Use of Social Interaction and Intention to Improve Motion Prediction Within Automated Vehicle Framework: A Review

Djamel Eddine Benrachou, Sebastien Glaser, Mohammed Elhenawy, and Andry Rakotonirainy

Abstract— **Human errors contribute to 94%(±2.2%) of road crashes resulting in fatal/non-fatal causalities, vehicle damages and a predicament in the pathway to safer road systems. Automated Vehicles (AVs) have been a potential attempt in lowering the crash rate by replacing human drivers with an advanced computer-aided decision-making approach. However, AVs are yet to progress in handling the unprecedented situations involving interactions with other road users. This raises a need for a sophisticated and robust methodological framework to predict human driver interaction and intention. It is of prime importance to develop a constructive knowledge on the existing literature for a proficient forward leap in the field. To address this, we aim to conduct a comprehensive review on motion prediction methods in automated driving context with a special emphasis on model-based and data-driven approaches. Over a hundred studies related to the motion prediction for AVs have been extensively reviewed. This study recommends that the field requires more intricate classification of motion prediction methods, as the conventional three-level categorisation scheme should be upgraded to a profound and present-day context. Therefore, we attempt to provide a clear categorisation of existing motion prediction solutions by adopting four principal strategies: 1. Prediction methods, 2. Classes, 3. Algorithms and 4. Datasets. An all-inclusive summary of the reviewed studies with their respective pros and cons are also presented. Furthermore, we summarise the standard evaluation metrics applied for road users' intention estimation and trajectory prediction tasks. It is found that the recent studies are built upon multi-agent learning systems with interaction among multiple road users in the same road environment. These methods can provide reliable prediction performance in highly interactive situations over long periods of time. However, the limitation could be at the cost of higher computational complexity in comparison to conventional methods, which are simpler to design and computationally effective. It is also observed that the conventional methods can only operate over a narrow prediction horizon and seldom consider the interactions among the road users. This review contributes to knowledge in validation, addresses the discrepancies, to explicate the ambiguities and to streamline current research for a futuristic perspective beneficiary in motion prediction field.**

Manuscript received 31 December 2021; revised 9 June 2022; accepted 13 August 2022. Date of publication 28 September 2022; date of current version 5 December 2022. This work was supported by the Queensland University of Technology Australian Research Council (QUT ARC) Linkage and the Motor Accident Insurance Commission (MAIC) Queensland. The views expressed herein are those of the authors and are not necessarily those of the MAIC. The Associate Editor for this article was B. Fidan. *(Corresponding author: Djamel Eddine Benrachou.)*

The authors are with the Centre for Accident Research and Road Safety—Queensland (CARRS-Q), Queensland University of Technology (QUT), Kelvin Grove, QLD 4059, Australia (e-mail: djameleddine.benrachou@hdr.qut.edu.au; sebastien.glaser@qut.edu.au; mohammed.elhenawy@qut.edu.au; r.andry@qut.edu.au).

Digital Object Identifier 10.1109/TITS.2022.3207347

Index Terms— **Automated vehicle, social interaction, motion prediction, intention prediction, trajectory prediction.**

I. Introduction

ANNUALLY around 1.35 million deaths are occurred as a result of road crashes [1].The AAA[1] reported about 1,140 road deaths in 2018 [2]. The 2021 Road Crash Report [3] found a recent spike in deaths on the Australian roads since the last year. Likewise, in the United States, the NHTSA[2] investigated the critical reasons−the last events in the crash causal chain−for road crashes and reported about 94%(±2.2%) of sever road crashes can be assigned to the driver's errors [4]. Likewise, statistics from the European Commission have shown that human error is one of the main factors in road crashes. This distress is preventable and delineates to be a technological challenge for researchers, vehicle manufacturing companies and policy makers. New motorised vehicles provide advanced technology to assist human-drivers in avoiding critical situations. These technologies assisting human drivers in providing a safe human-machine interface driving conditions are known as Advanced Driver Assistance Systems (ADAS). Though ADAS is a renowned concept first bought into light in 1950s has been changing the automation industry rapidly. However, researchers have been exploring and testing its efficiency in warning the human driver to act accordingly and avoid a crash. Nonetheless, the accelerated technological progress characterised by innovations whose rapid application and diffusion cause a rapid change in our society during this crucial technology acceptance phase.

Highly automated solutions help tackle most road challenges adapting high end sensors being inexpensive but fairly efficient, the rise of powerful Machine Learning (ML) techniques, and innovations introduced in recent years by tech companies like Google and Tesla Motors. Automotive and tech companies have shown the Automated Driving System (ADS) practicability through well-engineered prototypes, when in 2018, Google−Waymo's test fleet logging successfully completed over five million miles driving in automated mode on public roads across the United States [5], [6]. The Society of Automotive Engineers (SAE) in assignment with the U.S. Department of Transportation (DoT) effectively classifies ADS into six levels (0-5) of vehicle automation [7] according

[1]Australian Automobile Association.

[2]National Highway Traffic Safety Administration.

to its capacity to assist the human driver. This paper will follow the DoT/ SAE convention focusing on full automated operation, which requires the accuracy needed for active control. In this context, Automated Vehicles (AVs) will refer to SAE level 3+ systems. AVs can be advantageous on a social and individual standards [8]. On a social viewpoint, the AVs can potentially contribute to: (a)-improving road safety by reducing human-related elements and errors affecting the driving efficiency [9], [10]; (b) lowering the traffic fatalities by up to 90% by mid-centuries [11]. However, it is unrealistic to mention the probity of the collision prevention rate can be highly significant due to sole replacement of human-driver by the machine. The individual standards can address: (a)-a stress free driving environment for the human driver with a machine intervention [12], [13], [14]; (b)-mobility access enhancement for elderly and physically impaired users; (c)-lowering the carbon footprint, energy consumption and optimising traffic flow [15].

This is evidence of a successful milestone towards automation. The California Department of Motor Vehicles (CaDMV) has made mandatory for many years the recording when testing AV solutions, and has made open access to the disengage and crash data. The CaDMV's data is yet another testament showing that the lack of interaction prediction of the AV with other traffic participants is a major contributor to the high number of collisions. Favarò *et al.* [16] found that AV were over-represented in the front position in rear-end collisions. The detailed CaDMV data shows that collision often results from human driver misunderstanding of the AV behaviour. Events such as "emergency braking when nothing happens in the environment" or "stopping at an intersection when the roads are clear" have been mentioned in the CaDMV's collision report – mentioning such collisions occur mostly at low speed, with no injuries reported [17]. These examples demonstrate limitations related to decision-making in automated navigation systems and also the lack for adaption of the human driver to override the AV during emergency. They are mainly the effects of lack of understanding of the complex road environment, and also the difficulty in being able to infer the intention of other road participants. Brown [18] conducted in-depth analysis of publicly available video footage of real AV driving trials (Waymo—Google and Tesla AVs). His work focuses on emphasising ambiguities and misunderstanding of a real shared space in an urban design that integrates vehicles and AVs. He pointed out that the driving task is not a mere mechanical action but a complex activity that should be integrated in the AV software ("brain").

During navigating in the real-world, AVs generate their future trajectory according to multiple interactions - observed (past) and predicted (future) interactions – among neighbouring objects (static and dynamic obstacles). This process helps AVs to take better decisions, prevent hazards and overall functionality. Thus, adding to the robust perception and control requirements for proper functioning of the AV brain. The capacity of the AV to generate safe and effective interactive path is also critical to precedent tasks. Towards these goals, AVs should be effective at predicting motions of other road users while being able to infer their subsequent intention.

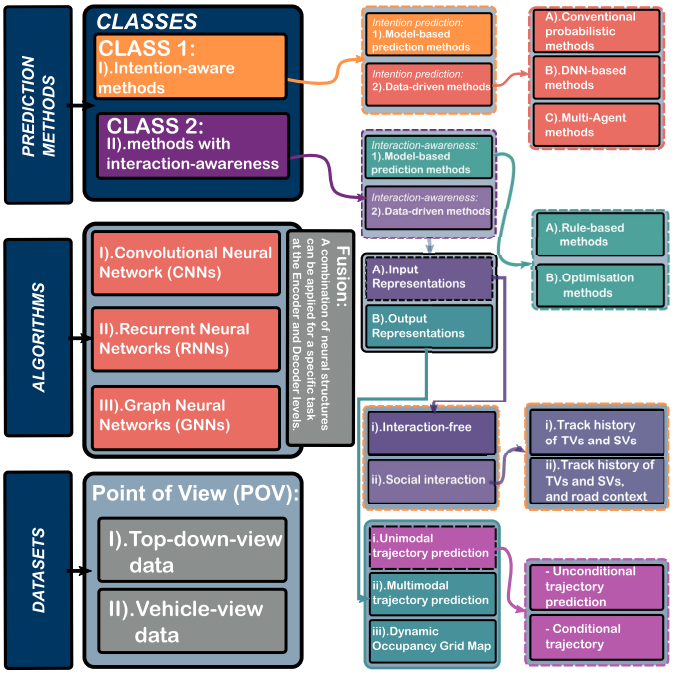

Fig. 1. Overview of the proposed categorisation scheme of state-of-the-art motion prediction approaches.

Motion prediction is a critical task to perform intelligent and robust decisions in simple and complex environments. This task refers to predicting future locations of objects/other road users over time in the driving scene by considering their past observations. The ability to make an effective prediction is at the core of robust and safer decisions, in particular to predict the position of all affected road users, regardless of the driving environment in which they operate.

This paper reviews recent works of motion prediction in an automated driving context and introduces a new classification scheme based on four main criteria: *prediction methods*, *classes*, *algorithms* and *datasets*. This paper also contributes in identifying the practical limitations of implementing recent solutions either based on data-driven methods or model-based methods. In summary, Fig. 1 depicts the proposed classification scheme which highlights the criteria in each of the four general groups and explained in the following sections.

The remainder of the paper is organised into eight sections: Section II defines basic terminologies and describes the generic problem. Section III presents challenges on motion prediction. Section IV provides an overview on related research and the implemented methodology. The automated vehicle motion prediction literature is reviewed in Section V. Datasets useful for developing and testing original motion prediction frameworks are presented in the section VI. Section VII describes metrics applied for assessment tasks. Current research gaps in the literature are given in the same section. The main concluding observations and potential new research insights and directions are given in Section VIII.

II. Problem Description

Predicting the trajectory of other road users in an automated driving scheme is a tedious task as the AV's software must

TABLE I
TERMINOLOGY: LOCATION INFORMATION AND VARYING TOPOLOGY PROBLEM

Term	Definition	Variable	Work
Position	Exact location of an observed object	(X, Y, Z, Ψ)	[19, 20]
Trajectory	Successive positions over time	Φ	
Manoeuvre	A set of trajectories achieving a common short-term goal, e.g., lane change (LC) can be performed by multiple trajectories with fast and slow lateral speeds	LK, LCR, LCL, etc	[21, 22, 23, 22]
Intention	Future manoeuvres performed by a vehicle		[24, 25, 26, 27]
Intention prediction	Predicting whether a vehicle will perform a specific manoeuvre		[25, 26, 27]
Social	Refers to a group of interacting individuals		[28, 19]
Interaction	A relationship between two or more systems, people, or groups that results in mutual or reciprocal influence		
Social interaction	Mutual behavioural/manoeuvring effects among vehicles and VRUs in a unit time and space on the road matrix		[29, 30]
Target Vehicle	A set of vehicle whose behaviour is to be predicted	TV	
Ego Vehicle	The automated vehicle which observes surrounding environment through gathered sensory data to predict the TV's behaviour	EV	
Surrounding Vehicle	Vehicles whose behaviour is explored by the prediction model. They are expected to affect the future behaviour of the TV	SV	

Legend: X = longitudinal positioning , Y = lateral positioning, Z = vertical positioning, Ψ = heading angles, LK= lane keeping, LCR= lane change right, LCL= lane change left, TV= target vehicle, EV= ego-vehicle, SV= surrounding vehicle.

show a spatio-temporal understanding of the world, while capturing the past states of observable surrounding road users, and their interaction patterns regardless of their number and types. It should also integrate different constraints related to the road scene and the stochastic nature of human behaviour. Motion prediction requires two key steps:(1)-track all associated information to neighbouring road users with the aim to get precise and reliable trajectories, while interpreting higher level behaviour, such as intention and interactions; and (2)-predict future motion of neighbouring road users based on sensed knowledge. To fulfil these tasks, the AV's software should have access to mapping data of the road scene and the area (or road context) where the AV or ego-vehicle (EV) is driving in (i.e., road and crosswalk locations, lane direction, and other relevant map-related information) including the surrounding road users (e.g., a surrounding vehicle (SV)) and targeted ones (e.g., a target vehicle (TV)).

The prediction design defines a reasoning about likely outcomes based on the history observations (or *the history states*) to predict *the future states* [31]: the joint state of M road users,[3] $\mathcal{D} = \{\mathbf{X}^m, \mathbf{Y}^m\}_{m=1}^{M}$ at time t as $\mathbf{X}_t \in \mathbb{R}^{M\times d} \doteq \{x_t^1, x_t^2, \ldots, x_t^M\}$, where X is the state vector and d is the size of each state,[4] and $x_t^n \in \mathbb{R}^d$ is the $n-$th road user at timestep t.

- *The history states* of the $n-$th road user $\mathbf{X}^m \doteq \{x_{t-\tau}^m, x_{t-\tau+1}^m, \ldots, x_t^m\}$ and $\mathbf{X} \doteq \mathbf{X}_{t-\tau:t}^{1:M}$ denoting the joint states in a temporal range; $t-\tau$ to t, where at any time instant t, τ is a fixed (past) time horizon. i
- *The future states* of all considered road users can be defined over a fixed (future) time horizon T, at any time t and denoted by $\mathbf{Y}^m \doteq \{y_t^m, y_{t+1}^m, \ldots, y_T^m\}$, $\mathbf{Y} \doteq \mathbf{Y}_{t:t+T}^{1:M}$ which is the joint state of all observations across few seconds in the future.

Table I provides useful terms and definitions used throughout this document. Table VIII describes the abbreviations/acronyms employed in this paper.

III. CHALLENGES

The ability of anticipating the behaviour of other road users occurs subconsciously and effortlessly for human drivers. This natural anthropomorphic skill has inspired many studies on automated driving [32], [33], [34], [35], and is a leading scope of scientific research aiming to achieve higher prediction performance, and thus promoting safer AV prototypes. This would be an eminent achievement for an AV to function with full human comprehension and decision-making capacity. Nevertheless, for an AV, its ability to predict the behaviour of other users is not merely about achieving the human performance level, but also to show the capability to consider other key parameters useful to achieve higher prediction accuracy as human driving is evidently imperfect. For instance, the AV should be able to model the spatio-temporal relationships (or interactions) among other road users by considering the stochastic and multimodal nature of their subsequent intention and movements in dynamic and heterogeneous driving conditions. These factors make the task of prediction multidimensional in space and time creating a complex multi-headed problem.

