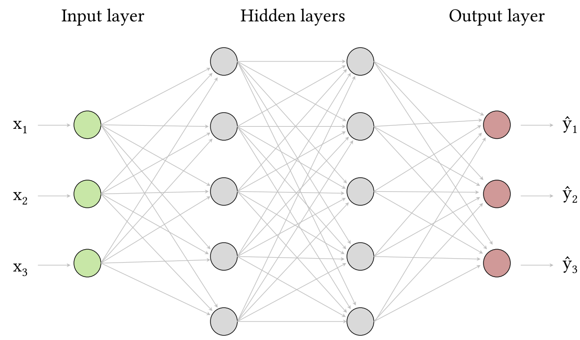

FIGURE 5. Feed-Forward Neural Network with the input variables and three output categories. The colors highlight the different layers (green for the input layer, grey for the hidden layers, and red for the output layer).

TABLE 1. Transition matrix A which corresponds to the example in Figure 4. S_0 represents the initial state and has an initial probability over the other states.

		S_0	$S_{lane\ change\ intention}$	$S_{lane\ keeping}$
	S_0	0	0.5	0.5
A =	$S_{lane\ change\ intention}$	0	0.75	0.25
	$S_{lane\ keeping}$	0	0.1	0.9

the stationary process assumption supposes that the conditional distributions are stable over time, and the independent output assumption states that the current observations stand alone from the preceding observations.

The transition probabilities between the unobservable states, illustrated by the a_{ij} edges in Figure 4 and equation 2, regulate how the new state of the driver at time step t+1 will be based on the state at time step t [29]. The probabilities together form the state transition matrix A (Table 1).

$$a_{ij} = P\big(q_{t+1} = s_j | q_t = s_j\big) \tag{2}$$

Where:

a_{ij} = Transition probability from state i to state j
q_t = State at time step t
s = State

$$b_{i(k)} = P\big(o_t = v_k | q_t = s_j\big) \tag{3}$$

Where:

$b_{j(k)}$ = Emission probability for observation k given state j
q_t = State at time step t
s = State
o_t = Observation at time step t
v = Observation

For every unobservable state, the emission probability represents hhow likely it is to observe variables that indicate a certain state. Refer to the b_{jk} edges in Figure 4 for the visual representation of the emission probabilities, and equation 3 for the formal notation. Every state has an emission matrix, consisting of the emission probabilities for that state. In the example, two emission matrices can be constructed (one for the lane change intention state and one for the lane-keeping intention state). If the emission probability for using the turn indicator is 80% for the lane change intention state and 20% for the lane-keeping intention state, then the probability that the driver uses the turn indicator is higher when the driver intends to change lanes.

Various estimation techniques can be used to establish the transition and emission matrices (parameters A and B) of an HMM. For example, the Baum-Welch expectation-maximization algorithm tries to find the maximum likelihood of the parameters given the observations, and the Viterbi max-sum algorithm aims to find the most likely assignment of the states [30], [31].

Every HMM is a single variable DBN and every discrete DBN is an HMM. The difference is the number of parameters of the models. The transition matrix of an HMM requires a connection between all variables and the hidden states, which results in an increasingly large transition matrix. Suppose, instead of having two Boolean state variables like in the example (lane change intention and lane-keeping intention), there are ten Boolean state variables. The corresponding HMM has 2^{10} states which leads to 2^{20} probabilities (over a million) to be estimated. In this case, a more practical solution would be to use a DBN transition model which allows modeling only the relevant relations between the variables. This reduces the number of probabilities to be computed but requires domain knowledge for the modeling task.

2) ARTIFICIAL NEURAL NETWORKS

Recent DIR studies rely often on (deep) artificial neural networks (ANNs) for recognizing driver intentions (e.g., [32], [33], [34]). Networks with at least two hidden layers are considered to be deep neural networks [35]. Refer to Figure 5 for a visual representation of a feed-forward neural network (FFNN). To handle sequential input data, the recurrent neural network (RNN) architecture can be used (Figure 6). A RNN reuses the output of a previous time step. The reuse of the output extends a RNN which can be regarded as a form of memory [36].

To train ANNs Rumelhart *et al.* [37] introduced the backpropagation training algorithm. Backpropagation is used for computing the gradients of the network and algorithms like stochastic gradient descent are used for learning the gradients [38], [39]. A loss function computes how well a network

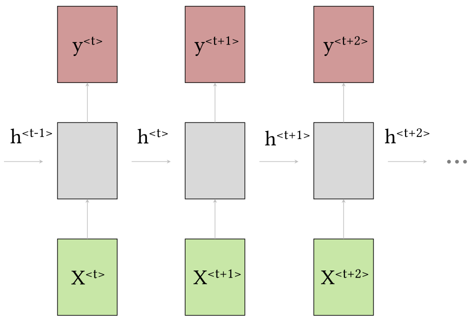

FIGURE 6. Unfolded recurrent neural network for three time steps. At every time step t, a vector of the previous time step is shared between the hidden layers to re-use the output in the next time step.

performs. For example, ANNs that perform binary classification can use the cross-entropy loss, which measures the difference between two probability distributions:

$$\mathcal{L}(z, y) = -\big[y\, log(z) + (1 - y)\, log(1 - z)\big] \tag{4}$$

Where:

$\mathcal{L}$ = Loss function to evaluate the model performance
z = Correctly predicted output
y = Actual output

The weights of an ANN are updated by computing the gradient of the loss for each weight (equation 5). For an RNN, the loss function is based on the sum of the loss at every time step, and the derivatives of the loss function for each weight are also computed at every time step [38], [39].

$$w \leftarrow w - \frac{\partial \mathcal{L}(z, y)}{\partial w} \tag{5}$$

Where:

α = Learning rate
w = Weight
$\frac{\partial \mathcal{L}(z,y)}{\partial w}$ = Gradient of the loss for each weight

The FFNN (Figure 5) and RNN (Figure 6) do not show the relationship between the variables, in contrast to, for example, a DBN. Through the training procedure, the network learns how to represent the relations between the variables in the hidden layers. Subsequently, the visual representations of an FFNN and RNN have no meaning other than displaying the architecture of the network. Even though DBNs, HMMs, and RNNs can be used for estimating an unobserved variable given a sequence of observations, RNNs deduce this latent variable deterministically based on the observed sequence instead of making use of, for example, a Markov chain [40].

3) EVALUATION METRICS

To avoid making misleading conclusions, it is crucial to select proper evaluation metrics and to determine the class balance of a dataset [41], [42]. A severe imbalanced distribution leads to poor performance of a classifier [43]. The standard evaluation metrics used by the DIR studies are briefly introduced below. Refer to Zheng [44] for an introduction of basic evaluation metrics.

The true positives (TP), and true negatives (TN) are the cases that are correctly classified. False negatives (FN) are cases that should have been positives, and false positives (FP) should have been labeled as another class. The fraction of the total number of correct predictions is obtained via computing the $accuracy = (TP+TN)/(TP+TN+FP+FN)$. The fraction of correctly predicted classes is calculated with the $precision = (TP)/(TP + FP)$. The number of a correct predictions given the total number of items present in a class is defined as the $recall = (TP)/(TP + FN)$. Last, the $F_1 score = (2 \cdot TP)/(2 \cdot TP+FP+FN)$ balances the precision and recall curve via combining the harmonic mean.

The Receiver Operating Characteristic (ROC) is constructed by plotting the true positive rate against the false positive rate. The ROC curve shows how many correct classifications can be yielded if more false positives are allowed [45], [46]. For discrete classifiers, only one confusion matrix is produced, however a logistic regression produces a value to which degree an instance belongs to a class. A threshold for that value can be applied to produce a binary classification. For each value of a certain threshold, a confusion matrix is produced, and thus a ROC point. The optimal threshold of the classifier depends on whether or not it is desired to accept more false positives to increase the true positive rate. Hanley and McNeil [47] note that the ROC area under the curve (AUC) value can be used to compare the performance of models. When using the ROC-AUC to compare different models, one should analyze whether the curves overlap rather than only comparing the AUC value [48]. A model can be disregarded if it is outperformed at every threshold step, but if there is an overlap it might indicate that an ensemble or fusion approach is beneficial for the performance. In cases with more than two classes, a multi-class ROC analysis should be performed where a pairwise comparison (one vs all) could be made to assess the performance of the model [49], [50]. Altogether, the evaluation metrics allow assessing model performance from different perspectives. Depending on the comparison needs, multiple metrics can support the assessment of the model performance.

Lastly, time is an essential parameter for the evaluation metrics applied to assess a DIR method. The overall goal of DIR methods is to predict whether the intended driving maneuvers are safe to execute and to pro-actively warn a driver if necessary. Previous surveys already highlighted the need for complementary evaluation metrics to consider the required time horizon that enables a driver to act upon the predicted information [9], [10]. Thus, the performance of a DIR model should be assessed over time to understand how well and far in advance an intention can be recognized.

C. TECHNOLOGY READINESS LEVELS

In the 1970s NASA introduced technology readiness levels (TRLs) to monitor and assess the maturation of technology to use in a particular application domain [51]. TRL1 (the

TABLE 2. Overview of the used search queries and databases.

Search query	Records retrieved	Date accessed	Database	Search type
(goal OR intention) AND (recognition OR prediction) AND (traffic OR lane OR road OR driver OR highway OR intersection OR pedestrian OR urban AND (bayesian OR markov OR "deep learning" OR hmm OR dbn or "hidden markov model" OR "neural network" OR "recurrent neural network" OR lstm OR fuzzy OR logic OR probabilistic))	787	2020-10-07	Scopus	Document search
TITLE-ABS-KEY(((goal* OR intent* OR intend* OR action* OR activity* or plan*) W/2 (recogni* OR predict*)) AND (driver* or driving*) AND (traffic* OR lane* OR road* OR "advanced driver assistance system") AND (highway* OR intersection* OR urban* OR city* OR bayes* OR dbn OR "dynamic bayesian network" OR markov OR "deep learning" OR *hmm OR "hidden markov model" OR "neural network" OR "recurrent neural network" OR lstm OR rnn OR fuzzy OR logic OR probabilistic))	368	2020-11-12	Scopus	Advanced search
TITLE-ABS-KEY(((goal* OR intent* OR intend* OR action* OR activity* or plan* or maneuver* or manoeuvre*) W/2 (recogni* OR predict*)) AND (driver* or driving*) AND (traffic* OR lane* OR road* OR "advanced driver assistance system") AND (highway* OR intersection* OR urban* OR city* OR bayes* OR dbn OR "dynamic bayesian network" OR markov OR "deep learning" OR *hmm OR "hidden markov model" OR "neural network" OR "recurrent neural network" OR lstm OR rnn OR fuzzy OR logic OR probabilistic OR "decision trees" OR "random forests" OR "ensemble"))	483	2022-03-22	Scopus	Advanced search
(traffic | urban | lane | highway | road | driver | pedestrian | bayesian | markov | "deep learning" | hmm | "neural network" | lstm | "fuzzy logic" | probabilistic) AND (goal | intention | plan) AND (recognition | prediction) allintitle: (driver AND intention)	163	2020-11-15	Scholar	Application: publish or perish.

lowest level) corresponds to the state where the respective technology is merely an existing idea to solve a problem or to improve existing technology, whereas TRL9 (the highest level) refers to the state when the technology is certified and deployed in industry (refer to [51], [52] for more extensive details on the TRL framework). An example of the TRL process was the journey to reduce jet engine noise, which led to the development of chevrons. In the 1980s, fundamental research on air-mixing devices was conducted (TRL1–2) and the first lab tests and concepts on paper were explored during the early 1990s (TRL3). TRL 4–5 was reached between 1995 and 1997 by conducting acoustic model tests and the first full-scale tests followed in the year after (TRL6). After validating the concept from 2001 till 2005 (TRL7), it was fully deployed into the market (TRL8–9). The gap between the TRLs indicates the number of conceptual studies, modeling, testing, or integration that has to be done before the technology is ready for real-world use. Apart from space and aero-engineering, other industries (e.g., oil, renewable energy, technology, and defense) have adopted the TRL framework to monitor their technological advances [53], [54], [55], [56], [57].

The application of TRLs has been criticized concerning validity, clarity, and completeness. Smith [58] highlighted that TRLs mainly focus on successful testing and integrating technology into real-world environments. In software engineering, changes are released continuously, but the technology ages quickly. In this scenario, it is unrealistic to assume that the technology remains at TRL9 in the absence of improvements and changes. Olechowski *et al.* [56] introduce 15 challenges for the TRL assessment, and noted subjectivity and imprecision of the scale as validity threats. When using a TRL classification to select a new technology, an individual who favors a certain technology might interpret the TRL higher than it is. Furthermore, the interpretation of the assessment itself, even when it is adjusted for a particular industry, could generate different results. In our case, the level should be interpreted as a guidance together with the identified gaps to understand what needs to be done before integrating DIR technology in an ADAS and for the field to move to the next TRL.

III. SYSTEMATIC REVIEW METHOD

We followed the guidelines of Moher *et al.* [59], Grant and Booth [60] and Snyder [61] to conduct this literature review. First, the used data sources are disclosed, followed by the rationale of the search parameters. At last, the screening and eligibility procedure describes the criteria for selecting which studies to include.

Google Scholar covers well over 300 million records [62], [63], and the Scopus database contains more than 75 million records [64]. Both platforms also include records from the IEEE Xplore, Sciencedirect, and Springerlink databases. To find as many relevant studies as possible, both Scopus and Google Scholar were used. The search queries considered: the title, the keywords, the abstract of the study and aimed to include experimental studies that have the objective to infer or predict the intention of a car driver but exclude trajectory prediction studies.

As discussed in Section II-A, the terms action, activity, goal, intention, and plan recognition are closely related and sometimes used interchangeably in the literature. To avoid missing any record that used a different term but does conduct an intention recognition experiment, we chose to include all closely related terms and to evaluate the eligibility. Previous literature reviews [8], [9], [10] highlighted several methods and environmental settings that were used in DIR experiments. As shown in Table 2 the first query had a wider perspective to include more studies, whereas the second query focused more specifically on DIR studies.

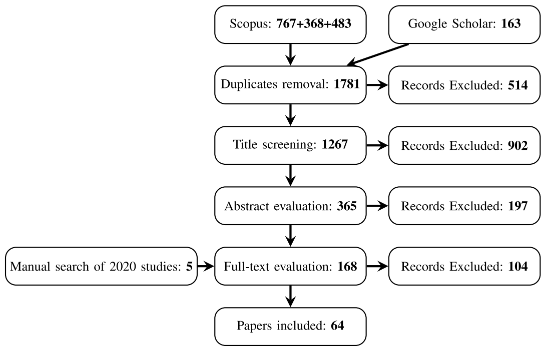

FIGURE 7. PRISMA diagram [59] to visualize the reporting of the identification, screening, eligibility and inclusion process of this literature review.

The third Scopus query also aims to include driver maneuver prediction studies, and the last query searches for driver intention studies on Google Scholar.

Figure 7 shows the screening and eligibility process of the search. 1781 papers were identified after removing the duplicates from the search results. Content analysis [65] is performed to systematically analyze the search results and to make valid and replicable inferences to understand how and what kind of results are reported in the DIR experiments. The papers are screened based on the title to determine whether the paper had a connection to the topic of DIR or inference. An abstract evaluation is conducted to understand if a driving intention recognition experiment is performed. The full-text review checks if the papers reported the results of the experiments. This step is necessary to exclude work-in-progress papers and to disregard papers that did not report their findings in detail. Subsequently, the reported intentions in the papers were evaluated. The most researched driving maneuver intentions were left lane changes, right lane changes, turning right, turning left, and driving straight/lane-keeping. Other driving intentions (such as u-turns, emergency lane changes, emergency turning, yielding, curvy road lane changes, lane change preparations, and sharp turning intentions) were excluded due to a limited coverage by the included studies which prohibits a high-level analysis. Papers not written in English, with zero citations two years after the publication, or only cited by the authors are also excluded. No limits were set on the publication year because we wanted to be able to identify trends in the literature over time. The initial search was conducted in October and November in 2020, an initial manual search was conducted in April 2021 to add studies published in 2020, and a final search query was conducted in March 2022 to include maneuver prediction studies up until 2020. In the end, 64 studies were used for the literature review.