In the literature, the prediction of the behaviour of other road users (e.g., a vehicle) is often referred and is rather

[3]Vehicles and Vulnerable Road Users.
[4]The state of each road user is assumed to be fully observable coordinates on the ground plane.

misinterpreted as the vehicle manoeuvre, i.e., "behaviour" and "manoeuvre" are sometimes used interchangeably [36], [37], [38], which can easily lead to ambiguity.

In this paper, we consider on-road behaviour prediction [21], [23], [38], [39] as a general term which implies the intention of a road user or its trajectory or its motion as a whole, depending on the formulation of the problem. In the automated driving context, the on-road behaviour prediction has three salient limitations: (1)-finite or constrained observability of the road scene [40], [41], [42], [43]; (2)-asymmetric dependency and inter-vehicular dependency; (3)-uncertainties and multimodal outputs [39], [44].

A. Constrained Observability of the Road Scene

The AVs (SAE level3+) require reliable solutions to provide an accurate mapping of the surrounding traffic. This is only possible with multi-sensor perception systems relying on a combination of sensors [45]. These sensors are key instruments in detecting nearby obstacles, distance, speed and provide a high resolution 3D depiction of the world in real-time. Today, available ADS research platforms make use of a complex sensor fusion strategy for automated operations. This includes LiDAR, radar, cameras. Since no technology can meet the requirements in all scenarios, this will require multimodal sensing and their intelligent combination to achieve safer operation requirements. However, despite the perceived operating simplicity of these sensors, it is not always simple to reach precise and reliable representation of the world. This technology is quite sensitive to different changes, nuisance and external disturbances occurring in the real-world, such as noise, lighting variation, viewpoint change, non-rigid object changes, rotational object deformation (e.g., tracking nuisance of non-rigid objects, such as pedestrians and derived rotation and scaling problems [46]), ambient occlusion, and adversarial weather conditions. Although being assumed as a game-changer in the automated driving paradigm, sensors are highly sensitive to adverse external conditions, which can severely affect the AV's perception system reliability and thus the prediction performance. Despite the progress made and the solutions proposed in the state of the art contributing on processing LiDAR data and other sensory data under adverse weather [47], [48], [48] and denoising frameworks [40]. It is, however, important to point out that is an open challenge and an investigation matter. Limited sensor range, can also reduce the road scene observability of the AV (as shown in Fig. 2). Solutions tried to remedy this issue by using a top-to-bottom view of the world via fixed surveillance devices, such as CCTV cameras [49] or unmanned aerial vehicle (UAV) [50]. Sensors' reaction time, may cause a delay in information processing and thus will provide insufficient time to the AV to avoid imminent collisions.

B. Asymmetric Dependency [24] or Inter-Vehicular Dependence

When it comes to vehicles, is that inherent spatial relationship that coexists between vehicles whose sharing the same road space and which can mutually influence their subsequent

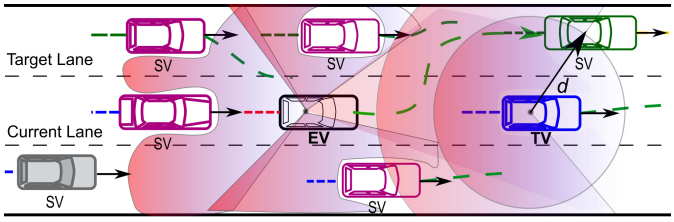

Fig. 2. Representation of restricted observability range of the on-board sensors in a three-lane freeway driving scenario with categorisation of different vehicles: the AV is set as the ego-vehicle (EV), surrounding vehicles (SVs) located within the range of visibility at a distance d are assumed affecting the current target vehicle (TV) behaviour. On the other hand, the grey-coloured SV (or Non Effective surrounding Vehicle [41]) is not covered by the sensor range, although it is potentially effective to the on-road behaviour of other vehicles.

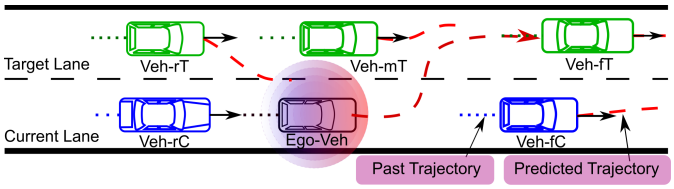

Fig. 3. A left lane change scenario [52]: The black vehicle (Ego-Veh) is the ego-vehicle, which will execute a left-lane change to the target lane. Ego-Veh needs to predict the future movements of Veh-mT and those of the Veh-fC, in order to achieve socially consistent on-road behaviour prediction. The prediction takes into consideration the leading vehicles (Veh-fT) and the tailing vehicle (Veh-rC), which is the targeted vehicle : the immediate left vehicle(Veh-mt), its leading vehicle (Veh-fT) and its following vehicle (Veh-rT).

behaviour on the road [18], [51]. In other terms, the behaviour of one vehicle can affect the behaviour of other vehicles around and vice versa.

Figure. 3 illustrates the inter-vehicular dependence in a motorway scenario and how vehicles can influence the behaviour of each other. In such situations, the AV's prediction framework should be mature enough to manage complex scenarios, such as ramp-merging and lane changes (as shown Fig. 3), which involve tactical negotiations, intention understanding of the behaviour of surrounding road users and their *interactions*. As vehicles' states are interdependent and do not evolve independently from one another, but rather they interact with each other [32]. In this paper, we refer to interactions among road users as *social interactions*.

C. Uncertainties and Multimodal Outputs

In highly interactive driving scenarios, the AV should perform a precise decision, while being aware of one's surroundings and different uncertainties. These uncertainties originate from the environment and/or noisy sensor data (as shown in Fig. 4(a)) and also from the fact that the intention of human drivers is inherently stochastic and cannot be directly measured [39]. This problem is generally formulated as a partially observable Markov decision process with latent variables [53] and multiple representations of the future intent, e.g., a vehicle may turn left or right in an intersection, as depicted in Fig. 4 (b).

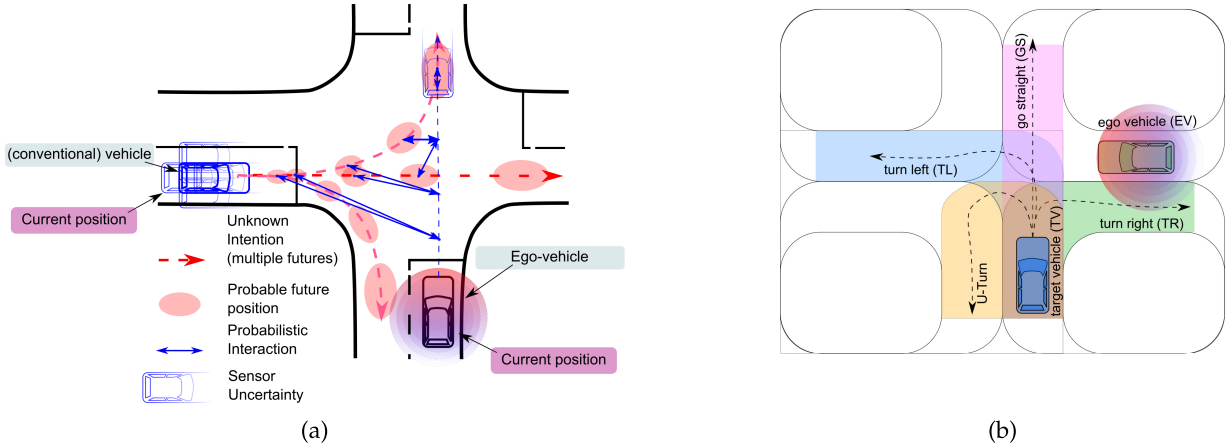

Fig. 4. Illustration of the multimodality and uncertainties in the urban road environment. (a) A typical situation in which the ego-vehicle has to make a decision for its subsequent manoeuvre under different uncertainties regarding the predicted motion of conventional vehicle [53]. (b) Multimodality in the prediction with multiple possible future intentions.

IV. Related Work and Methodology

A. Related Work

Motion prediction plays a key role in automated driving application. This task is a support for automated decision-making [28] and useful to assess risks associated with the planning process [54]. Motion prediction refers to predicting subsequent locations of one or multiple objects (such as cars, trucks and VRUs) in the road scene taking into account any spatio-temporal changes in their states based on historical observations.

Prior work focused on the motion prediction task using either *model-based* methods or *data-driven* methods. Model-based methods include conventional filtering techniques (i.e., kinematic-based models, such as Kalman Filter) and optimisation techniques, such advanced methods of process control. While data-based methods tend towards probabilistic and learning techniques. There are several review papers devoted to motion prediction and human-driver behaviour analysis. Leon and Gavrilescu [55] provide a joint review on tracking, prediction and decision-making for effective automated driving policies. Shirazi and Morris [56] published a review study on vehicle monitoring, behaviour and safety analysis at intersections. A review of unsupervised methods for vehicle behaviour analysis with a special focus on trajectory clustering and topic modelling methods published in [57]. Yurtsever *et al.* [58] provide an overview of state-of-the-art automated driving system with software-hardware related practices. Their paper covers few key functions of the automation driving system framework (connected systems, localisation and perception, planning, Human-Machine Interaction (HMI)), as well as newly released datasets. Lefèvre *et al.* [28] give a comprehensive survey on vehicle behaviour analysis and motion prediction with risk assessment in the automated driving context. Lefèvre [28] pioneered the categorisation of motion prediction approaches that relied on three types of methods: physics-based, manoeuvre-based, and interaction-aware prediction methods that set the interaction between neighbouring vehicles at the core of motion prediction task. Physics-based prediction methods undergo on the physical laws, manoeuvre-based methods apply probabilistic and "Old-Fashioned" ML [e.g., support vector machine (SVM) [59]] and interaction-aware methods that consider the reactive part of multiple vehicles in the prediction.

However, at the time of publishing their study in 2014 fewer studies were implementing manoeuvres and interactions to predict motion of other vehicles. Since then, several advanced approaches have emerged for predicting motion.

Mozaffari *et al.* [41] provide a systematic and comparative review of ML and Deep Neural Networks (DNNs) applied for predicting other vehicles' behaviour for AV and Connected and Automated Vehicles (CAV). Their classification of deep learning based motion prediction approaches applies three criteria: input representation, output type and prediction method. Leon and Gavrilescu [55] published a literature review on dynamic objects tracking and trajectory prediction methods of the AV in mixed traffic conditions.

However, to the best of our knowledge, fewer survey papers are published on vehicles and VRUs behaviour analysis, and none of them has a special focus on the motion prediction of different road users, in highly interactive automated context and identification of principal challenges to deal with for SAE (level3+).

This paper has given the main focus on intention-aware methods and those with interaction-awareness. Intention-aware motion prediction methods are presented in Subsection V-A. These methods are of two *classes*, model-based and data-driven methods.

B. Methodology

The scientific literature supporting this paper was collected from a variety of online databases (e.g., Google Scholar, Web of Science, ResearchGate, etc) using various keywords: [Auto* **OR** Self* **OR** Self-drive* **OR** Car* **OR** Vehicle*] **AND** [Intent* **OR** Manoeuvr* **OR** Maneuver*] **AND** [motion prediction **OR** motion planning **OR** Traject* Predict* **OR** Traject* Forecast*]. The search started in 2015 and ended in

2021, with the result that more than a hundred contributions related to the field were found. From the reviewed literature are three main critical remarks:

1) In this paper covers studies on data-driven-based algorithms and model-based algorithms. Both aim to predict the future motion of the surrounding traffic.
2) Must be on high-level AVs.
3) Must be on the task of motion prediction in high-level automation context.

The selection criteria for theory-oriented articles are as follows.

1) Must be published in a scientific journal or proceedings.
2) Published studies applying conventional prediction approaches and those based on deep learning approaches were covered.
3) Published studies modelling interactions and do not modelling interactions were covered.

The selection criteria for application-oriented papers are as follows.

1) Reviewed works have been published in a scientific journal or proceedings.
2) The automated driving discipline.
3) The aimed field of study covers motion prediction in the automated driving context.
4) Reviewed works are appraised in simulation or naturalistic driving conditions.

This paper encompasses the most recent studies on the field-of-study of motion prediction in AVs, we present a review with the objects presented below:

- *Prediction methods* in Section. V, summarises efforts of most recent studies in comprehension and implementation of motion prediction approaches in automated driving context.
- *Classes* in Subsection V-A and Subsection V-B, classifies motion prediction methods into two classes, model-based class and data-driven class. It also shows relevant contributions geared towards motion prediction with different parameter settings in various driving scenarios.
- *Algorithms* in Subsection V-B.2.b, analyses mostly applied and recent prediction algorithms in the state of the art.
- *Datasets* in Section. VI, concisely describes utilised datasets to implement reviewed algorithms.

V. Motion Prediction Methods

This section classifies *prediction methods* into two dominant classes, intention-aware methods and those relying on interaction-awareness. Each of these classes are then classified into two sub-classes, model-based methods or data-driven methods.

A. Intention-Aware Motion Methods

This paper will use "intention prediction" to denote the prediction of intended manoeuvres. Intention prediction theorises the trajectory prediction as being highly correlated with the executed manoeuvres, that is to say, current manoeuvres of a vehicle are relative to its subsequent trajectory. Pentland and Liu [60] is a pioneering scientific paper in the field of intention prediction. Their approach consists of setting up the on-road decision-making at the centre of the human factor in the vehicle control chain. Thus making the intention prediction crucial in automated driving, which ease the prediction of other vehicles future locations in the scene and help to understand how effective the prediction is, in practical cases from road safety stand point. For instance, a lateral manoeuvre prediction error does not have the same impact as a longitudinal manoeuvre prediction error in roadway design. A safe intention prediction is an important factor to avoid crashes [61]. Scholarly research on road safety over the past years has provided substantial literature on the topic. Predict the intention of other vehicles gives a high-level description of their future displacements. For instance, manoeuvres (e.g., lane change, overtaking or re-acceleration and braking) during stop-start driving helps improve other vehicles' trajectory prediction [42] and which is a prominent stage towards reliable path planning and traffic understanding [62]. Existing methods for predicting the intentions of other road users are generally based on the principle of early detection−fewer seconds in advance−where the human driver intends to go (or stay) on a particular lane. Recent work generally uses data-driven methods for prediction. However, other equally interesting methods such as model-based methods can be as effective as other methods. The two methods have their advantages and downsides. The aim is to predict early positions of the targeted vehicle a few seconds in advance, while improving the prediction capability over an extended prediction horizon, where the prediction response tends to deviate sharply from the safer trajectory to follow, as illustrated in Fig. 5.

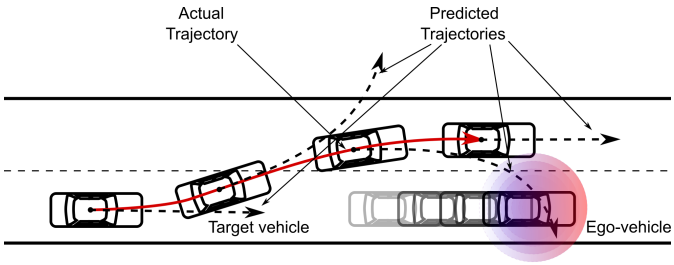

Fig. 5. Erroneous lane change prediction over a long term: Trajectory prediction challenge when the TV has non-monotonic movement. the prediction cannot always fit the movement pattern. The prediction can only be accurate over a short term, and begin to deviate from the right trajectory, once the lane change spans a longer term.

1) Model-Based Intention Prediction Methods: In [63], they proposed a fuzzy logic rules and Bayes filtering method for modelling probabilistic finite-state machines system and predict the intention. Kasper *et al.* [64] applied Bayesian networks and vision-based object detectors for classification and prediction of the SVs intention on freeways, using naturalistic data to test and validate. These approaches are effective to predict the intention in a short prediction horizon (i.e., less than 1s). Shin *et al.* [65] proposed an internal modelling method and Extended Kalman Filter (EKF) to predict intention in urban intersection. Houenou *et al.* [66] proposed an intention

prediction method based on a statistical distance performance indicator. It computes the distance between observed positions of the TV and those of the EV (a lower (dis) similarity value is preferable, a proper differentiation of the prediction propositions, so the next location is considered as being the current one). Otherwise, the TV is considered being in a lane change. In [66], they proposed a mixture of constant yaw rate and acceleration to construct a dual system for predicting trajectory and intention, although their approach is effective for predicting intention. It requires, however, a hand-defined cost function beforehand, in order to predict the likelihood trajectory over a prediction horizon of 4s. Likewise, the approaches proposed in [63] and [64] have shown less reliability in predicting intention over a long term, and may thus show a drop in performance beyond an extended prediction horizon. Most of model-based prediction methods operate on limited history (i.e., observation history) to achieve the Markov propriety−the predictive system depends on the current observed states−while being usable within an integrated prediction. Another drawback of model-based methods is that they perform well under controlled conditions, but would show performance drop in complex situations. In addition, these methods rarely focus on modelling interaction among vehicles. Furthermore−so far−most of them have only been successful in very limited prediction horizon.

Table II provides a summary of model-based studies on intention prediction. It also gives the data or simulator (if available) used to validate and test. We report advantages and disadvantages of each study.

2) Intention-Aware Based on Data-Driven Methods: The problem of predicting upcoming intentions using ML solvers can be achieved in two ways, formulate as a classification problem, or a regression problem, resulting in a probabilistic distribution of the sensor data with intention as the response variable. Regression approaches primarily try to predict subsequent short-term positions and long-term trajectories of a vehicle. This is to say that intention prediction model learns or encodes probability distributions of the manoeuvres rather than simply classifies them, which impedes the inherent stochasticity of the driving behaviour. In other words, the manoeuvre prediction has endogeneity or uncertainties which can be addressed by random terms in a regression problem rather than a discrete classified output labels. The prediction can be biased due to classification accuracy errors. The intention prediction may not be constrained or restricted to a number of class labels than to know the statistical distribution of exogenous factors with a highest likelihood of the possible driving intents. However, there have been several approaches focusing on predicting on-road vehicle behaviour by classifying manoeuvres. The key concept is that manoeuvres are determined from a given set of data, which is allocated into multiple classes. For instance, Lane Keeping (LK), Lane Change to the Left (LCL), and Lane Change to the Right (LCR) are three fundamental manoeuvre classes. Each of them can be labelled and processed in a discrete temporal sequence of measurement to feed learning models. Thus, sequences form a succession of critical manoeuvres that need constant scrutiny to optimise and increase the learning predictability of the algorithm. The types of manoeuvres and their variations in their number are an application-driven issue. They also often revolve around the driving scenario and can be roughly limited to recognised cases. For instance, in highway driving scenarios, three manoeuvre classes are typically chosen: LCR, LCL and LK. Whereas, in urban or city driving scenarios, typical manoeuvres can be set as: Turn Left (TL), Turn Right (TR), continuing straight/ Go Straight (GS) and U-turn in the intersections, which is often excluded or decomposed into recognised manoeuvre sub-classes (e.g., CL, LK, CL) [66]. To refine the prediction performance, recent studies explore additional manoeuvre sub-classes [67]. For example, in [68], they divided the LCL and LCR into three subclasses/ sub-levels (preparation, insertion and adjustment). This stratified categorisation of the class labels shows a clear improvement in the prediction performance of the intention. The higher the number of class categories is the higher prediction rate as the imperceptibility or the latency of the manoeuvre is decreased. In the following, intention prediction methods are classified into three subcategories: (A)-simple probabilistic methods; (B)-DNN based methods; and (C)-multi-agent interactions methods.

a) Conventional probabilistic methods: These methods apply conventional (shallow) ML methods, like Multilayer Perceptron (MLP) and extended random decision forests in [69], [70], and [71] to predict the intention of the SVs.

In [23], they predicted the intention of the SVs by adopting a "divide and conquer" strategy during prediction. They combined MLP with Gaussian mixture regressor−the former classifies intentions and the latter estimates the trajectory of the SVs. The Gaussian mixture gives a multimodal prediction on a non-public dataset. Their approach models prediction with complex multimodalities on motorways. Wirthmüller *et al.* [72] proposed a straightforward learning approach that operates on highways. Their approach uses publicly available top-down view data for training and testing, which becomes de facto standard sets of data for developing predictive methods. Performance-wise of their intention classifier achieved an Area Under the Receiver Operating Curve (AUC-ROC) above 97 % over a prediction horizon of 5s. One drawback, their approach drastically lacks generalisation. In [73], they tried to solve intention prediction problem as a multi-class classification problem by using SVM and MLP methods. SVM-MLP combination is trained and tested on NGSIM dataset, with MLP's prediction performance greater than SVM for 98.8% of precision and the ability of MLP for predicting a lane change 2.4s before occurring.

Wissing *et al.* [69] proposed a probabilistic intention prediction approach trained and validated on simulation and naturalistic driving data that can achieve a root mean square error (RMSE) less than 1m before a lane change occurs. Wissing *et al.* [69] predict the intention of the SVs and estimates their lane change intent by using cubic polynomials and quantile regression forests. Li *et al.* [74] applied the Dynamic Bayesian Network (DBN) to predict the intention of SVs on highways. DBN can handle multiple cues of the historical states of all observed SVs, including their occupied road space and probabilities of low-level interactions between

TABLE II

SUMMARY OF RECENT WORKS ON INTENTION PREDICTION METHODS BASED ON MODELS: CLASS, DATA OR SIMULATOR (DATA/SIM), RELATED WORK (WORKS), PROS AND CONS OF EACH WORK (PROS/CONS) AND SUMMARY OF EACH WORK

Class	Data /Sim	Works	Summary	Cons	Pros
Intention-aware prediction based on models	N/A	Shin et al. [65] 2020	Intersection-target-intent prediction based on IMM and EKF. These models incorporate kinematic and dynamic models of the TVs. The positioning performance is improved demonstrating high accuracy in different urban operations.	1)-IMM realised a quite good performance for intention prediction;2)-the EKFs perform multi-target state estimation in urban intersection.	1)-the intention prediction is performed only on the TVs;2)-the SVs' interaction is not considered;3)-the intention prediction is performed over a short prediction horizon (less than 1s).
	N/A	Li et al. [62] 2020	The MPC controller is applied to predict lane changes and performs a short-term prediction of the trajectory on highways.	1)-an online optimal inputs are generated from real driving environments;2)-the trajectory prediction is performed on the SVs and TVs;3)-the algorithm defines gaps between the leading and trailing SVs in target lanes to initiate a safer lane change.	1)-The algorithm requires a handcrafted cost function;2)-the algorithm looks only at few SVs' history states to predict their intention;3)-the algorithm provides the EV centred trajectory;4)-the prediction is performed over a prediction horizon of 1s;5)-all observed the SVs are assumed not to initiate a lane change;6)-the algorithm is mainly based on computer vision.
	Self-gathered data	Xie et al. [184] 2018	IMM trajectory prediction (IMMTP) algorithm investigates lane change intention on highways by combining IMM, UKF and DBN; UKF predicts the vehicle's trajectory and DBN predicts intention with uncertainty. UKF/DBN outputs are averaged by the IMM to refine prediction.	1)-The prediction is performed on the SVs over short- and long-terms and includes the vehicle dynamics;2)-UKF can handle the uncertainty of nonlinear systems and showed robustness to noise;3)-physics-based model is only effective in 1s, while intention-based and IMMTP models maintain a good performance over 8s.	1)-The algorithm assumes constant vehicles' dynamic (velocity and acceleration) parameters in the near future, which is not the case in the real world;2)-an optimal structure of the DBN is complex to obtain even with optimisation algorithms (e.g., genetic algorithm);3)-the algorithm does not consider the SVs' states.
	NGSIM	Yang et al. [185] 2018	A constrained optimisation based approach generates and update a time-independent polynomial for different lane change situations, such as the existence of gap between the vehicles on the target lane and sudden acceleration of the SVs which closed the gap, by adjusting the weights of the TTLC and lateral acceleration in the cost function.	1)-The algorithm uses a naturalistic driving data for validation;2)-the algorithm can handle the SVs' state changes at different time steps	1)-The cost function needs to be tuned for each situation;2)-the algorithm is not fully automated with limited adaptation to the environment change.
	CarSim	Luo et al. [186] 2016	TTLC and distance-to-lane-change (DTLC) criteria are incorporated into the intent inference framework for a prediction with low latency and high accuracy, and also keeping a safe distance between the SVs and EV.	The algorithm considers the surrounding information during the prediction.	1)-The algorithm assumes the SVs with constant dynamic values;2)-the algorithm has a limited adaptation ability to the environment changes.
	N/A	Nilsson et al. [108] 2015	The algorithm deals with the intention prediction problem as a longitudinal planning problem, which is solved by the MPC and thus generating the future longitudinal trajectories. If such a longitudinal trajectory exists, then the algorithm will generate the corresponding lateral trajectory.	the algorithm provides separate predictions of the lateral and longitudinal manoeuvres, and thus can provide better understanding of the prediction effectiveness different operation scenarios.	1)-The algorithm can be less effective to predict longitudinal manoeuvres;2)-the algorithm may show high computation time;3)-the algorithm assumes that all SVs are travelling at constant longitudinal speed without performing lane change manoeuvres on the prediction horizon;4)-the algorithm focuses only on the TV and assume constant parameters of the SVs.
	N/A	You et al. [61] 2015	The algorithm adopts a series of dynamic circles representing the geometric coordinates of the EV and those of the SVs in a collision avoidance scenario.	1)-The approach is simple to model with a performance-wise comparable or slightly better than other model-based methods;2)-the algorithm generates collision-free trajectories by mapping different obstacles in the road scene.	1)-The algorithm may show poor performance-wise in different driving situations;2)-the algorithm does not consider interaction between the SVs;3)-the algorithm assumes the SVs with a constant speed.

the SVs. Their approach achieved an F1-Score of 80% in the lane change prediction, on NGSIM data with an average prediction horizon of 3.75s. Sefati *et al.* [75] proposed a tactical decision-making approach based on intention prediction. Their approach is formulated as reinforcement learning mechanism (an agent learns through trial and error. When it takes a desired action, the model receives a reward) for predicting the intention of other road users (or agents) on urban driving

scenes. Their approach uses continuous Partial Observable Markov Decision Process (POMDP) to operate in automated driving context, and incorporates the uncertainty estimation. The uncertainties in intention inference are designed using a DBN. In their work, they decoupled the vehicle's behaviour into lateral behaviour and longitudinal behaviour. When the former regards scenarios with tactical manoeuvres, such as lane keeping and lane changes to the left or right of a vehicle, the latter regards acceleration, breaking or keeping a constant speed.

While the abovementioned methods described an intention prediction approach, they were not complemented by adaptive temporal constraints that are critical for achieving motion efficiency and reliable prediction. Successful AV navigation in dynamic, unpredictable and ambiguous real-world scenarios such as non signalised intersections or roundabouts needs firstly the ability of predicting the states of surrounding road users with a better time-space understanding of their possible behaviours using contextual information and secondly the ability to incorporate temporal constraints in behaviour prediction process that leads to the AV achieving safer motion, driving efficiency and acceptable driving comfort. Additionally, these methods can be good at predicting intent over short-term, but would be less applicable to time-series forecasting problems. Consequently, they are less effective over an extended prediction horizon, as the temporal dependency being rarely considered and mostly bypassed.

b) DNN-based intention prediction methods: With the recent success of deep learning in many domains, recent studies have focused on DNN based methods for intention prediction. Specifically, the Recurrent Neural Networks (RNNs) has shown promising performance for sequence learning, and many studies have employed Long Short-Term Memory (LSTM) [76] and Gated Recurrent Unit (GRU) networks for time-series intention forecasting given past sequential observations. Methods based on RNNs are often combined with Convolutional Neural Networks (CNNs) that are known for their space invariant property and particularly applied to encode scene context features by using pixel-wise segmentation and image analysis techniques.