IV. EVALUATION OF DRIVER INTENTION RECOGNITION STUDIES

Section IV-A introduces the considered scenarios and the used feature types in the included studies. Section IV-B covers how the studies collected data, and which open-source datasets are available. Section IV-C reviews the sample sizes and the number of instances per maneuver type in the DIR studies, followed by an overview of the applied methods and the reported performance. Section IV-D lists the open challenges mentioned by studies published between 2016–2020. Lastly, Section IV-E presents a high level estimation of the current TRL of the driver maneuver intention recognition field, and an indication of challenges that must be addressed to advance to the next TRL.

A. OVERVIEW OF THE INCLUDED STUDIES

Figures 8 and 9 show an overview of the maneuvers considered by the included studies. The number of studies that consider DIR has increased in recent years, and the majority is focused on lane change intentions. Roughly 17% of the included studies consider both lane change and turn maneuvers. Although, we acknowledge the effect that preprocessing strategies of the features can have on the DIR performance, we limit the feature analysis to a high-level categorization. To infer the intentions of a driver to perform a maneuver, three high level categories of features are used:

- Vehicle dynamics (e.g., steering wheel angle, pedal state, turn signal, acceleration, velocity, yaw rate).
- Driver state (e.g., eye tracking, head pose estimation, facial expression, drowsiness).
- Driving scene cues (e.g., lane detection, GPS map information, (vulnerable) road user detection, distance to other vehicles, or relative speed).

Figure 10 shows an overview of the feature categories that have been used over time. Note, that a study can include features from several categories. Driver state features are least often used to classify the driver's intentions (Figure 11). Driving scene cues have become the most popular features to include since 2015, but this is partially due to the increased use of the Next Generation Simulation (NGSIM) open-source dataset.

B. EXPERIMENTAL SET-UP

How and where researchers collect data can affect the results of an experiment. One example of the differences between the studies is the number of involved participants. For example, Gerdes [66] involved two participants, where Li *et al.* [67] recruited 47 participants. Apart from the varying characteristics of the involved participants across DIR studies, two methods are used to collect data: a driving simulator or a real-world driving experiment. The data collection setup can differ greatly across the included studies. For example, Lethaus *et al.* [68] built a simulator with an actual car chassis in a lab to perform simulated DIR experiments, whereas Kim *et al.* [69] used a screen, and a steering wheel. Of the included studies, two of the turn intention experiments and 13 of the lane change intention experiments used a driving simulator to collect data. In total roughly 20% of the included studies used a simulator.

Another way to research driving intentions is to leverage an open-source dataset. The NGSIM database was the only

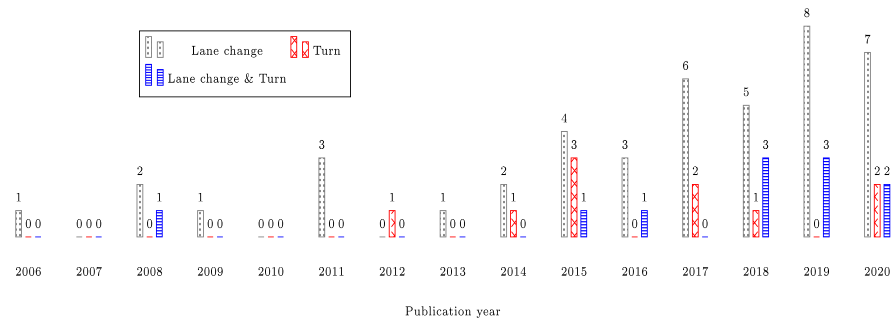

FIGURE 8. Overview of the types of maneuvers explored in the included DIR studies over time.

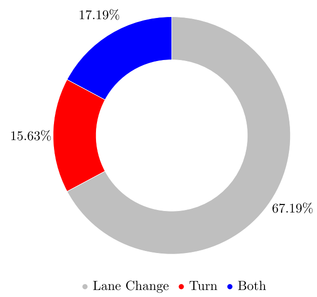

FIGURE 9. The maneuver type as a percentage of the total number of included studies (N=64).

available open-source dataset for years. A major drawback of the NGSIM database is that the dataset was collected through static cameras that recorded highways and intersections. Therefore, access to the driver state or vehicle dynamics features (apart from the direction and velocity) is impossible. However, in recent years multiple datasets have been published with intention labels, multiple environments, and video footage of the driver and the driving scene. The following open-source datasets are found in the included studies:

- For the NGSIM datasets [70] stationary cameras were placed at highways and intersections in Los Angeles in 2007. The datasets were constructed with the purpose to enable the development of intelligent transportation systems based on real-world data. The database consists of 90 minutes of bird's-eye-view video footage, vehicle trajectories, and lane change features.
- For the Brain4Cars dataset [71] thousands of miles were driven around San Francisco in both highway and urban environments. 274 lane change scenes and 131 turn scenes were captured. The provided data consist of driving intention labels and videos of the driver and the driving scene.
- The highD dataset [72] was collected on a German highway around Cologne by drone. Over 5600 highway lane changes were captured. The provided data consists of 16.5 hours of bird's-eye-view highway video footage, vehicle trajectories, time headway, and lane changes made by a vehicle (when the lane markings are crossed).
- The Honda Research Institute Driving Dataset (HDD) [73] consist of 104 hours of lane branch, lane change, merge, park, passing, and turn maneuvers in the San Francisco Bay Area. The provided data include exterior camera and lidar footage, GPS, vehicle dynamics, and behavioral labels.
- For the collection of the PREVENTION dataset [74], a test car was equipped with a camera, lidar, and radar sensors, and drove for six hours on a highway. In total 2000 left lane changes, 2000 right lane changes, and 1309 lane keeping sequences were captured. The provided data consist of exterior video footage, radar and lidar footage, road user detections, lane changes, and vehicle trajectories.

LIMITATIONS. Simulations of driving environments can be useful for testing novel model architectures or to create examples of rare scenarios, but the question remains how representative the results are if the involved participants execute driving maneuvers on a screen in a lab setting. As mentioned by Xing *et al.* [10], there is a potential bias if the driver has to execute a maneuver on request. In such a scenario, the interaction with other road users might not be realistic which limits the generalization of the results. Lastly, driving experiments are expensive and the captured data are potentially subject to privacy infringements. Thus, it is a positive trend that more datasets have become publicly available in the past five years.

C. SAMPLING, METHODS, AND PERFORMANCE

The included DIR studies mainly consider two types of maneuvers: lane changes and turns. The majority of the highway lane change intention studies aim to infer whether a driver intends to perform one of the following maneuvers: a left lane change, right lane change, or whether the driver continues to drive straight and to stay in the same lane (also referred to as lane keeping). For the intersection turn intention studies the driver intention classification task consists of the following three maneuvers: turning left, turning right, or driving straight ahead at a crossroad. To enable the recognition of driver intentions for a maneuver, a dataset must have enough instances per maneuver and must be balanced across the maneuver types. The study of Philips *et al.* [75] used a dataset where 8355 instances (21%) were left turn maneuvers, 2106 instances (5%) right turn maneuvers, and 29086 instances (71%) were examples of driving straight.

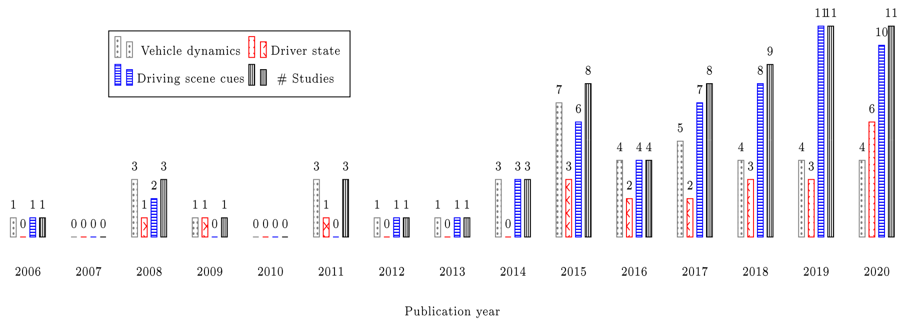

FIGURE 10. Overview of the feature types over time. The three feature categories indicate which dimensions are considered to recognize the intentions of a driver.

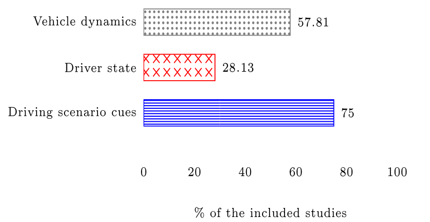

FIGURE 11. The feature types as a percentage of the total number of included studies. Note, studies can use features from multiple categories.

The RNN implementation recognized the driver's intention to perform a right turn maneuver poorly (precision: 3.9%), in contrast to recognizing the intention to drive straight (precision: 96.80%), or the intention to perform a left turn (precision: 68.20%). This example highlights that having thousands of instances per maneuver type does not directly lead to a high performance if the dataset is imbalanced.

To understand the reliability of a DIR experiment, the studies are divided into three groups based on the number of instances used to conduct the experiment: 0–99 instances, 100–999 instances, or a 1000 or more instances. Figure 12 visualizes the number of instances used over time, and Figure 13 shows a total overview of the used number of instances. Note, studies that consider lane change maneuvers and turn maneuvers are included separately per maneuver type. A dataset with three types of maneuvers (e.g., left lane change, right lane change, or lane keeping) would be perfectly balanced if every maneuver type would make up for 33.33% of the instances in the dataset. To understand the balance distribution of the included studies, we analyzed the papers that report the number of instances per maneuver type. If the number of instances of a maneuver type is between 30% and 40% of the total number of instances in the dataset, the class is considered to be balanced for that study. In eleven lane change intention studies the lane keeping and right lane change maneuver instances are imbalanced in 45% of the studies. The left lane change instances are imbalanced in 55% of the studies. For nine of the turn maneuver intention studies, the turn left and turn right classes are imbalanced 50% of the time. The driving straight instances are imbalanced in 63% of the studies.

The usage of deep learning methods to recognize driver intentions rose from 2016 onward. Since 2016, 78% of the turn maneuver studies and 54% of the lane change maneuver studies applied a deep learning method to infer the driver's intentions. Figure 14 shows the usages of different methods over time. Note, if a study compares multiple methods, every method is included separately. For example, Tang *et al.* [76] compared the result of an HMM, SVM, and DBN, and Driggs-Campbell and Bajcsy [77] evaluated an SVM, RF, and Logistic Regression. Therefore, the total number of applied methods exceeds the total number of included studies.

Figure 15 shows the performance bandwidth of studies that disclose the precision per maneuver type and were published between 2016–2020. The precision is based on the time step closest to the execution of the intended driving maneuver. Figure 15 illustrates that the performance of the applied methods is skewed towards 100%. The skewed performance hints towards, regardless the method, a relatively high performance for recognizing the driver's intention to perform a right lane change maneuver compared to a left lane change maneuver.

For the ADAS to infer the driver's intention, it is essential to complete the inference task before the driver starts to act. However, not every study reports the performance of the method at multiple time steps and some studies do not use seconds to indicate the time step. For example, Tran and Firl (2012) [78] reported the number of frames without mentioning the frequency, and Tang *et al.* (2015) [76] reported the relative distance in meters to an intersection. In recent studies the performance varies and the results are presented with different metrics. Li *et al.* (2019) [67] performed an experiment with 47 participants, acquired 2759 driving scenarios from which vehicle dynamics and driving scene features were constructed. A neural network architecture that combined a convolutional neural network and a RNN (LSTM) yielded an accuracy of 77.2%. Xu *et al.* (2019) [79] used the open-source dataset from [73] and proposed a temporal recurrent neural network architecture. The precision for the left and right turn intentions were respectively 77.0% and 76.6%, while the lane change precision was 45.9% for the left lane change and 43.6% for the right lane change maneuver. Girma *et al.* (2020) [80] used 2970 data points of

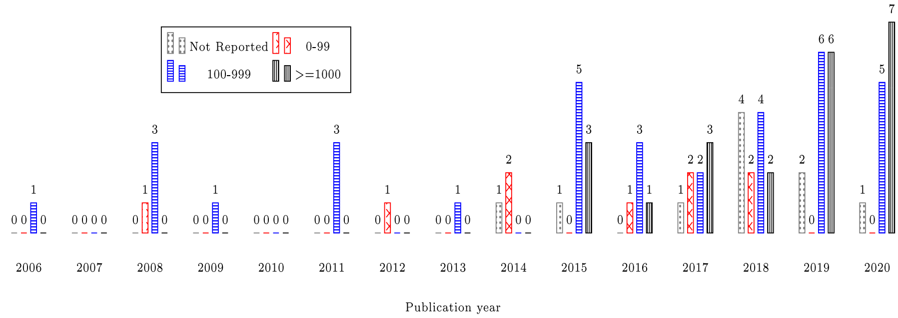

FIGURE 12. Overview of the number of reported instances used for inferring driver intentions over time.

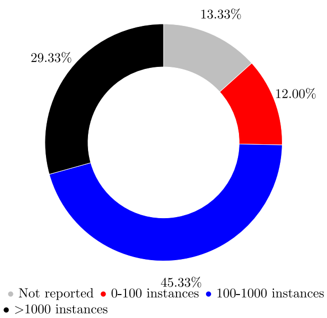

FIGURE 13. Total overview of the number of reported instances used in the DIR studies.

which 990 were labeled as driving straight, 770 as stopping, 660 turning right, and 550 turning left at an intersection. Only the velocity and yaw rate were used as categorized sequence features as input to a bidirectional LSTM with an attention mechanism and yielded an accuracy of 99.65%. Xing *et al.* (2020) [34] conducted a study with three participants and constructed a dataset consisting of 65 left lane change maneuvers, 70 right lane change maneuvers, and 66 lane-keeping instances. Vehicle dynamics, driver monitoring, and driving scene features served as input for an ensemble of three RNNs and they reported the precision for multiple time steps. The left lane change, right lane change, and lane-keeping precision were respectively, 95.6%, 95.6%, and 94.9% at half a second before the maneuver, 95.9%, 93.2%, and 94.7% at two seconds before the maneuver, and 93.8%, 88.9%, and 87.2% at three and a half seconds before the maneuver. At five seconds before the maneuver, the precision drops to 70% for all three maneuver types. Rong *et al.* (2020) [81] only use the video frames from the Brain4Cars dataset [71] as input to their convolutional based encoder-decoder architecture. The fusion of the interior and exterior achieved the best result. Similar to Xing *et al.* [34] the performance is reported per time step before the maneuver. The accuracy was 83.98% at zero seconds before the lane change or turn maneuver, 72.09% at two seconds before the maneuver and 59.13% at four seconds before the maneuver.

LIMITATIONS. The imbalance across the maneuver types disqualifies accuracy alone as a sufficient evaluation metric. The relatively low performance of left lane changes, regardless of the method, can motivate extra research efforts to explain the difference in detail (Figure 15). Although the results of some studies look promising, it is necessary to assess the performance on multiple datasets to understand whether the methods are capable to generalize. Driving scenarios and interactions can be highly contextual, which can lead to biased models.

DIR studies that explicitly state to study intentions and that do not report the performance of the method before the actual maneuver is executed are classifying maneuvers rather than recognizing what a driver aspires to do. Besides the different terminology usage, a unified approach to report performance over time is non-existing, but the performance leading up to a maneuver is essential for DIR research.