To predict the Time To Lane Change (TTLC), Dang *et al.* and [77] applied LSTM with two dense layers. Their approach predicted the TV's intent, and the SV features are not taken into account during the prediction stage. It achieved a low average prediction error of 0.3m, over a prediction horizon of 3s, when feeding the LSTM with history observations of 3s. Based on their early studies, Wirthmüller *et al.* and [22] combined the LSTM structure and dense layers for better prediction performance, instead of using only an MLP network to classify and forecast the time to the next lane change of SVs on highways [23], [72]. In their study, the built regressor predicts the intention of the SVs by using a prediction criterion of the time to an upcoming lane change before it occurs, i.e., TTLC on highways. They have used the point in time where a TV must cross the lane marking [23]. The investigated TTLC problem on highways (highD dataset) based on RMSE metric. The approach achieved lower prediction errors below 0.25m over a prediction horizon of 3.5s before an lane change occurs. In [71], the SVs' intention is predicted using a mixture weight technique. Altché and de La Fortelle [42] proposed an MLP-based approach to classify and predict the lateral intentions of the SVs on freeways. Their study combined the MLP with LSTM encoder-decoder to generate the subsequent trajectories of SVs, using the NGSIM dataset for training and testing which achieved an RMSE of 0.09m over a prediction horizon of 5s. In [78], they applied the Quadratic Discriminant Analysis (QDA) and LSTM layer to predict the intention of a vehicle approaching a T-junction. Unlike the QDA system, LSTM has the capacity to handle multiple consecutive time steps during prediction. Messaoud *et al.* [79] proposed an algorithm based on LSTM encoder-decoder that infers the intention of SVs on freeways. The LSTM encoding block learns the spatial probability distributions of the history states of the observed SVs in the road scene. The LSTM decoding block interprets and predicts the parameters of bivariate Gaussian distributions. They trained and assessed their method on two datasets (NGSIM and highD datasets) over a prediction horizon of 5s. It realises an RMSE of 4.3m and 2.9m in each dataset, respectively. Benterki *et al.* [80] designed a two-output deep neural structure based on LSTM and GRU [81] to analyse the spatial-temporal features of the history observations. First, two deep neural structures are built and trained (LSTM to predict the lateral manoeuvres and GRU to predict the longitudinal manoeuvres), their outputs are then merged through a Fully Connected (FC) layer, to finally predict the intent of 6 SVs around the TV over different horizons of prediction. Through sequential learning, the system is fed with 10s of history observations from NGSIM dataset and predicts lateral manoeuvres with an RMSE of 0.41m, and 0.42m for the longitudinal manoeuvres over a prediction horizon of 5s. Later, Benterki *et al.* [21] improved their earlier study [73] by combining an MLP network to classify manoeuvres and LSTM to generate future trajectory and predict the intention of 6 SVs around the TV on highways. Precedent studies showed reliable performance over an extended prediction horizon (generally, 5s of prediction horizon). However, the modelling of safety critical motion models that are underrepresented in trajectory training data are seemed to be skewed and lack to model efficiently. In addition, they are constrained in generalisation, and their effectiveness is limited to training data [82]. Hu *et al.* [83], recently, proposed a "generic" intention prediction framework based on Semantic-based Intention and Motion Prediction (SIMP). Their approach adapts to various driving situations, such as roundabouts, highways, or intersections. SIMP generalises quite well in different driving scenarios. The algorithm considers lower order interactions between the SVs (up to 7 SV). However, it is first order Markov model, i.e., the input features depend only on the current time-step.

c) Multi-agent intention prediction methods: These methods attempt to solve the problem of intention prediction as a multi-agent learning system [84], [85], [86], [87], [88] and also examine the temporal and spatial relationships among road users interacting in the same road space.

Alahi *et al.* [89] proposed the social pooling design that was a seminal work in intention and trajectory predictions

by modelling interactions. Deo and Trivedi [67] improved the social pooling design by adding a 2D-CNN layers to better capture the multimodal future behaviours of the agents. Since LSTM can only capture the information temporal structure from the trajectory sequences, but may fail to handle the spatial relationships among the vehicles. Social pooling remedies this issue by pooling LSTM hidden states of spatially proximal sequences in a form of target-centric map or social tensor. In other words, the social pooling design of [67] provide additional improvement to the original structure [89] by adding a convolutional with max-pooling layers to the social tensor. The two methods can learn the spatial relationship among trajectories observed in the surrounding.

These methods learn the vehicles' intention through the LSTM decoder with a softmax layer, which is built upon the Social LSTM (S-LSTM) [89] and Convolution-Social LSTM (CS-LSTM) [67] to predict the future locations for predefined manoeuvre classes. Moreover, these approaches achieved competitive performance on NGSIM data and highD data over a prediction horizon of 5s. For instance, CS-LSTM achieved an average RMSE of 2.07m and 1.16m at 3s before a lane change can occur on NGSIM data and highD data, respectively.

As mentioned earlier, the work of [67] combines the LSTM and CNN to capture the temporal and spatial features among the SVs. Their approach encodes and decodes intention of the SVs. Recent intention prediction approaches combine the LSTM and Generative Adversarial Network (GAN) structure for prediction [90], [91], [92]. The LSTM network is effective to learn time series data. However, GAN counts two main limitations [93]: (1)-in the learning stage, GAN often shows a mode collapse, which negatively affects the model convergence to the distribution and thus unable to capture and produce diverse outputs [94]; (2)-GAN's generative and discriminative blocks are likely to create learning conflicts, while inducing unstable learning [95], [96].

Inspired by [97] which showed superior prediction performance over that of the LSTM structure, by introducing temporal convolution layers (i.e., 1D-CNN layer). Mersch *et al.* [51] proposed to replace the LSTM layer with a spatio-temporal 2D-CNN and 3D-dense layers to predict the intention of up to 7 SVs. In terms of performance, their technique can be quite effective over a prediction horizon of 5s with an RMSE of 1.34m before the start of the lane change execution phase on HighD data. However, their method performs less well on NGSIM data, as this dataset is more challenging than HighD data, and achieved an RMSE of 4.05m over a prediction horizon of 5s.

The precedent studies predict the intention of vehicles using a multi-agent system of interacting entities in a shared environment. In a multi-agent system, agents are independent in that they have independent access to the environment. Therefore, each agent should incorporate a learning algorithm to learn and/or explore the environment. Notwithstanding their clear advantage in producing a rational prediction of the intention, most multi-agent algorithms for intention prediction are complex and often lack of explainability because of their black-box nature. However, these methods overlap with interaction-aware methods. Therefore, the three-level categorisation scheme proposed by Lefèvre [28] needs to be updated.

Table III summarises data-driven works on intention prediction. It also reports different datasets used for training and testing, with advantages and disadvantages of each study.

B. Methods of Predicting Motion With Interaction-Awareness

Existing motion prediction methods with interaction awareness are deepened with game theory approaches [98]. These are effective at analytical strategic behaviour of rational decision makers. Recently, interaction-awareness approaches have been massively investigated and often framing the motion prediction problem as a multi-agent system, which makes the problem more tractable. Motion prediction methods based on interactions have shown effectiveness when operating in highly interactive environments [99], [100]. This paper classifies the motion prediction methods, built with interaction awareness, into two classes: model-based methods and data-driven methods.

1) Model-Based Prediction Methods: These methods apply conventional techniques, such as Bayesian formulation [101], Monte Carlo simulation [102], Hidden Markov Model (HMM) [103], Kalman Filter [104], MPC [105], Gaussian Mixture Regression, DBNs, and so forth. In the following, we divided model-based methods into two groups: (A)-optimisation-based methods; and (B)-rule-based methods.

a) Optimisation-based methods: Model Predictive Control (MPC) is a popular optimisation utility-based method widely applied for trajectory planning [106], lane change prediction [62] and lane merging [107] tasks in high-speed structured environments. Nilsson *et al.* [108] tried to simplify lane change inference problem by checking first if the EV is positioned in a gap in the target lane and then MPC is applied to generate loose trajectories in both lateral and longitudinal positions to perform a lane change. Suh *et al.* [109] applied MPC for lateral manoeuvring that uses pre-defined cost function to choose whether a lane change manoeuvring is required and, if so, in which direction (to the left or right). Nilsson and Sjöberg [110] and Nilsson *et al.* [111] dealt with intention prediction as an optimisation problem using MPC. They optimised the weighted effects of acceleration and braking of the vehicle according to the shape and feasibility of the manoeuvre and thus trajectory. They gave a straightforward way to turn the problem into a well-defined optimisation problem by applying a specific solver. Manual change of weights is challenging in addition to optimisation formula that is handcrafted. While MPC can handle disturbances by rejecting noise, their drawback is they cannot address the uncertainty as they usually assume the predictions to be accurate. Also, these methods do not consider information of the surrounding traffic [112].

Kalman filter is also largely adopted for optimally predict motion with associated uncertainty. Xu *et al.* [113] and Deo and Trivedi [67] applied two different constant velocity Kalman filters to predict the trajectory on highways. These give different prediction outcomes but fewer explanation specific to their implementation or parameters choice which is

TABLE III

SUMMARY OF RECENT WORKS ON INTENTION PREDICTION BY USING DATA-DRIVEN METHODS: CLASS, DATA OR SIMULATOR (DATA/SIM), RELATED WORK (WORKS), PROS AND CONS OF EACH REPORTED STUDY AND RELATED SUMMARY

Class	Data /Sim	Works	Summary	Cons	Pros
Intention-aware prediction based data-driven methods	NGSIM US-101	Benterki et al. [21] 2020	A hybrid method combines the intention classification and trajectory generation of up to 6 SVs : MLP classifies the intentions of all vehicles and the LSTM encoder-decoder encodes the vehicles' track history and then decoders the subsequent SVs' trajectories.	During the training phase, the temporal consistency of the vehicle intent can be biased, because the MLP cannot solve time series forecasting problem.	1)-The algorithm considers up to 6 SVs;2)-it predicts the trajectory and classify three manoeuvre classes;3)-the algorithm performs a prediction over different horizons of prediction (1s, 3s and 5s);3)-tests are performed in real hardware and software systems under different environment constraints;4)-the vehicles' intent can be predicted 2.2s before a manoeuvre occurs and anticipate future trajectory with 0.3m and 3.1m of lateral and longitudinal position errors, respectively.
	NGSIM	Benterki et al. [80] 2019	Two-output RNN-based structures (LSTMs and GRUs) and a dense layer with a softmax layer to predict probabilities of lateral/longitudinal intentions of the SVs.	The lateral model yields an RMSE of 0.41m and longitudinal model yields an RMSE of 0.42m when both predict over a prediction horizon of 5s.	The dual predictor is effective at predicting longitudinal positions of the SVs than their lateral positions. This is a critical in automated driving context.
	NGSIM, HighD	Benterki et al. [68] 2021	The MLP is trained for predicting the intention of up to 5 SVs on highways. A three-level intention predictor applies at each level the dynamics and kinematic states, Yaw and YawRate of the TVs in the MLP training/testing stages. Tests are performed over different horizons of prediction (3s, 4s and 5s). The MLP achieved an accuracy of 93.23%, 97.10% and 98.47% of correct classified intention, respectively.	The MLP is unable to map time sequences of historical states of other vehicles to predict their future behaviour.	1)-The algorithm is effective in predicting correct intentions over a prediction horizon of 5s;2)-The algorithm applies SHAP (SHapley Additive exPlanations) [188] to better understanding the prediction effectiveness of each feature ($x-y$ positions, velocity, acceleration, TTC, Yaw, YawRate).
	NGSIM	Benterki et al. [73] 2019	A three-class manoeuvre classification algorithm based on the MLP and SVM kernels.	1)-The algorithm does not establish interaction between vehicles;2)-the algorithm does not consider the temporal consistency during the learning and prediction stages.	1)-The SVM and MLP are simple to train;2)-The SVM and MLP achieved a precision of 97.1% and 98.8%, respectively;3)-the MLP (2.4s) predicts LCL earlier than SVM (2.0s) and MLP ($\approx$ 2.2s) and SVM (1.9s) for RLC.
	NGSIM	Deo et al. [143] 2018	The algorithm improves the LSTM social pooling by adding a 2D-CNN layers and capture multimodal future behaviours (i.e., it decodes 6 manoeuvre classes for predicting the intention).	Local information extracted from the LSTM hidden state is not always sufficient and the generated context vector is independent of the TV state.	1)-the neural structures are trained with six manoeuvre classes representing the vehicles' intention;2)-the algorithm performs a multimodal prediction of the intention;3)-the algorithm is partially-scalable [189];4)-the algorithm establishes interactions among the SVs.
	NGSIM	Wirthmüller et al. [22] 2021 [23] 2020	The MLP applied in [23, 72] is replaced by LSTM [22] for seq-2-seq learning and long-term prediction of the vehicle intention.	The performance-wise comparison of the LSTM is quite biased because their algorithm is essentially based on a set of RNNs with complex "gated" functions, against those of a dense layers, with no recurrent connections, nor handle the data temporal aspect.	1)-The labelling step is simplified;2)-unlike the MLP, LSTM can handle temporal dependencies between sequences and solve the problem of vanishing and exploding of the gradient;3)-the LSTM achieved promising results on unbalanced learning data;4)-the algorithm encodes interactions among the SVs.
	NGSIM I-80, NGSIM US-101	Li et al. [21] 2020	The DBN is applied for long-term trajectory prediction and lane change intention prediction of the TVs within an "attention map".	1)-The algorithm does not consider interactions among the vehicles;2)-an optimal DBN model is difficult to obtain.	1)-The algorithm is effective for predicting the lane change intention with an F1-Score of 80% and a prediction time of $3.75s$;2)-the algorithm outperforms most of conventional approaches and RNNs for intention prediction;3)-the HMM captures the spatial and temporal variability of the trajectory [190].
	NGSIM US-101	Hu et al. [83] 2018	The intention prediction scheme applied goal position and temporal context into each insertion area or probable future location of the TV. The semantic based SIMP uses GMM and multi density mixture to predict multimodal semantic manoeuvres, while generating trajectories of TVs, under different operational scenarios.	1)-The algorithm requires few parameters to be manually tuned;2)-SIMP's overall architecture includes several FC layers that break up the time consistency of the learning data.	1)-The algorithm considers up to 7 SVs at each input frame during prediction;2)-the algorithm provides a quite stable performance in different driving scenarios;3)-the algorithm can handle uncertainty of the driving environment;4)-the algorithm can predict the intention 0.3s before lane change occurs within 3s of prediction horizon.
	Self-gathered data	Zyner et al. [78] 2018	Intention prediction on the destination of travel at a roundabout and a T-junction by using multiple LSTM cells.	1)-The algorithm predicts the vehicles' intent in a specific driving scenario;2)-the algorithm is evaluated over different prediction horizons: 5(0.2s), 15(0.6s) and 25(1.0s)time-steps. This means that the intention prediction can be only effective over a short-term.	The algorithm is computationally effective.
	NGSIM, ETH-UCY [92], Stanford Drone [191]	Tang et al. [51] 2021	End-to-end trainable MATF system combines multi-agent interactive framework and encoder-decoder neural structure : encodes the spatial relationship among agents and the scene context, as well as the history tracks of all agents. MATF uses 2D-CNN to encode the scene context, while preserving the spatial layout of all of them within a fusion tensor−MATF decoding segment interprets the social consistency and scene context.	1)-MATF is not adaptive to the arbitrary number of the SVs;2)-MATF achieved an RMSE of 5.12m at 5s of prediction horizon, which performs slightly less than most methods in the literature.	1)-The algorithm provides a multimedia prediction of the future;2)-the algorithm learns explicit latent behaviour modes of other road users, which helpful to provide consistent prediction over longer time horizon (e.g., 5 seconds);3)-the algorithm is based on the GRU, which is slightly faster computationally than the LSTM.

common in the literature. This makes the result interpretation unclear and biased.

Mercat *et al.* attempted to provide a distinct explanation of the parameter selection related to the constant velocity model in [104]. They proposed a hybrid approach which combines multimodal constant velocity predictor and LSTM module to control predictions. Their approach operates on freeway situations and learns from 3s of history observations at 5Hz, then attempts to predict trajectory over a prediction horizon of 5s. As a result, this approach achieved better Negative Loglikelihood (NLL) than other physics-based models on NGSIM US-101 and I-80 datasets.