D. OPEN CHALLENGES

Figure 16 shows the high level classification of the open challenges recommended by the included studies published between 2016 and 2020. Nine studies do not indicate any directions for future work (e.g., [26], [67], [71], [82], [83], [84], [85], [86]), whereas other studies mention multiple challenges. For example, Benterki *et al.* [32] described the need for a larger dataset with more instances, different driving scenes, and new feature representations (e.g., a way to represent the interactions between vehicles). Ramanishka *et al.* [73] stated that adding data could improve the performance, but a better representation is required to understand the relationship between behaviors in different layers. Only two of the included studies considered a layered representation: Ramanishka *et al.* [73] decomposed driver behavior in a four-layer representation (goal, stimulus, cause, and attention) and Li *et al.* [67] analyzed human driving data on multiple levels (driving action recognition, attention, driver intentions and cause inference).

LIMITATIONS. The mentioned challenges are valid areas to explore, but they are not unique for DIR research. For most machine learning research more data can result in performance improvements or saturation. Figure 15 indicates that the reported precision of lane change intention studies is skewed to 100%. However, similar to what Ramanishka *et al.* [73] stated, applying different ML algorithms without changing the underlying modeling approach will not support the progress.

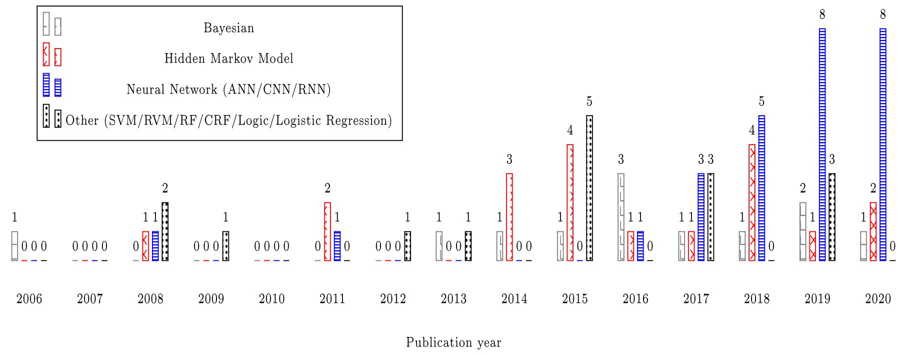

FIGURE 14. Overview of the methods applied for inferring driver intentions over time. Some of the included studies explore multiple methods.

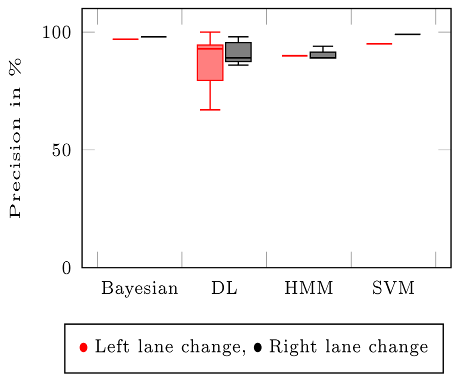

FIGURE 15. Overview of the reported precision by 15 lane change intention studies since 2016 per method. Only studies that reported the precision for both the left and right lane intention are included.

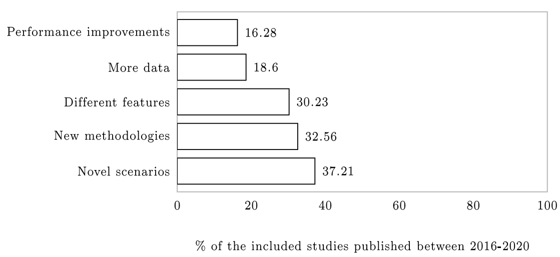

FIGURE 16. Overview of open challenges listed by studies published between 2016–2020 (N=43). Note, studies that mention multiple challenges, and 22% of the studies do not mention future work.

E. TECHNOLOGY READINESS LEVEL ESTIMATION

The TRLs are translated to criteria that enable a high level estimate of the current TRL of DIR technology [51]. TRL1 (basis principles observed) is the level where the initial research hypothesis is formulated. The environment where the technology should operate is still hypothetical and initial scientific observations are reported in journals or proceedings. TRL2 (concept formulation) marks the stage where experiments are performed on synthetic data, a desktop environment is used, and that the individual parts of the technology work without an actual integration attempt. For TRL3 (the experimental proof-of-concept phase), experiments have to be conducted on small representative datasets and the algorithm should be implemented on a surrogate processor in a lab environment. TRL4 (broad laboratory validation) requires formal systems architecture development, system requirements to be known, and experiments with full-scale representative datasets. One of the criteria for TRL5 (validation in a relevant environment) is that the algorithms run on a processor with similar characteristics as the future target environment.

The included studies have not yet demonstrated their DIR methods in a computational environment similar to a car. Several studies reported the hardware used to conduct the analysis (e.g., [27], [34], [79], [86], [87], [88], [89], [90], [91]). For example, Zhou *et al.* [86] used the NVIDIA Titan X GPU, and Xu *et al.* [79] used the NVIDIA Quadro P6000 GPU for processing the data. For future cars, most car manufacturers announced to integrate a GPU to enable an ADAS or ADS to process multiple (visual) data streams in parallel [92]. Embedding a DIR module in the ADAS that can infer the driver's intention of several different maneuvers in real-time multiple time steps ahead is required for the field to advance to the next TRL (TRL5 and above).

Besides from not testing algorithms in a computational environment similar to a car, there is also a lack of understanding on how the intention recognition methods will support the driver. For example, Han *et al.* [26], who used an open-source static highway dataset, claim that the developed system can support the decision making process of the car but it is unclear what decisions are supported. Leonhardt *et al.* [24] combined driving scene cues and driver state features to detect the driver's intention to perform a lane change maneuver and suggest that these functions are the basis for an 'adaptive driver assistance system' without clarifying what that system would do.

In summary, the TRL of DIR technology is estimated to be between TRL3 and TRL4. To progress to TRL5 and above, the actual application of the technology must be clear and the algorithms should be tested in an environment with similar computational power as a car, and comply with the regulatory requirements for ADAS and AI based functionalities.

V. DISCUSSION

This section highlights challenges that need to be tackled to advance the DIR research field. Section V-A covers the lack of performance evaluation comparisons between methods and highlights why estimating a minimal required time horizon is a complex task, Section V-B discusses the lack of uncertainty representations, and Section V-C considers a hierarchical modelling approach.

A. EVALUATION

It is unlikely that future DIR studies will always share their datasets. Thus, a true comparison between studies and the applied methods is impossible. Therefore, constructing a benchmark dataset from existing open-source datasets enables the possibility to understand the effectiveness of different DIR methods better. A drawback of using a benchmark dataset is that not all desired sensor data are available. After the construction of a benchmark dataset, the performance comparison approach also requires attention. A common performance comparison approach is to construct an overview table and highlight which method yields the best precision and recall (e.g., [88]). To establish SOTA performance for a method for a driving scenario, statistical significance testing is required. For example, after computing the evaluation metrics, the approach described by Demšar [93] can be followed to establish significant performance differences between methods.

Following the integrated safety model of Jiménez *et al.* [94], a DIR system aims to assist or warn a driver to avoid potential traffic conflicts. The system must recognize a driving intention well in advance before the first action. This means that a driver should get enough time to process and act on the notification to alter the plan of actions if necessary.

Green (2000) [95] argued that a standard or general perception reaction time does not exist, and Muttart (2003) [96] showed that the use of a mean response time is inappropriate. Thus, the estimation of how far ahead a DIR method should recognize an intention remains a challenging task. The required minimal time horizon depends on multiple factors, ranging from the time the system needs to recognize the driving intentions, the time the driver needs to change the physical position, to the time the car needs to respond to the driver's action. The driver's perception to register the scenario is also subject to multiple factors [97]. For example, the mean reaction time increases with age [98], [99], [100], [101], results are indecisive on whether there is a difference between sexes with regard to reaction time ([99], [101], [102] did not observe any difference, where [103], [104] did find a difference). Multiple studies found evidence that distractions such as phone usage influence the reaction time while driving [105], [106], [107], also the response time of a driver to a detection task increases when the driving context becomes more complex (e.g., a rural area without other vehicles compared to a busy intersection with pedestrians, cyclists, and multiple traffic signs) [108], [109]. While the list of examples is non-exhaustive, it illustrates the complexity and dynamic nature of estimating the required time horizon for a DIR system.

B. UNCERTAINTY REPRESENTATION

Guo *et al.* (2017) [110] and Wilson and Izmailov (2020) [111] highlighted that modern deep learning approaches are often not accurately calibrated. This means that the prediction confidence of the classifier is not aligned with the misclassification rate to express the uncertainty of the model. To assess how well a model is calibrated, a comparison between the predictions and actual outcomes can be made. For example, the number of times a model predicts driving intentions for a maneuver and the number of times that maneuver is actually performed by the driver. However, in this specific case, the abandoned intentions are not considered.

The two most common uncertainty types are aleatoric and epistemic uncertainty. On an abstract level, aleatoric uncertainty covers the natural variation in a considered environment, whereas epistemic uncertainty describes the lack of knowledge or information during the modeling phase [112]. In a machine learning context, this means that the aleatoric uncertainty deals with the randomness of the input data and that the epistemic uncertainty (also referred to as model or parameter uncertainty) expresses to what extend the model is certain about the produced output [113]. In a driving context, a modeling approach does not account for all unique scenarios. Given the safety-critical environment in which a car and a driver operate, a method should be capable to express when it is uncertain about the produced result [114]. If a method is capable to express the epistemic uncertainty, it helps to understand the limitations and gaps of the modeling approach.

C. HIERARCHICAL MODELING

Driving is a complex behavior where a driver can possibly entertain multiple intentions at the same time. It could be beneficial to introduce a hierarchy of driving intentions [67], [73]. Consider the following example, a car is stuck behind a slower truck on a two-way highway. As long as it is impossible to change lanes due to the upcoming traffic on the left lane, the driver might still have the intention to overtake the truck, however for the time being the driver intents to keep the lane until overtaking becomes possible. This scenario covers an intention without an immediate priority to act. A driver can be stuck behind the truck for half a minute, while nothing changes about the intention and the commitment to overtake the truck. In sum, it might be worthwhile to investigate if short, medium, and long-term driving intentions improve performance.

VI. CONCLUSION

A significant amount of research has been performed on single agent DIR maneuver classification. Recently, more open-source DIR datasets have become publicly available [70], [71], [72], [73], [74]. The number of instances to train a DIR method have increased and deep learning algorithms have become the preferred approach. The imbalance of the number of reported instances of a maneuver type in a dataset and the selected evaluation metrics require attention to avoid misleading conclusions. Several challenges must be addressed for DIR technology to proceed to TRL 4–5. The intention recognition methods should be implemented in a computing environment similar to a car. A DIR system needs a clear goal of how it can support a driver and in which driving scenarios it can offer support. Complementary evaluation metrics must be applied to review if the approach is capable to timely recognize the driving intentions and significance testing should be executed to establish performance differences between methods.

REFERENCES

[1] "Global Status Report on Road Safety 2018." WHO. 2018. [Online]. Available: https://apps.who.int/iris/bitstream/handle/10665/277370/WHO-NMH-NVI-18.20-eng.pdf?ua=1

[2] S. L. James *et al.*, "Morbidity and mortality from road injuries: Results from the global burden of disease study 2017," *Injury Prevent.*, vol. 26, no. S2, pp. i46–i56, 2020.

[3] "European Regional Status Report on Road Safety 2019." WHO. 2019. [Online]. Available: https://www.euro.who.int/en/publications/abstracts/european-regional-status-report-on-road-safety-2019

[4] A. Jarašūniene and G. Jakubauskas, "Improvement of road safety using passive and active intelligent vehicle safety systems," *Transport*, vol. 22, no. 4, pp. 284–289, 2007.

[5] K. Hojjati-Emami, B. Dhillon, and K. Jenab, "Reliability prediction for the vehicles equipped with advanced driver assistance systems (ADAS) and passive safety systems (PSS)," *Int. J. Ind. Eng. Comput.*, vol. 3, no. 5, pp. 731–742, 2012.

[6] A. H. Ahangarnejad, A. Radmehr, and M. Ahmadian, "A review of vehicle active safety control methods: From antilock brakes to semiautonomy," *J. Vib. Control*, vol. 27, nos. 15–16, pp. 1683–1712, 2021.

[7] K. Xie, K. Ozbay, H. Yang, and C. Li, "Mining automatically extracted vehicle trajectory data for proactive safety analytics," *Transp. Res. C, Emerg. Technol.*, vol. 106, pp. 61–72, Sep. 2019.

[8] A. Doshi and M. M. Trivedi, "Tactical driver behavior prediction and intent inference: A review," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, 2011, pp. 1892–1897.

[9] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH J.*, vol. 1, no. 1, pp. 1–14, 2014.

[10] Y. Xing *et al.*, "Driver lane change intention inference for intelligent vehicles: Framework, survey, and challenges," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 4377–4390, May 2019.

[11] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.

[12] L. M. Pereira and T. A. Han, "State-of-the-art of intention recognition and its use in decision making," *AI Commun.*, vol. 26, no. 2, pp. 237–246, 2013.

[13] F. Sadri, "Logic-based approaches to intention recognition," in *Handbook of Research on Ambient Intelligence and Smart Environments: Trends and Perspectives*. Hershey, PA, USA: IGI Global, 2011, pp. 346–375.

[14] G. Johansson, "Visual perception of biological motion and a model for its analysis," *Percept. Psychophys.*, vol. 14, no. 2, pp. 201–211, 1973.

[15] S.-J. Blakemore and J. Decety, "From the perception of action to the understanding of intention," *Nat. Rev. Neurosci.*, vol. 2, no. 8, pp. 561–567, 2001.

[16] V. Gallese, L. Fadiga, L. Fogassi, and G. Rizzolatti, "Action recognition in the premotor cortex," *Brain*, vol. 119, no. 2, pp. 593–609, 1996.

[17] V. Gallese and A. Goldman, "Mirror neurons and the simulation theory of mind-reading," *Trends Cogn. Sci.*, vol. 2, no. 12, pp. 493–501, 1998.

[18] F. Heider and M. Simmel, "An experimental study of apparent behavior," *Amer. J. Psychol.*, vol. 57, no. 2, pp. 243–259, 1944.

[19] H. M. Wellman, *The Child's Theory of Mind*. Cambridge, MA, USA: MIT Press, 1992.

[20] J. W. Astington, *The Child's Discovery of the Mind*, vol. 31. Cambridge, MA, USA: Harvard Univ. Press, 1993.

[21] J. Perner, *Understanding the Representational Mind*. Cambridge, MA, USA: MIT Press, 1991.

[22] B. F. Malle and J. Knobe, "The folk concept of intentionality," *J. Exp. Soc. Psychol.*, vol. 33, no. 2, pp. 101–121, 1997.

[23] L. Fridman, "Human-centered autonomous vehicle systems: Principles of effective shared autonomy," 2018, *arXiv:1810.01835*.

[24] V. Leonhardt, T. Pech, and G. Wanielik, "Data fusion and assessment for maneuver prediction including driving situation and driver behavior," in *Proc. 19th Int. Conf. Inf. Fusion (FUSION)*, 2016, pp. 1702–1708.

[25] J. Liu, Y. Luo, H. Xiong, T. Wang, H. Huang, and Z. Zhong, "An integrated approach to probabilistic vehicle trajectory prediction via driver characteristic and intention estimation," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, 2019, pp. 3526–3532.

[26] T. Han, J. Jing, and Ü. Özgüner, "Driving intention recognition and lane change prediction on the highway," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2019, pp. 957–962.

[27] D. Tran, W. Sheng, L. Liu, and M. Liu, "A hidden Markov model based driver intention prediction system," in *Proc. IEEE Int. Conf. Cyber Technol. Autom. Control Intell. Syst. (CYBER)*, 2015, pp. 115–120.