Ju *et al.* [114] proposed a hybrid approach based on Kalman Filter using predicted accelerations from LSTMs. Acceleration predictions and Kalman filter are implemented in two separate modules, and then combined. LSTM captures history observations as input and provides acceleration prediction sequences which are then fed to the Kalman Filter. Interactions between vehicles are implemented using a mix of LSTM encoder-decoder and 2D-CNN regarded as a social tensor encoder and a max pooling to pool social features and FC layers regarded as a merger of the social features and merging with LSTM encoder. Finally, the LSTM encoder-decoder interprets all encoded information to predict the vehicles' intention. Another combination of the Kalman Filter and LSTM in [115] that substitutes the state update of the LSTM mechanism for predicting the vehicles' motion. Usually, Kalman Filter-based approaches are inefficient to model complex interactions among the vehicles. Recent studies attempt to overcome this limitation by proposing hybrid approaches such as in [114] and [115].

One of the limitations of optimisation techniques lies in the generation of multiple local minima. Usually model-based methods apply tracking filter over time. However, the growing uncertainties often cause future positions to end up at non-accessible or unrealistic locations. Thus, making prediction biased over a long-term.

b) Rule-based methods: The rule-based approaches have been proven to be reliable in the DARPA Urban Challenge [116], [117], [118]. They are simple to implement and showed robustness in simple driving scenarios. However, rule based approaches can be made tailor made solutions for complex navigation tasks and thus violates the criteria of generalisation. They also falls in short-terms of handling the uncertainty and partial observability [75]. Moreover, these approaches are unable to consider the SVs and become increasingly unreliable for longer-term prediction as the vehicle will change control inputs.

2) Data-Driven Prediction Methods: We classify data-driven prediction with interaction-awareness according to three criteria: (a)-input representations; (b)-output types; and (c)-algorithms.

a) Input representations: This subsection categorises the reviewed studies based on the adopted input representation for on-road behaviour prediction. We consider three sub-classes with the increase of the abstraction levels and type of tracked objects: (1)-track the history of TVs (or interaction-free methods);(2)-track the history of the TVs and SVs; and (3)-track the history of the TVs and SVs and road context.

i) Interaction-free methods: These methods dwell in simple tracking of the history observations of the TV to predict their motion without dealing with any form of interactions.

Zyner *et al.* [119] used the history observations of the vehicles ($x-y$ position, speed and heading) for a particular track snippet leading up time, spatial coordinates to the road intersection, velocity, and orientation are encoded by a recurrent mixture density structure. Their approach operates in an unsignalised urban intersections and provide future positions of the TV in a multimodal mixture form with trajectory of highest likelihood via a clustering method. Xing *et al.* [120] proposed a computer vision based solution that combines different signals of multiple low-cost mounted sensors and data acquisition system. Their algorithm uses a set of Enhanced Bidirectional LSTM (EBiLSTM) for learning and prediction of the TV's motion. Performance-wise, using EBiLSTM architecture can be quite fast, while achieving higher prediction accuracy. However, vision-based intention prediction methods are not always accurate [121].

The abovementioned works track observation history of the TVs, but none of them handle interactions between the SVs. Prediction results can mislead the prediction performance, and thus generate unfeasible trajectories especially under complex conditions [41], and eloquently delineates the performance capabilities of the prediction system. These approaches are unable to reason about on-road dynamics constraints, which may lead to an inconsistent and unrealistic prediction, i.e., the prediction is often not achievable by the underlying control variables (e.g., a vehicle moving sideways).

ii)Motion prediction under social interaction: Social interaction is closely illustrated to that of Social Robots (SR), where robots interact with humans and between each other in the conventional social code [122], [123]. They can smoothly navigate through many social interaction scenarios. SR design was inspired from the human intrinsic behaviour "theory of mind" [5]), which is the capacity of reasoning about people's actions based on their mental states [124]. Reasoning about future motion of the road users is an important prerequisite for a safer, dynamic, and socially aware robotic navigation systems. Social interaction shows several benefits, including stronger understanding of the AV for its situation. This can be reinforced by adding semantic information and heterogeneous variables to the prediction algorithm [31], [33], [125], [126]−complex interactions can occur in heterogeneous traffic situation in dense traffic flow [127]. Setting up interactions helps the AV aggregate important information within the prediction framework. Thereby, it can learn interaction patterns from heterogeneous participants, resulting in accurate prediction of their future trajectories, and retrospectively make reasonable navigation decisions. As a result, imbuing automated schemes with interactions would strengthen decision-making and refine proactive responses to be drawn towards nearby road users. Moreover, the probability of critical situations

[5]In multi-agent settings, ontological approaches define a set of rules that agents are expected to follow, or an analytic design representing an agent's internal decision-making strategy.

can be mitigated by the AV through a social interaction prediction design. Where each road user can be adequately informative of its past and current states. Social interaction designs have become the core component of the modern robotic system [33], especially in critical safety applications such as automated safety frameworks. In practical situations, participants simultaneously interact with each other. Modelling interactions between road users is an open challenge. Previous studies have been devoted for this task by either modelling a representation of one or a group of road uses [128] or by modelling a static environment scheme, such as LiDAR beam simulation [91]. Nonetheless, these approaches have shown some limitations, e.g., a static environment representation induces to a learning problem because the data that might be extracted upstream cannot be used. On the contrary, traffic representations scale computationally with the quantity of feasible interactions, which generally require expertise to define the features to use

- **Track the history of the TVs and SVs.** These approaches rely on modelling interactions among the SVs and TVs to predict motion, thus can enhance the effectiveness of the prediction task. A central capability for any prediction model is to be able of predicting as many interacting vehicles as are available in the vicinity of the AV. The increase in the number of SVs have a higher likelihood for a better prediction performances [42], [129]. This will give the AV a better understanding of the world around; be aware of "who is doing what, when and where?" Thus makes it possible to build a more realistic prediction with accurate decision-making. Dong and Dolan [112] proposed a non-parametric lane change intention prediction by adding interactions between SVs using Reproducing Kernel Hilbert Space (RKHS). Their approach considers history states of up to 5 SVs simultaneously for predicting their future locations over a finite prediction horizon. Dong *et al.* [129] improved their previous method [112] by proposing a combination of Recurrent Meta-Induction Network (RMIN) and Conditional Neural Process (CNP) able to handle up to 7 SVs. Performance-wise of RKHS's prediction can be quite efficient with a maintained lower values of loss over a prediction horizon of 2s. Their approach showed that LSTM is not effective in predicting longitudinal lane change. Although the lateral performance using LSTM is comparable to that of other methods of the state of the art. However, the prediction performances are likely to deteriorate over an extended horizon of prediction. Depeweg *et al.* [130] used a Bayesian Neural Network (BNN) learns from a controllable learning space with latent variables. Their approach is scalable and flexible as a probabilistic design with the ability to regularise uncertainties during prediction (model uncertainty via a distribution over weights−epistemic uncertainty). They used latent variables to capture complex noise patterns in the data (distribution over the latent input variables−aleatoric uncertainty). Inspired by the finding of [130] for uncertainty components decomposition, [34] quantified uncertainties by using a cost-based DNNs and reinforcement learning to imitate human-driving behaviour and predict intention of all SVs on highways. Their prediction model learns from the history of all SVs to predict their future states. It also encodes the driving scene via a Dynamic Occupancy Grid Map (DOGM) and provides multimodal prediction of the vehicle intentions. However, when predicting, SVs do not respond to the controlled vehicle (i.e., the EV) and this makes interactions less reliable, as in some cases, SVs do not try to avoid collisions. Kuefler *et al.* [91] applied reinforcement learning and generative adversarial imitation learning [131] to imitate and predict human-driving behaviour. Their approach perform well only in a specific scenario because its performance crucially depends on the road context.

  Tables. IV summarises motion prediction methods with interaction-awareness which track the history states of the TVs and SVs. It also summarises the advantages and disadvantages of each reported work.
- **Track the history of the TVs, SVs and road context.** Interaction methods with a road context-awareness show effective performance in highly interactive driving scenarios [35], [99], [100]. These methods take into account multiple sources of information for the learning and prediction tasks. These include sequences of relative positions of the TVs and SVs, including their complex interactions and physical and logical representation of the road context, as well as the planning information of the EV. These systems are typically set up as a multi-head predictors, where each head learns specific features to predict a desired outcome.

  Multi-head predictors include a backbone network and mix various neural engineers (shallower or quite deeper networks) are applied. This comprises of RNNs, CNNs, GNNs or generative models with a controllable latent space like the Conditional Variational Autoencoder (CVAE) structure.

  Each of these networks has a special demand to fulfil, and their combination sums up their outcomes, for instance, RNNs are effective in capturing the temporal information consistency and 2D-CNNs are shift-invariant or space-invariant and can therefore preserve the spatial features during training/prediction. While graph variants, such as Spatial-Temporal Graph Convolutional networks (ST-GCN) [132] to learn both spatial and temporal data of joint traffic dynamics. CVAE yields a controllable latent space, which is useful for encoding and decoding motion with a minimum of information loss, and generates a Gaussian mixture distribution, and so predicting all possible future behaviours.

  In [31], [43], [129], and [125], a physical and logical representations of the world are applied, with the aim to enhance the motion prediction performance. Besides the history states gathered from other road users at different time-steps in a driving sequence with the scene context, such as the raster image representation of the road scene with $\mathbb{R}^{H\times W\times C}$, where H is the height, W is the width and C is the channel dimension of the raster imaging

TABLE IV

SUMMARY OF RECENT WORKS ON SOCIAL INTERACTIONS BASED ON DATA-DRIVEN METHODS: INPUT REPRESENTATIONS (INPUTS), DATA OR SIMULATOR (DATA/SIM), RELATED WORK (WORKS), PROS AND CONS OF EACH REPORTED STUDY AND RELATED SUMMARY: TRACK THE HISTORY OF THE TVS AND SVS

Inputs	Data/ Sim	Works	Summary	Cons	Pros
Track the history of the TVs and SVs	NGSIM US-101, I-80	Dong et al. [129] 2019	The algorithm is a generative-based model – RMIN and CNP to learn from 10s of track history to predict the next 5s of lane change of the TV and up to 5 SVs.	1)-The prediction error can result in a larger deviation in the longitudinal direction than in the lateral one;2)-the algorithm includes FC layers which do not take into account the temporal motion information.	The algorithm predicts the lane change intention with an MSE of 0.248m for both longitudinal and lateral positions over an extended prediction horizon of 5s.
	NGSIM I-80 and US-101	Deo et al. [179] 2018	The algorithm is trained on 8s of trajectory segments (3s of track history and 5s of post trajectory segments) for predicting the intention of other vehicles.	The algorithm works well only on highways but can expect a performance drop in other driving scenarios.	1)-The algorithm achieved an RMSE of 4.66m over a prediction horizon of 5s;2)-the algorithm improved the social pooling design by adding a 2D CNN layers;3)-the algorithm provides a multimodal prediction of the manoeuvres;4)-the algorithm can take into account up to 7 SVs during the prediction.
	NGSIM I-80, US-101	Dai et al. [140] 2019	The ST-LSTM predicts the behaviour of multiple vehicles. ST-LSTM is a two-layer LSTM – one layer predicts the TVs' trajectory and another layer designs the SVs' interaction. The algorithm can handle up to 6 SVs around each TV.	ST-LSTM is trained and tested on small sets of data, so achieved performance can be biased.	1)-The algorithm Embeds spatial interactions into basic LSTM to model the spatial relationships among the SVs;2)-the algorithm adds shortcut connections between inputs and outputs of the two consecutive LSTM layers in order to overcome gradient vanishment.
	NGSIM I-80, US-101	Dong et al. [112] 2018	The algorithm applies a non-parametric generative model for predicting lane change manoeuvres on highways. They used NGSIM dataset to train/test the neural structure with 10s of track history and 10s of post trajectory segments.	1)-The predictor is less effective for predicting the longitudinal manoeuvres;2)-the prediction error may fluctuate over extended prediction horizons.	1)-The algorithm predicts lateral manoeuvres with an average MSE of 0.251m over a prediction horizon of 3.5s;2)-the algorithm takes into account up to 5 SVs around each TV.
	NGSIM US-101	Altche et al. [42] 2017	The algorithm applies one LSTM layer with two dense layers for encoding and decoding the lateral and longitudinal manoeuvres of the TV. The algorithm can take into consideration up to 9 SVs in the three adjacent line lanes.	The prediction is less effective over extended prediction horizons.	1)-The LSTM provides better results for longer prediction horizons;2)-the NGSIM dataset is used to train and test the TVs' trajectory prediction over different prediction horizons and achieved an RMSE of 0.65m for lateral position prediction over a prediction horizon of 10s;3)-the algorithm achieved competitive performance compared to the multlayer counterparts in pedestrian motion prediction [192].
	NGSIM I-80, US-101	Li et al. [126] 2019	The algorithm combines graph operations, 2D-CNN and LSTM encoder-decoder for predicting social interactions of the SVs and TVs.	The implemented graph-based predictor neglects the static scene context during the modelling procedure.	1)-The algorithm provides a better data scalability with a more accurate performance-wise than other similar approaches;2)-the algorithm designs explicit interactions among vehicles.

scheme. These methods provide useful information of road context [34].

By using probabilistic models, the road scene context can be denoted as $\mathcal{I}_t$ observable at any time-steps T. Trajectory prediction can be performed by incorporating $\mathcal{I}$ into the probabilistic learning model $p(\mathbf{Y}|\mathbf{X}, \mathcal{I}_t)$. The aim is to build a multi-head trajectory prediction model [35] that can accurately model $p(\mathbf{Y}|\mathbf{X}, \mathcal{I}_t)$, in a sequential learning design [133] while modelling interaction between other road users. Li *et al.* proposed GRIP [126] and GRIP++ [134] algorithms that applied GNNs to model interactions among the SVs for predicting their trajectory. GRIP/GRIP++ achieved high precision and recall values in the prediction. However, GRIP/GRIP++ take much efforts to learn the interaction and are computational burdensome. Jeon *et al.* [135] overcame the computational complexity shown by GRIP algorithms, by proposing a deterministic prediction approach which designs the interactions between SVs and aggregates the road context in the prediction. The study built a dual-channel DOGM of SVs and a lane markings data to feed the Enhanced Graph Convolutional Network (EGCN) [136] combined

with LSTM. The EGCN is inherently scalable and predict motion of up to 10 SVs at once, with lower residual kept across 5s of prediction horizon. Li *et al.* [137] addressed issues related to temporal consistency retrieval during prediction by applying Conditional Generative Neural System (CGNS). The CGNS is based on two deep generative systems: CVAE and Conditional Generative Adversarial Network (CGAN). The study linked these two conditional latent space training and variational divergence minimisation criteria. The CGAN is effective to predict trajectory by learning interactions among road users.