[28] D. Ramage, "Hidden Markov models fundamentals," CS229 Section Notes, vol. 1, 2007.

[29] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, vol. 3, no. 1, pp. 4–16, Jan. 1986.

[30] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[31] L. E. Baum and L. Welch, "Statistical estimation procedure for probabilistic functions of finite Markov processes," *Proc. Nat. Acad. Sci. USA*, to be published.

[32] A. Benterki, M. Boukhnifer, V. Judalet, and C. Maaoui, "Artificial intelligence for vehicle behavior anticipation: Hybrid approach based on maneuver classification and trajectory prediction," *IEEE Access*, vol. 8, pp. 56992–57002, 2020.

[33] Z. Hao, X. Huang, K. Wang, M. Cui, and Y. Tian, "Attention - based GRU for driver intention recognition and vehicle trajectory prediction," in *Proc. 4th CAA Int. Conf. Veh. Control Intell. (CVCI)*, 2020, pp. 86–91.

[34] Y. Xing, C. Lv, H. Wang, D. Cao, and E. Velenis, "An ensemble deep learning approach for driver lane change intention inference," *Transp. Res. C, Emerg. Technol.*, vol. 115, Jun. 2020, Art. no. 102615.

[35] A. Géron, *Neural Networks and Deep Learning*, O'Reilly, 2018.

[36] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D, Nonlinear Phenom.*, vol. 404, Mar. 2020, Art. no. 132306.

[37] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[38] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016, http://www.deeplearningbook.org

[39] A. Amidi and S. Amidi. "Deep learning cheatsheet." 2018. Accessed: Nov. 21, 2021. [Online]. Available: https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-deep-learning

[40] A. Salaün, Y. Petetin, and F. Desbouvries, "Comparing the modeling powers of RNN and HMM," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, 2019, pp. 1496–1499.

[41] N. V. Chawla, N. Japkowicz, and A. Kotcz, "Special issue on learning from imbalanced data sets," *ACM SIGKDD Explorations Newslett.*, vol. 6, no. 1, pp. 1–6, 2004.

[42] M. Zhao, D. Käthner, and M. Jipp, "Modeling driver intention and behavior at roundabouts," in *Proc. 4th Interdisziplinärer Workshop Kognitive Systeme*, 2015, pp. 10–35.

[43] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline oversampling for imbalanced data classification," *Int. J. Knowl. Eng. Soft Data Paradigms*, vol. 3, no. 1, pp. 4–21, 2011.

[44] A. Zheng, *Evaluating Machine Learning Models: A Beginner's Guide to Key Concepts and Pitfalls*. Sebastopol, CA, USA: O'Reilly Media, 2015.

[45] J. A. Swets, *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. New York, NY, USA: Psychol. Press, 2014.

[46] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006.
[47] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, no. 3, pp. 839–843, 1983.
[48] D. M. Powers, "The problem of area under the curve," in *Proc. IEEE Int. Conf. Inf. Sci. Technol.*, 2012, pp. 567–573.
[49] T. C. Landgrebe and R. P. Duin, "Approximating the multiclass ROC by pairwise analysis," *Pattern Recognit. Lett.*, vol. 28, no. 13, pp. 1747–1758, 2007.
[50] P. Flach, "The many faces of ROC analysis in machine learning," in *Proc. ICML Tutorial*, 2004, pp. 1–49.
[51] J. C. Mankins, "Technology readiness levels," Washington, DC, USA, NASA, White Paper, Apr. 1995.
[52] J. Banke. "Technology Readiness Levels Demystified." Aug. 2010. [Online]. Available: https://www.nasa.gov/topics/aeronautics/features/trl_demystified.html
[53] *Technology Readiness Assessment (TRA) Guidance*, United States Dept. Defense, Washington, DC, USA, 2011.
[54] D. McGarvey, J. Olson, S. Savitz, G. Diaz, and G. Thompson, *Department of Homeland Security Science and Technology Readiness Level Calculator (ver. 1.1)*, Arlington, VA, USA: Homeland Security Stud. Anal. Inst., 2009.
[55] D. N. Veritas, "Qualification of new technology," DNV, Høvik, Norway, Rep. DNV-RP-A203, 2011.
[56] A. Olechowski, S. D. Eppinger, and N. Joglekar, "Technology readiness levels at 40: A study of state-of-the-art use, challenges, and opportunities," in *Proc. Portland Int. Conf. Manag. Eng. Technol. (PICMET)*, 2015, pp. 2084–2094.
[57] G. A. Papafotiou, G. D. Demetriades, and V. G. Agelidis, "Technology readiness assessment of model predictive control in medium- and high-voltage power electronics," *IEEE Trans. Ind. Electron.*, vol. 63, no. 9, pp. 5807–5815, Sep. 2016.
[58] J. D. Smith, "An alternative to technology readiness levels for non-developmental item (NDI) software," in *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci.*, 2005, p. 315a.
[59] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *PLoS Med.*, vol. 6, no. 7, 2009, Art. no. e1000097.
[60] M. J. Grant and A. Booth, "A typology of reviews: An analysis of 14 review types and associated methodologies," *Health Inf. Libraries J.*, vol. 26, no. 2, pp. 91–108, 2009.
[61] H. Snyder, "Literature review as a research methodology: An overview and guidelines," *J. Bus. Res.*, vol. 104, pp. 333–339, Nov. 2019.
[62] E. D. López-Cózar, E. Orduña-Malea, and A. Martín-Martín, "Google scholar as a data source for research assessment," in *Springer Handbook of Science and Technology Indicators*. Cham, Switzerland: Springer, 2019, pp. 95–127.
[63] A. Martín-Martín, M. Thelwall, E. Orduna-Malea, and E. D. López-Cózar, "Google scholar, microsoft academic, Scopus, dimensions, Web of science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations," *Scientometrics*, vol. 126, no. 1, pp. 871–906, 2021.
[64] J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quantitative Sci. Stud.*, vol. 1, no. 1, pp. 377–386, 2020.
[65] K. Krippendorff, *Content Analysis: An Introduction to its Methodology*. Los Angeles, CA, USA: Sage Publ., 2018.
[66] A. Gerdes, "Driving Manoeuvre recognition," in *Proc. 13th World Congr. Exhibit. Intell. Transport Syst. Serv.*, 2006, pp. 1–8.
[67] M. G. Li *et al.*, "DBUS: Human driving behavior understanding system," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2019, pp. 2436–2444.
[68] F. Lethaus, M. R. Baumann, F. Köster, and K. Lemmer, "Using pattern recognition to predict driver intent," in *Proc. Int. Conf. Adapt. Nat. Comput. Algorithms*, 2011, pp. 140–149.
[69] I.-H. Kim, J.-H. Bong, J. Park, and S. Park, "Prediction of driver's intention of lane change by augmenting sensor information using machine learning techniques," *Sensors*, vol. 17, no. 6, p. 1350, 2017.
[70] "Next Generation Simulation (NGSIM)." FHWA. 2007. [Online]. Available: https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm
[71] A. Jain, H. S. Koppula, S. Soh, B. Raghavan, A. Singh, and A. Saxena, "Brain4cars: Car that knows before you do via sensory-fusion deep learning architecture," 2016, *arXiv:1601.00740*.
[72] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The HighD dataset: A drone dataset of naturalistic vehicle trajectories on German highways for validation of highly automated driving systems," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, 2018, pp. 2118–2125.
[73] V. Ramanishka, Y.-T. Chen, T. Misu, and K. Saenko, "Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7699–7707.
[74] R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "Experimental validation of lane-change intention prediction methodologies based on CNN and LSTM," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, 2019, pp. 3657–3662.
[75] D. J. Phillips, T. A. Wheeler, and M. J. Kochenderfer, "Generalizable intention prediction of human drivers at intersections," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2017, pp. 1665–1670.
[76] B. Tang, S. Khokhar, and R. Gupta, "Turn prediction at generalized intersections," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2015, pp. 1399–1404.
[77] K. Driggs-Campbell and R. Bajcsy, "Identifying modes of intent from driver behaviors in dynamic environments," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, 2015, pp. 739–744.
[78] Q. Tran and J. Firl, "A probabilistic discriminative approach for situation recognition in traffic scenarios," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 147–152.
[79] M. Xu, M. Gao, Y.-T. Chen, L. S. Davis, and D. J. Crandall, "Temporal recurrent networks for online action detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 5532–5541.
[80] A. Girma, S. Amsalu, A. Workineh, M. Khan, and A. Homaifar, "Deep learning with attention mechanism for predicting driver intention at intersection," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2020, pp. 1183–1188.
[81] Y. Rong, Z. Akata, and E. Kasneci, "Driver intention anticipation based on in-cabin and driving scene monitoring," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, 2020, pp. 1–8.
[82] P. Gebert, A. Roitberg, M. Haurilet, and R. Stiefelhagen, "End-to-end prediction of driver intention using 3D convolutional neural networks," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2019, pp. 969–974.
[83] S. Wang, Q. Yu, and X. Zhao, "Study on driver's turning intention recognition hybrid model of GHMM and GGAP-RBF neural network," *Adv. Mech. Eng.*, vol. 10, no. 3, 2018, Art. no. 1687814018764985.
[84] F. Bocklisch, S. F. Bocklisch, M. Beggiato, and J. F. Krems, "Adaptive fuzzy pattern classification for the online detection of driver lane change intention," *Neurocomputing*, vol. 262, pp. 148–158, Nov. 2017.
[85] D. S. González, V. Romero-Cano, J. S. Dibangoye, and C. Laugier, "Interaction-aware driver maneuver inference in highways using realistic driver models," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, 2017, pp. 1–8.
[86] D. Zhou, H. Ma, and Y. Dong, "Driving maneuvers prediction based on cognition-driven and data-driven method," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, 2018, pp. 1–4.
[87] A. Benterki, M. Boukhnifer, V. Judalet, and M. Choubeila, "Prediction of surrounding vehicles lane change intention using machine learning," in *Proc. 10th IEEE Int. Conf. Intell. Data Acquisition Adv. Comput. Syst. Technol. Appl. (IDAACS)*, vol. 2, 2019, pp. 839–843.
[88] B. Rekabdar and C. Mousas, "Dilated convolutional neural network for predicting driver's activity," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, 2018, pp. 3245–3250.
[89] J. Li, X. Li, B. Jiang, and Q. Zhu, "A maneuver-prediction method based on dynamic Bayesian network in highway scenarios," in *Proc. Chin. Control Decis. Conf. (CCDC)*, 2018, pp. 3392–3397.
[90] Y. Xing, Z. Hu, Z. Huang, C. Lv, D. Cao, and E. Velenis, "Multi-scale driver behaviors reasoning system for intelligent vehicles based on a joint deep learning framework," in *Proc. IEEE Int. Conf. Syst. Man Cybern. (SMC)*, 2020, pp. 4410–4415.