Benchmark tests have been carried out on publicly available datasets of pedestrian trajectory dataset and set of self-collected data in a roundabout driving scheme. Their method takes advantage of density learning in a unified generative system, including forward block attention mask to extract and interpret the most probable features of each agent, and a 2D Gaussian mixture attention block. These two blocks select probabilistic maps to be prioritised during the training stage.

The two blocks are merged to predict the conditional distribution of the future trajectories of all interactive road users in the road space by considering their history states and the road context. They pointed out that leveraging the image context on the learning, trajectory can be predicted over an expanded prediction horizon with a slight decrease of performance.

Road context becomes more profitable, resulting in a substantial performance increase on trajectory prediction. Salzmann *et al.* [33] combined the CVAE, GRU and GNN to design social interactions among heterogeneous road users with a spatiotemporal consistency. Their approach can learn the spatio-temporal relationships between road users and provide a multimodal prediction. Lee *et al.* [31] proposed an approach based on front-view semantic image to learn and predict SVs trajectories, including semantic context of the road. The study applied a mixture of multi-layer GRU and CNN with CVAE encoding. However, this method is burdensome for a one iteration and is not effective to capture the distinct characteristics of the manoeuvres.

In [138], they incorporated planning information into the trajectory prediction process. This conditioned the trajectory prediction on present goals on the EV. Their study focuses on predicting the trajectory by incorporating planning information into the learning model during the prediction stage. Nonetheless, it uses only the goal location information inducing some restrictions.

For instance, in an assumed EV control scenario that can be inserted in congested traffic conditions, even given the same goal of the EV, the future behaviour of the SVs will greatly fluctuate according to the time profile and how the EV manages to achieve its goal. Carrasco *et al.* [43] proposed SCOUT algorithm which uses GNNs enhanced with attention module to extract most likely features during the encoding step. SCOUT models interactions among heterogeneous road users, while aggregating the traffic scene (road context) during the motion prediction in a socially consistent way and under mixed traffic conditions. InD and ApolloScape datasets were used to training the trajectory prediction benchmarks. RounD dataset was used to test the generalisation capability of the algorithm. SCOUT performs slightly better than GRIP/GRIP++ on the InD dataset, which was inaccurate for S-LSTM, CS-LSTM and Social Generative Adversarial Network (S-GAN) [139], as these models perform better on highways. However, the S-GAN might be unstable during the learning stage. The model can hardly converge and therefore will generate a slightly biased performance.

Table. V summarises recent studies that the track the history of the TVs, SVs and road context.

b) Algorithms: This section classifies existing *algorithms* into three classes, namely Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs).

i) Recurrent neural networks (RNNs): Recurrent Neural Networks (RNNs) have been largely applied on solving the problem of sequential design of trajectories and on-road behaviour prediction because of their capacity of learning time-series data structure and availability of many publicly accessible databases. The LSTM is one of the most implemented for on-road behaviour prediction of the vehicles and VRUs. LSTM can solve two inherent problems in RNNs, the gradient vanishing and exploding of the weights [76]. The GRU is inspired from LSTM and thus embeds similar advantages with less complexity thanks to their recurrent gated architecture. However, LSTM is unable to simultaneously describe the spatial interactions between the SVs and the temporal relations between the trajectory time-series. Therefore, existing approaches cannot precisely estimate the inter-vehicular influence or dependencies, because to some extent, basic LSTMs are subject to a vanishing gradient and are therefore complex to train for long-term time series data. This often lead to large prediction residual errors in vehicle trajectory prediction. Recently, Dai *et al.* [140] proposed a Spatio-Temporal LSTM (ST-LSTM) to predict trajectory of SVs. Two main enhancements have been implemented: (1)-embed spatial interactions into LSTM to implicitly design interactions between SVs; and (2)-induce shortcut connections that link between inputs and outputs of two consecutive LSTM cells to decrease the gradient vanishment effect in the backpropagation. The shortcut enhancement of their architecture gives LSTM mode la better flexibility to train with a low residual error obtained. However and despite improvements brought to the LSTMs, integrating a notion of social interactions, which results in spatio-temporal consistency in predicting trajectories by using the ST-LSTM. This method has been trained and tested on small sets of data. In addition, the ST-LSTM is deterministic by nature.

Table. VI summarises the most recent motion prediction algorithms, classified by the prediction used methods.

TABLE V

SUMMARY OF RECENT WORKS ON MOTION PREDICTION UNDER SOCIAL INTERACTION BY TRACKING HISTORY OF THE TVS, SVS AND ROAD CONTEXT : INPUT REPRESENTATIONS (INPUTS), DATA OR SIMULATOR (DATA/SIM), RELATED WORK (WORKS), PROS AND CONS OF EACH REPORTED STUDY AND RELATED SUMMARY

Inputs	Data/Sim	Works	Summary	Cons	Pros
Track the history of the TVs, SVs and road context	ETH [193], UCY [194], SDD [191], ID [195], [196]	Li et al. [137]2019	A multi-head trajectory prediction algorithm called CGNS that combines two conditional generative models with 2D-CNN and GRU for encoding and decoding motions.	1)-The algorithm applied latent/compressed learning space to encode original features and reconstruct new features in order to get as realistic of a reconstruction as possible;2)-The algorithm applied a pre-trained VGG-19 which can increase the computational burden;3)-CGNS does not explicitly model interactions among the road users;4)-CGNS does not perform well over a short term;5)-The algorithm does not predict the intention.	1)-The algorithm applied the unsupervised learning techniques without explicit labelling;2)-CGNS achieved slightly better prediction than the state of the art methods, with a displacement error of 0.49m and 0.97m, respectively;3)-CGNS shows good generalisation performance.
	self-collected data (roof-mounted LiDAR)	Casas et al. [142] 2018	IntentNet applied a BEV dynamic, which maps semantic objects in the environment, and performs joint object detection and intention prediction.	The algorithm is not always reflective of the real world conditions.	IntentNet considers the influence of the road scene upon the prediction framework and model interactions among vehicles to predict their behaviours.
	ETH [197], UCY [198], nuScenes [199]	Salzmann et al.[33] 2020	Trajectron++ applied GNN for spatio-temporal representation of the traffic scene, combined with GRU and 2D-CNN to interpret the road semantic. CVAE is implemented to generate a multimodal prediction of the trajectories. The algorithm achieved a best average ADE and FDE of 0.38m and 0.93m for a one-shot prediction respectively and a best average ADE and FDE of 0.19m and 0.41m for a multi-shot prediction (N=20), respectively.	1)-Trajectron++ allocates a GRU to each road user for the prediction, thus inducing an important computational load;2)-The algorithm cannot accurately reason about the key manoeuvres, such as sudden turns and lane change.	1)-Trajectron++ considers various types of road users (vehicles and VRUs);2)-Trajectron++ implements explicit interactions between road users;3)-Trajectron++ can handle multimodality by leveraging the CVAE architecture;4)-The algorithm uses CVAE that learns a smooth representation of the latent space and produces a robust prediction of the future trajectory.
	self-collected data (roof-mounted LiDAR)	Luo et al. [200] 2018	End-to-end DNN based structure that uses LiDAR data for prediction. The algorithm implements FaF Nets [200] for predicting the trajectories over a prediction horizon of 1s.	1)-The environment constraints and weather conditions are not taken into account by the algorithm;2)-FaF Nets can significantly increase of the computation burden;3)-The prediction is performed over a short term.	The algorithm facilitates fusing information gathered from the car on-board sensors.
	KITTI [161], SDD [191]	Lee et al. [31] 2017	DESIRE is a non-deterministic prediction system that considers interactions among the road users and operates on urban environments. KITTI [161] and SDD [191] are used to train and validate the algorithm. These datasets include vehicles and VRUs, and thus useful to learn interactions among the road users in heterogeneous traffic conditions.	The algorithm is computationally burdensome:high-resolution radar image generation and target recognition are two computationally demanding processes.	The algorithm gives a realistic representation of the traffic flow.
	nuScenes [199], CARLA [165]	Rhinehart et al. [125] 2019	PRECOG algorithm designs interactions among road users by combining their previous observed states, LiDAR and road scene images.	PRECOG is not easy to re-implement and computationally burdensome.	1)- PRECOG considers multiple SVs simultaneously (from2 to 5 SVs); 2)-The algorithm learns from different information sources with a minimum-loss;3)-The algorithm explicitly encodes multimodality during prediction;4)-The algorithm operates in heterogeneous driving conditions.
	LOKI [201]	Girase et al.[201] 2021	LOKI is a goal-oriented network-based and socially aware framework uses RGB images/LiDAR point clouds and planning goals to jointly predict the intention and trajectory of the SVs in top-down view images. The intention prediction is performed over a short-term of 0.8s. The algorithm achieved a best average ADE and FDE of 2.76m and 7.26m for a one-shot prediction (N=1) respectively and a best average ADE and FDE of 1.54m and 3.59m for a multi-shot prediction (N=20), respectively.	1)-The intention of the SVs and VRUs can be predicted over a short-term of 0.8s;2)-Two-layer MLP is unable to handle the temporal consistency during the learning process.	1)-LOKI can consider three key criteria:i)-ego-centric agent planned future;ii)-social interactions among the SVs and VRUs;iii)-road context;2)-LOKI operates in heterogeneous and highly interactive traffic conditions;3)-The algorithm generates a multimodal trajectory prediction (N=20) in one shot, i.e., multi-shot prediction setting;4.road context is encoded by using GNNs.
	ApolloScape [202], NGSIM, KITTI 2019, TRAF [127], ETH [193], UCY [194], Cityscapes [158]	Chandra et al.[164]2019	TraPHic is a vision-based algorithm based on LSTM and 2D-CNN for predicting trajectory in heterogeneous conditions with weighted interactions between the road users. TrafficPredict [127] improves TraPHic algorithm by increasing numer of road users to handle in the prediction (i.e., up to 13 road users with 8 vehicles and VRUs that can simultaneously interact).	1)-The algorithm requires a large amount of training data for generalisation because of the high complexity of the model;2)-The algorithm may fail to prediction interactions in some rare driving configurations, such as red light offences.	1)-The algorithm operates in heterogeneous traffic conditions;2)-The algorithm outperforms few state-of-the-art methods on different traffic datasets and achieves 0.78m of RMSE over a prediction horizon of 5s. It can learn from the trajectory history of (2s-4s), and provide a prediction of the trajectory over different prediction horizons (3s-5s).
	NGSIM, CARLA	Zhao et al. [148] 2019	The algorithm applies conditional probability density function on the history tracks of the SVs with semantic latent variables to learn/predict multimodal intentions without explicit labelling.	The algorithm does not explicitly encode the interaction among the SVs.	1)-The algorithm provides a multimodal prediction of the future;2)-The algorithm predicts the intention and generate trajectory of the interactive SVs;3)-The algorithm can handle the road context.
	CARLA [165]	Ding and Shen [203]2019	The algorithm combines high-level intention prediction with a low level contextual reasoning. It applies the LSTM encoder-decoder to predict the SVs' intention.	1)-Modelling and validation of ADS in a simulation environment may show "safe" behaviour that is variable under real-world conditions. This is called the "realistic" gap between the simulation and the real conditions.;2)-The algorithm does not consider the interaction between the VRUs.	1)-The validation tests are performed in different urban traffic configurations;2)-The algorithm encodes the environment feature map and learns the interaction between vehicles with contextual factors;3)-The LSTM decoder provides a multi-shot prediction with a Softmax layer to predict the maximum likelihood manoeuvres.
	self-gathered data	Deo et al. [190] 2018	The algorithm operates on real highway data with the aim to joint estimate future locations with uncertainty of the SVs over a prediction horizon of 5s.	1)-The algorithm can be computationally burdensome by the HMMs applied for each class of manoeuvre;2)-By considering more SVs, the computational complexity of the algorithm may exponentially increase.	1)-The algorithm models the interactions between the SVs and includes the road context in the learning process;2)-The algorithm uses naturalistic driving data captured on Californian freeways for evaluation;3)-The algorithm applies variational Gaussian mixture models (VGMMs) for each manoeuvre class in the trajectory prediction;4)-The algorithm achieves a classification accuracy of 87.19%.

TABLE VI

DATA-DRIVEN ALGORITHMS : PREDICTION SCHEMES [RECURRENT NEURAL NETWORKS (RNNS), CONVOLUTIONAL NEURAL NETWORKS (CNNS), GRAPH NEURAL NETWORKS (GNNS)]

Nets	works	con(-)/pros(+)	Methods
RNNs	[204] 2017	The LSTM handles longer sequence **(+)** and its prediction is more accurate than other RNN variants **(−)**.	A stacked Long LSTM network for sequence classification of the TVs' states and predicts their future motion.
	[78] 2018	The LSTM variants to handle time-series data and predict motion over a long time range (5s of prediction horizon) **(+)**.	A two-layer LSTM for predicting acceleration distribution of TVs.
	[205] 2017	The LSTM is unable model interaction, and usually needs additional neural models to handle spatial information **(−)**.	Multiple LSTM to classify three manoeuvre classes.
	[203] 2019	The GRU comparing to LSTM needs fewer parameters for training and is less memory consuming **(+)**.	A GRU-based approach to design the pairwise interaction among the TVs and treat one SV at the time.
	[42] 2017	The GRU units provide comparable results to LSTMs with slightly faster performance [206] **(+)**.	An LSTM layer to estimate the target lane combined with another LSTM layer to predict the trajectory based on prior predictions.
	[169] 2018		An LSTM encoder-decoder to predict the occupancy probabilities of target and surrounding vehicles on a BEV grid.
	[119] 2018		A combination of LSTM and GMM to predict mixtures of Gaussian distribution.
	[179] 2018		An LSTM encoder-decoder: The LSTM encoder encodes information of observed vehicles as input sequences, then the encoder's hidden state is fed into an LSTM decoder to predict the manoeuvres.
	[80] 2019		A GRU-based approach that predicts the trajectory of 6 SVs.
	[120] 2020		A multi-layer LSTM combined with a fully connected neurons for predicting lane change manoeuvres.
	[32]2021		MFP jointly learns all agents history locations by using GRU encoder and the "world" context by encoding the road map via latent modes and predicts multiple futures.
CNNs	[142] 2018	Requires a massive amount of calculation [121] $(-)$; CNN captures the spatial interactions among road users and the traffic context by encoding the road scene $(+)$.	Use 2D-CNN with residual connections to calculate LiDAR features from the bounding boxes of the voxelized shape car and rasterized map [207] for detecting vehicles and predict future intention and thus trajectory.
	[31] 2017	DESIRE encodes the temporal and spatial information consistency $(+)$ by using 1D-CNN and 2D-CNN, respectively. It implements also a pooling layer to aggregate the social context, with a dense layer for the prediction.	A 2D-CNN encodes the spatial information of the road and two-layer RNN encoder-decoder with GRU for interpreting past positions of the road users in heterogeneous driving conditions. A CVAE is applied to generate explicit latent variables and refines the prediction outputs.
	[143] 2018	The LSTM and 2D-CNN with social-pooling layers is a good combination to capture temporal and spacial consistency of the TVs and SVs for prediction.	The algorithm implements the LSTM hidden states with a social-pooling layer and CNN to encode a (25×5)pixels DOGM and thus can capture different interactions between the vehicles.
	[51] 2021		The algorithm encodes semantic depiction of the driving scene around the TVs and performs a joint aggregation of higher-level features during the prediction. The memory-efficient and dense 3D tensor encodes the time, SVs locations, and their history states. The algorithm implements 2×2D-CNNs to jointly capture spatio-temporal feature from observed inputs.
	[142] 2018		A 2D-CNN based backbones and voxalized LiDAR data with DOGM encoding for a high-level representation of the road users' intention.
	[137] 2019		A combination of 2D-CNN and fully connected layers with GRU which learns from an explicit latent space and sets interactions among the vehicles. The prediction of the trajectory is performed on roundabouts and urban environment under mixed traffic conditions.
GNNs	[127] 2019	The GNN may not robust to noise in graph data - node-feature perturbation, edge addition/deletion $(-)$.	A GNN and LSTM based algorithm that models the interactions between the vehicles and VRUs.
	[201] 2021	The GCN may fail to distinguish between two graphs, especially in the classification stage$(-)$ - when input features are uniform.	LOKI implements GNNs for modelling the explicit interactions among the vehicles and VRUs. The MLP is trained to predict the intention. The CVAE and GRU are applied for trajectory inference of the surrounding road users.
	[189] 2020	GCN may fail to properly capture the interaction structure of graphs $(-)$.	The EGCN and LSTM are implemented set explicit interactions among the SVs and infer their trajectories on highways.
	[144] 2019	The GCN encodes explicit interaction among road users $(+)$.	The GCN and GAT are applied for modelling explicit interactions among vehicles on highways.
	[126] 2019	The GCN is a fully scalable network and can handle arbitrary number of entering/exiting vehicles in the road scene $(+)$.	A combination of CNNs and GNNs for setting up the interaction among the SVs. The LSTM encoder-decoder is also implemented to predict vehicle trajectories on highways.
	[43] 2021		The algorithm is based on GAN for modelling interactions among the vehicles and VRUs on roundabout under mixed traffic situation.
	[146] 2021		A two-level HGNN models the interactions the vehicles: a combination of LSTM and GCN for encoding spatial relationships among vehicles.
	[33] 2020		A 2D-CNN based algorithm that encodes the road semantic map. The LSTM and GNN are also combined for setting up intersections and predict the road users' behaviour in heterogeneous traffic conditions.
	[30] 2021		A GNN-based algorithm which models the interaction among the SVs and infrastructures. The algorithm also capture the spatio-temporal information consistency from the track history of the SVs by using the RNN and 2D-CNN in urban driving situations.
	[147] 2021		A combination of the GNN and RNN for predicting the SVs' trajectory on highways.

ii) Convolutional Neural Networks (CNNs): Convolutional Neural Networks (CNNs) can solve the problem of spatial representation of the sequential information encountered by LSTMs. These two prediction models are generally combined.

In [141], they applied a top-down or simplified Bird's-Eye View (BEV) images to encode the road context in the neighbourhood of the TVs and SVs. The extracted features are used to feed an LSTM encoder-decoder for time-series data learning and interpretation. The LSTM decoding block is fed to a deconvolutional layer to carry out image data, thus representing the road context. Lee *et al.* [31] combined GRU encoder-decoder and CNN to encode the road scene or context. In [142], a CNN is applied to encode LiDAR features, extracted from the bounding boxes of the voxelized shape car and rasterized map for predicting intention and trajectories of the SVs. In [143], they combined CNN and LSTM encoder-decoder to model social interaction between the SVs in a road scene around the TV, which will evolve in the following time-steps. Deo and Trivedi [67] combined 1D-CNN and 2D-CNN structures with the LSTM for encoding all dynamics states from the temporal trajectory of the observed SVs.

The CNN-based approaches can implicitly encode interactions between road users. However, CNNs are unable to explicitly extract and learn explicit social interactions between surrounding road users. Table. VI (CNNs) provides CNNs-based prediction methods.

iii) Graph Neural Networks (GNNs): Graph Neural Networks (GNNs) have become popular to model social interactions [6] among the surrounding road users observed in the road scene [33]. They express features each road user by graph nodes and spatial relationships between them by graph edges. Social interaction modelling by GNNs attracted great interest these past two years [126], [134], [144]. Diehl *et al.* [144] proven that by encoding the traffic scene with graphs it is possible to considerably increase the quality of the prediction on highly an interactive highway data. Mo *et al.* [145] proposed Recurrent Convolutional and Graph Neural Networks (ReCoG) uses GNNs for designing explicit interactions between SVs and infrastructures as a heterogeneous graph. ReCoG adopted GNNs for interaction features by focusing on one SV trajectory prediction in urban environment where geometry of the road significantly impacts the on-road behaviour of the SVs. Li *et al.* [126] and [134] modelled explicit interactions between the SVs to predict the on-road behaviour of the targeted ones. First, a graph convolution network (GCN) is applied to encode each of the SVs' features and their reciprocal interaction. The GCN aggregates and transforms the features of nodes through convolution operations and nodes' weight sharing via graph operations using adjacent matrix. GCN outputs are fed into an LSTM encoder-decoder to predict trajectory. In [126] and [134] interactions are also explicitly modelled between the SVs. However, precedent methods are deterministic with a single-shot prediction. Jo *et al.* [146] designed a Hierarchic GNN (HGNN) based method to predict interactive intentions among the SVs. Their procedure is in two levels: an intention-aware multimodal trajectory prediction and an interaction-aware trajectory prediction network design. The first level applies LSTMs and GCN encoder and their products are merged to supply LSTM decoder with a dense layer; the second level generates future trajectories of the SVs and setups interactions among them. It runs an LSTM encoder with graphs and mixed outputs. GCN encoder is given to an LSTM decoder for generating subsequent interactive trajectories of all observations. Mo *et al.* [147] combined CNNs, RNNs and GNNs to handle explicit interactions between varying number of vehicles. They used LSTM encoder-decoder for calculating dynamics features of all SVs and GNNs to design their social interactions. Their approach handle inter-vehicular consistency. LSTM decoder interprets mixtures of the TV dynamics and SV interactions to predict the on-road behaviour of the vehicles.

Table. VI provides social interaction approaches based on GNN network.

c) Output representations: This subsection categorises the reviewed studies based to their adopted output representation for on-road behaviour prediction. We consider three sub-classes:(1)-unimodal trajectory prediction;(2)-multimodal trajectory prediction; and (3)-Dynamic Occupancy Grid Map (DOGM).

i) Unimodal trajectory prediction: • **Unconditional trajectory prediction approaches.** These approaches describe the possible behaviours of other vehicles by considering their viable intentions as being decoupled from their future trajectories. The prediction can be performed in discrete or continuous spaces over a predefined and finite time-frame [42].

In the automated driving use case continuous spaces are much more intuitive because the real-world is continuous. On the other hand, discrete space makes the solution space compressed and much easier to solve. On the contrary, predicting the trajectory while neglecting the intention can provide more precise information on the future on-road behaviour over a short-term prediction horizon. However, it is often difficult to interpret the efficiency of the prediction because without manoeuvring, modelling or interpretation, it is difficult to evaluate lateral displacement errors. This does not have the same effect on the consistency of the prediction as a longitudinal prediction error, particularly in a motorway scenario.

In addition, given a definite driving scenario and the past states of a vehicle, it would be feasible for them to traverse different trajectories at once. Therefore, the corresponding distribution has multiple modes (or multiple futures) [32]. Unimodal trajectory predictors can only predict the maximum likelihood which is picked from all possible predicted trajectories. However, these approaches cannot cover all possible modes and thus wrongly assume that the most probable prediction is always correct and therefore simplifies the uncertainty [75].

• **Conditional trajectory prediction.** These approaches describe the future on-road behaviour of other vehicles by

[6]Social interaction is expressed at a higher level, as GNNs provide explicit interactions between surrounding road users, by modelling their interactions in a pattern form based on nodes and edges. Instead of just implicitly encoding the history states of the TVs and SVs in the surrounding traffic.

predicting their trajectory conditioned upon intention inference to yield a unimodal distribution of the prediction probabilities. They consider the highest probabilistic prediction outcomes among those of several predefined manoeuvre classes [142], [143] to estimate a vehicle trajectory. These approaches showed accurate recognition of the intended manoeuvres in simple driving scenario [32] and are mostly combined with improved version of the LSTM encoder-decoder to predict future trajectory. Nonetheless, these approaches lack precision in the prediction of the trajectory in a given driving scenario, and count two major limitations: (1)-they cannot accurately predict the trajectory of a vehicle if the vehicle intention has not predefined in the manoeuvre classes. This problem usually occurs in complex driving scenarios, as it is difficult to predetermine all possible intentions in such environments; and (2)-they apply supervised learning techniques, which require the intent classes to be labelled manually to train/test the model. This task is time consuming and prone to errors. Therefore, the prediction results can be biased.

ii) Multimodal trajectory prediction: Heretofore, the unimodal prediction presented above can be effective in predicting the on-road behaviour of vehicles over tighter prediction horizons. However, there will always be a possibility of converging towards an average of different modes of behaviour [140]. To overcome this limitation, more advanced methods predict multiple motion hypotheses by yielding "multiple future" [32], which can be implemented as a conditional probabilistic design with a multimodal distribution conditioned on some inputs [31], [125], [139], [139], [148]. Casas *et al.* [142] implemented intention prediction as an eight-class classification framework based on CNNs and voxelized LiDAR scan and DOGM of a high-level representation of the world. Predicting the trajectories of other vehicles is conditioned on intention prediction scores. Song *et al.* [35] proposed a hybrid model with generative models, which uses joint probability of lateral and longitudinal manoeuvres conditioned on the likelihood trajectory candidates. Other equally effective approaches [31], [125], [139], [139] have adopted the multimodality paradigm with intention encoding. Although their higher computational complexity reported, they have shown the ability to calculate the uncertainty during prediction, which makes them effective under complex driving conditions.

iii) Dynamic occupancy grid map (DOGM): These methods [143], [149] build a 2D/3D channel Dynamic Occupancy Grid Map (DOGM), including locations and lane marking of the observed vehicles. The 3rd channel consists of particular intentions [34]. Future occupied location cells can be predicted by combining CNNs and LSTM encoder-decoder [35], [135], [150]. However, the limitation of these methods is that the prediction efficiency is restricted to the size of the cells. And adding more cells will lead to an increase of the computational cost [35]. Nonetheless, DOGM solutions are effective to exact the vehicle spatial locations with its attributes and intended manoeuvres. This type of approaches can improve prediction performance in a data fusion-based prediction framework [34], [35]. DOGM provides a high-level representation of the traffic, while giving a high-level understanding of future on-road behaviour, often defined for specific scenarios. DOGM contains the probabilities of occupancy at future time-steps [143], which helps to predict multiple future modes but lacks accuracy for larger grid cells resulting in less consistent trajectories.

VI. Datasets

Datasets have an essential role in data-driven methods and conventional methods based on models. Nonetheless, data-driven approaches require extensive datasets for end-to-end training and also transfer learning or fine-tuning tasks, as well as for testing prior to their deployment on the road. As data-driven approaches have emerged, so have the training datasets to support them.

This section classifies the *datasets* into two main classes according to the Point Of View (POV): (a)-Top-down-view data; and (b)-Vehicle-view data.

The two classes can help to speed up the development process and enable interdisciplinary research to be carried out with the aim of better understanding the road behaviour and derived social interaction between road users in the shared space.

A. Top-Down-View Data

Has a aerial POV and represents the traffic objects (vehicles and VRUs if any) by a colour-coded bounding boxes. For instance, in Uber data [142] vehicles are sketched as a voxelized bounding box and it plots the location history of vehicles using bounding boxes with transparent linear gradient colour code of similar colours. Top-down-view data includes BEV data with a simple representation that uses polygons, lines and other geometrical designs in a BEV images describe in a basic yet powerful way of the traffic scene.

It retains objects proportion, their spatial coordinates and roadway geometry. However, BEV representation neglects high-level representation of the traffic scene (such as textures, lighting, colours, reflection). Among the existing publicly accessible datasets, some of them provide a top-down view of the driving environment with different road environments (highway and urban).

For instance, on highways, few publicly available datasets are widely applied in the state of the art, among these datasets, Next Generation SIMulation (NGSIM) [49], HighD [50], Tongji Road Trajectory (TJRD) [151]. Whereas, in urban or street driving environment, few datasets are largely applied for motion prediction in automated driving context, among these datasets, INTERACTION [152], InD [153] and roundabout [154].

Top-down-view traffic data is useful for developing and testing original approaches because it provides a less complex representation of traffic while preserving the naturalistic driving aspect. This can match applications with quite less computational load and constraints. However, this representation also has some shortcomings. Unrealistic observations can be generated because obstruction in the Field Of Vision (FoV) of the vehicle cannot be taking into account. Top-down-view data can include duplicated objects (a vehicles) at overlapping locations and noise and thus generating measurement errors.

Therefore, failures can be expected in predicting trajectories since data is biased and contains measurement errors. For example, the NGSIM data contains perception errors, as stated by [155]. Thus, two sources of cumulative errors may affect the prediction efficiency. This is to say that perfect predictions would always produce positive error values [104].

B. Vehicle-View Data

Gives a higher level representation of the collected sensory data, and thus incorporates heterogeneous contextual contents of the driving environment. Vehicle-view data is of higher dimension than the top-down-view data, with a massive computing resources and terabytes of storage requirements.

Several vehicle-view datasets are publicly available and completely free of charge, such as Ford Autonomous Vehicles dataset [156], CityScape3d dataset [157], nuScenes dataset [158], KITTI dataset [159], [160], [161], [162], TRAF [163]. These sets of data have been collected in different road environments, at different times of the day (e.g., nuScenes dataset [158]) and under various weather conditions [164]. Vehicle-view datasets can have access to richer information of the surrounding which makes them useful to test algorithms under practical conditions of the real-world, thus giving a better evaluation of the algorithm performance-wise. However, computational burden will follow for processing high dimension input data, which is impractical for embedded implementation in the AV. Nonetheless, new computing techniques, such as parallel computing for resource sharing or the graphics processing unit (GPU) have made the processing of big data much easier and are quite useful tools for training complex models that merge several data sources.