[91] D. Richardos, B. Anastasia, D. Georgios, and A. Angelos, "Vehicle maneuver-based long-term trajectory prediction at intersection crossings," in *Proc. IEEE 3rd Connected Autom. Veh. Symp. (CAVS)*, 2020, pp. 1–6.

[92] "NVIDIA Self Driving Cars—Partners." NVIDIA. 2021. [Online]. Available: https://www.nvidia.com/en-us/self-driving-cars/partners/

[93] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Dec. 2006.

[94] F. Jiménez, J. E. Naranjo, and F. García, "An improved method to calculate the time-to-collision of two vehicles," *Int. J. Intell. Transp. Syst. Res.*, vol. 11, no. 1, pp. 34–42, 2013.

[95] M. Green, "'How long does it take to stop?' methodological analysis of driver perception-brake times," *Transp. Human Factors*, vol. 2, no. 3, pp. 195–216, 2000.

[96] J. W. Muttart, "Development and evaluation of driver response time predictors based upon meta analysis," *SAE Trans.*, vol. 112, pp. 876–896, Jan. 2003.

[97] P. Droździel, S. Tarkowski, I. Rybicka, and R. Wrona, "Drivers 'reaction time research in the conditions in the real traffic," *Open Eng.*, vol. 10, no. 1, pp. 35–47, 2020.

[98] B. U. Forstmann, M. Tittgemeyer, E.-J. Wagenmakers, J. Derrfuss, D. Imperati, and S. Brown, "The speed-accuracy tradeoff in the elderly brain: A structural model-based approach," *J. Neurosci.*, vol. 31, no. 47, pp. 17242–17249, 2011.

[99] G. Der and I. J. Deary, "Age and sex differences in reaction time in adulthood: Results from the United Kingdom health and lifestyle survey," *Psychol. Aging*, vol. 21, no. 1, pp. 62–73, 2006.

[100] D. Dykiert, G. Der, J. M. Starr, and I. J. Deary, "Age differences in intra-individual variability in simple and choice reaction time: Systematic review and meta-analysis," *PLoS One*, vol. 7, no. 10, 2012, Art. no. e45759.

[101] D. L. Woods, J. M. Wyma, E. W. Yund, T. J. Herron, and B. Reed, "Factors influencing the latency of simple reaction time," *Front. Human Neurosci.*, vol. 9, p. 131, Mar. 2015.

[102] R. Gottsdanker, "Age and simple reaction time," *J. Gerontol.*, vol. 37, no. 3, pp. 342–348, 1982.

[103] D. Dykiert, G. Der, J. M. Starr, and I. J. Deary, "Sex differences in reaction time mean and intraindividual variability across the life span," *Develop. Psychol.*, vol. 48, no. 5, pp. 1262–1276, 2012.

[104] K. J. Anstey, K. Dear, H. Christensen, and A. F. Jorm, "Biomarkers, health, lifestyle, and demographic variables as correlates of reaction time performance in early, middle, and late adulthood," *Quart. J. Exp. Psychol. Sect. A*, vol. 58, no. 1, pp. 5–21, 2005.

[105] W. J. Horrey and C. D. Wickens, "Examining the impact of cell phone conversations on driving using meta-analytic techniques," *Human Factors*, vol. 48, no. 1, pp. 196–205, 2006.

[106] J. L. Hendrick and J. R. Switzer, "Hands-free versus hand-held cell phone conversation on a braking response by young drivers," *Perceptual Motor Skills*, vol. 105, no. 2, pp. 514–522, 2007.

[107] P. Choudhary and N. R. Velaga, "Effects of phone use on driving performance: A comparative analysis of young and professional drivers," *Saf. Sci.*, vol. 111, pp. 179–187, Jan. 2019.

[108] G. Jahn, A. Oehme, J. F. Krems, and C. Gelau, "Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 8, no. 3, pp. 255–275, 2005.

[109] C. J. Patten, A. Kircher, J. Östlund, L. Nilsson, and O. Svenson, "Driver experience and cognitive workload in different traffic environments," *Accid. Anal. Prevent.*, vol. 38, no. 5, pp. 887–894, 2006.

[110] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1321–1330.

[111] A. G. Wilson and P. Izmailov, "Bayesian deep learning and a probabilistic perspective of generalization," in *Proc. Int. Conf. Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 4697–4708, 2020.

[112] W. L. Oberkampf, J. C. Helton, C. A. Joslyn, S. F. Wojtkiewicz, and S. Ferson, "Challenge problems: Uncertainty in system response given uncertain parameters," *Rel. Eng. Syst. Saf.*, vol. 85, nos. 1–3, pp. 11–19, 2004.

[113] A. Der Kiureghian and O. Ditlevsen, "Aleatory or epistemic? Does it matter?" *Struct. Saf.*, vol. 31, no. 2, pp. 105–112, 2009.

[114] R. McAllister *et al.*, "Concrete problems for autonomous vehicle safety: Advantages of Bayesian deep learning," in *Proc. Int. Joint Conf. Artif. Intell.*, 2017, pp. 4745–4753.