Overall, vehicle-view data and top-down-view data are of two different levels of abstraction in the AV data collection chain but yet complementary to each other. Top-down-view data is more convenient to represent "the world" including static and dynamic objects in the driving scene. It is also more useful to develop theories and algorithms using all the information available. Vehicle-view data is, however, useful for exploring raw sensory knowledge with learning models that often rely on millions of parameters [165], allowing multiple sensors to be used simultaneously to achieve a certain goal of prediction.

VII. Evaluation Metrics

Principally, this section presents the most applied evaluation metrics in literature to sum up the performance evaluation of the trajectory and intention predictions. Next, research gaps are identified and discussed. Lastly, future trends are presented. As previously stated, evaluation metrics of intention prediction is often contemplated as a classification problem, while the task of predicting the trajectory is perceived a regression problem. The performance efficiency of these two tasks is summarised separately using a set of metrics briefly described in Table. VII.

1) **Trajectory prediction metrics:**
   - **Distance-based metrics.** These metrics are often applied in the literature to evaluate the reliability of the trajectory prediction. These include Average Distance Error (ADE) and Final Distance Error (FDE), which compute the Euclidean distance between the predicted trajectories and ground truth data. ADE and FDE can evaluate multimodal prediction approaches when the average (avg) or minimum (min) values are calculated over K trajectory samples. For instance, minADE is the minimum average Euclidean error between the predicted trajectory and ground truth trajectory of targeted agents, while considering the top K predictions. Similarly, minFDE corresponds to the minimum Euclidean error between the predicted endpoint and the ground-truth endpoint. Park *et al.* [166] applied the ratio of avgFDE-to-minFDE (RF) metric for the performance-wise evaluation stage; RF is defined as avgFED/minFDE, which differentiate between prediction with multiple mode and a single mode. These metrics are applied for multimodal trajectory predictors.
   - **Root Mean Squared Error (RMSE).** RMSE measures the standard deviation of the residuals (predicted trajectories) which is a measure of how far from the regression line are the data points which are ground truth trajectory data points in this case. This measure shows the density of the residuals around the best fitted prediction/trajectory line.
   - **Cross-Entropy or Negative Log Likelihood (NLL).** For the prediction data distribution and the ground truth data distribution, the NLL can be applied for both prediction tasks, namely intention prediction and trajectory prediction. To a certain extent, NLL can be more effective in assessing multimodal prediction than RMSE or ADE metrics. RMSE can be biased in favour of models that predict the average of modes and thus it can prone be to large errors due to the usage of squared error in its formula, as noted by [104]. Nonetheless, NLL can penalise a multimodal prediction method as it is unable of covering all the prediction modes (i.e., a mixture of a finite number of Gaussian distributions).
   - **Kernel Density Estimate-based Negative Log Likelihood (KDE-NLL).** ADE and FDE metrics are effective for assessing deterministic prediction outcomes, but they are constrained in comparing mixture of a finite number of Gaussian distributions produced by generative models, hence neglecting multimodality aspects. In [167], they introduce KDE-NLL, which at each timestep has sampled trajectories from a probability density function (PDF) at a specific timestep. Later determining the NLL of the ground-truth trajectory and apply off-the-shelf KDE which estimates its own bandwidth for each mode separately.

2) **Intention prediction metrics:**
   - **Accuracy Score.** Existing studies often measure reliability of the intention prediction with accuracy

TABLE VII

EVALUATION METRICS APPLIED IN THE LITERATURE TO EVALUATE THE PERFORMANCE OF INTENTION (INT) AND TRAJECTORY (TRAJ) PREDICTION

Evaluation metrics	Description	Formula	Traj	Int
Accuracy Score	Detects the prediction precision: higher the scoring value is, the better is the prediction effect. This metric is defined as the total number of correctly classified data samples, divided by the total number of data samples.	–	×	✓
Confusion Matrix (CM)	Measures class correspondence and performance visualisation in supervised/ unsupervised learning; in unsupervised learning, it is generally called matching matrix, and used for the same purpose. Derivations from a CM; True Position (TP), False Positive (FP), True Negative (TN), an False Negative (FN).	–	✓	✓
Precision (Pre)	It is often desirable when the manoeuvre classification model achieves higher values of the precision and the recall, simultaneously.	$\text{Prec} = \frac{\text{TP}}{(\text{TP+FP})}$	×	✓
Recall (Rec)		$\text{Rec} = \frac{TP}{(\text{TP+FN})}$	×	✓
F1−Score	A harmonic mean of the precision and recall for further comparison of the classification results	$\text{F1} = 2 \times \frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}}$	×	✓
Receiver Operating Characteristic (ROC)	Assess the quality of the learning algorithms, by calculating the AUC, which shows the probability of correct discrimination between the manoeuvre classes.	–	×	✓
Area Under Curve (AUC)	AUC is calculated from the ROC−curve by using the $5-$ and $10-$*fold* cross-validation techniques.	–	×	✓
Average Inference Time	Calculates over a predefined sliding window of the input data series for predicting the subsequent manoeuvre of interest. This metric is calculated by averaging the prediction time of the first class of manoeuvre of interest predicted, over all data samples.	–	×	✓
Final Distance Error (FDE)	Calculates the $L2$ norm between the predicted final location $\hat{y}_{t_{\text{end}}}$ and the final location ground truth $y_{t_{\text{end}}}$. t_{end} is the end of the prediction horizon. But it does not calculate the prediction error in the prior time steps on the prediction horizon.	$\text{FDE} = \lvert\hat{y}_{t_{\text{end}}} - y_{t_{\text{end}}}\rvert$	✓	×
Cross-Entropy (a.k.a NLL)	ϕ and ψ The trajectory distribution and the ground truth data distribution, respectively. This metric can be applied in manoeuvre and trajectory prediction tasks.	$H(\phi, \psi) = \mathop{\mathbb{E}}_{x \sim \psi} -log(\psi(x))$	✓	✓
Summarised Cross-Entropy [138]	[138] proposes an enhanced form of the cross-entropy metric for an appropriated evaluation of multi-modal trajectory predictors; $\overline{\psi}$ is an approximation of ψ	$H(\phi, \psi) + H(\phi, \overline{\psi}) = \mathop{\mathbb{E}}_{x \sim \psi} - log(\phi(x)) + \mathop{\mathbb{E}}_{x \sim \psi} - log(\overline{\psi}(x))$.	✓	×
Minimum of K Metric [92, 208, 31]	Evaluates multi-modal predictors with K predicted trajectories for different parameters. Previous metrics are also applicable, but they are valid when evaluating one of the K predicted trajectories (the likelihood trajectory). These metrics cannot however qualitatively assess K prediction outcomes and mostly $K-1$ trajectories can not necessarily be examined during evaluation. This metric is extensively discussed in [138].	–	✓	×
MAE/RMSE	MAE averages the magnitude of the prediction error/ displacement error ϵ_t. RMSE calculates the square root of the average of ϵ_t, they are computed over a predefined time window t of the prediction horizon; where n is the number of samples and ϵ_t computed upon the residuals-differences between observed/ ground truth and predicted values of data. RMSE and MAE are quite similar and applied in regression-based approaches for trajectory generation.	$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\epsilon_t^2}$ $\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}\lvert\epsilon_t\rvert$	✓	×
Average Distance Error (ADE) [92]	The predicted data samples $\hat{y}$ and the ground-truth data samples y are the predicted and ground truth positions, respectively. *minADE-N* and *minFDE-N* error metric introduced in [92] are applied for multi-modal assessment. The metric is the minimum ADE and FDE out of N future trajectories predicted at test-time	$\text{ADE} = \frac{\sum_{j=t_{ob}+1}^{t_f} \lvert\lvert\hat{y}_j - y_j\rvert\rvert_2}{(t_f - t_{ob})}$	✓	×
Best-of-N (BoN) [33]	Computes the minumum ADE and FDE from N randomly-sampled trajectories.	–	✓	×
KDE-NLL [33, 168, 209]	Computes the mean log-likelihood of the ground truth trajectory under a distribution created by fitting a kernel density estimate on trajectory samples.	–	✓	×

TABLE VIII
DEFINITION OF THE MAIN ACRONYMS

Acronym	Description
AV	Automated Vehicle
ADAS	Advanced Driver Assistance System
ADS	Automated Driving System
ML	Machine Learning
CAV	Connected Automated Vehicles
AAA	Australian Automobile Association
NHTSA	National Highway Traffic Safety Administration
CaDMV	California Department of Motor Vehicles
DoT	Department of Transportation
SAE	Society of Automotive Engineers
EV	Ego-Vehicle
SV	Surrounding Vehicle
TV	Target Vehicle
VRU	Vulnerable Road User
UAV	Unmanned Aerial Vehicle
LK	Lane Keeping
LC	lane Change
LCR	Lane Change Right
LCL	Lane Change Left
TL	Turn Left
TR	Turn Right
GS	Go Straight
EKF	Extended Kalman Filter
MLP	Multilayer Perceptron
AUC	Area Under Curve
SVM	Support Vector Machine
RMSE	Root Mean Square Error
DBN	Dynamic Bayesian Network
POMDP	Partial Observable Markov Decision Process
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
GNN	Graph Neural Network
TTLC	Time To Lane-Change
QDA	Quadratic Discriminant Analysis
S-LSTM	Social LSTM
ST-LSTM	Spatio-Temporal LSTM
CS-LSTM	Convolution-Social LSTM
GAN	Generative Adversarial Network
MPC	Model Predictive Control
HMM	Hidden Markov Model
KF	Kalman Filter
EBiLSTM	Enhanced Bidirectional LSTM
RMIN	Recurrent Metainduction Network
RKHS	Reproducing Kernel Hilbert Space
BNN	Bayesian Neural Network
DOGM	Dynamic Occupancy Grid Map
CGNS	Conditional Generative Neural System
CVAE	Conditional Variational Autoencoder
CGAN	Conditional Generative Adversarial Network
GRU	Gated Recurrent Unit
GCN	Graph Convolution Network
HGNN	Hierarchic GNN
POV	Point Of View
BEV	Bird's-Eye View
NGSIM	Next Generation SIMulation
TJRD	Tongji Road Trajectory
FoV	Field of Vision
GPU	Graphics Processing Unit
NLL	Negative Log Likelihood
ADE	Average Distance Error
FDE	Final Distance Error
avgFED	average FED
minFDE	minimum FDE
V2X	Vehicle-to-Everything
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
MCM	Manoeuvre Coordination Message
MCS	Manoeuvre Coordination Service
VGMM	Variational Gaussian Mixture Model

score. The total of correctly classified ground-truth manoeuvre samples divided by the sum of all data samples describes this metric. Nonetheless, referring only on the accuracy score to interpret the prediction outcomes can be skewed, especially when the data is unbalanced. For instance, the AV operating on freeways, the longitudinal manoeuvre classes, such as LK, can often be of high significance of classes compared to other manoeuvres, such as LC class. Therefore, LK can show higher accuracy than other manoeuvre classes. Further prediction performance indicators are often considered to extend performance-wise evaluation of the predictors in understanding their effectiveness better. These metrics include confusion matrix, precision, recall and F1-Score. Note that for a given manoeuvre class, a low precision reveals a high ratio of false alarms. While a low recall indicates many ground-truth samples for a given manoeuvre class are incorrectly classified/predicted among the other classes. F1 is the harmonic mean of the precision and recall.

- **Average Inference Time.** In the interest of a given driving scenario, this metric is applied on a time-based sliding window (or time-frame) of a ground-truth time-series data for predicting correct occurrences of the predicted manoeuvres. This metric is calculated by averaging the instance of the first correctly predicted LC manoeuvre (start-point or origin) and its last predicted occurrence (end-point) within respective time-frames applied for all data samples.
- **Negative Log Likelihood (NLL).** This metric is effective for evaluating the performance of predicted manoeuvres as a multi-class classification task.

Table. VII presents evaluation metrics that are commonly used for intention and trajectory prediction in existing studies.

A. Research Gaps and New Trends

a) It is often desired to have a robust motion prediction scheme that is adaptable and can reliably operate on different driving scenarios, like freeways or urban scenarios. Most existing studies either concentrate on specific scenarios [168], [169] or train and apply their techniques in data from diverse scenarios without appraising their scenario-wise performance [33], [134], [170], [171] As a result, these approaches would have a performance decline when transferred to new scenarios of unseen model parameters and situations. These approaches may fall short in adapting the trajectory time frames to be instinctive manoeuvring behaviour rather than structured output proposition. They also may abstain from the primary motive to lower the road hazards which usually occur due to unseen model parameters. Several factors lead to such lack of flexible predictability, while the graphical working interphase of the AV system limited to physical parameters can be one of the foremost predicaments. The input features of these methods are usually in Cartesian coordinate frame and/or pixel-wise depiction of the road-map, which may lead to some drawbacks, such as scenario-specific information, including traffic signals and geometric roadway design, are insufficiently integrated, if not maybe bypassed. And these representations are not constant and will change each time it encountered a new scenario. These regulatory or demographic variables have unpredicted significance in the manoeuvre prediction which needs to be taken into account. Quite a few studies have tried to address these limitations. One recent work [172] achieves transferable prediction by designing generic depictions called Semantic Graph, where Dynamic Insertion Areas, instead of road entities, are regarded as the graph nodes. In such representations, scenario specific information, such as roadway geometry and traffic signals, is effectively and comprehensively aggregated. However, their approach only considers the intention prediction sub-task and requires further trajectory to decode from the downstream module for heuristic utilisation.

b) So far, reviewed studies assume full observability of the traffic environment with complete access to the past states of objects (e.g., a vehicle). However, this would not be practicable in many cases due to the unexpected events that the vehicle cannot deal with. Solutions based on the vehicle past states only are sometimes inaccurate and show early limitations when operating under different conditions. Therefore, behaviour prediction should always consider sensor deficiencies that may limit access to relevant environmental information, such as the number of observable vehicles around. Top-down view of environment traffic throughout infrastructures, like mounted sensors (e.g., CCTV or antennas) and roadside units is a relevant solution to access and use unseen and occluded traffic information. Yet, covering all road scenes with these sensors is unrealistic. AVs can sense their environment to predict other vehicles' behaviours. In fact, sensory information may not always be available in all situations. Thus, V2X (vehicle-to-everything) communications is used to provide this missed data through CAVs. V2I (vehicle-to-infrastructure) and V2V (vehicle-to-vehicle) protocols are utilised by CAVs, in combination with conventional sensors to detect the world. CAVs use cooperative manoeuvres to coordinate their future trajectories for efficient navigation. Current efforts define cooperative manoeuvres on a distributed approach where vehicles use V2V networks to exchange wirelessly information about their future intentions. Manoeuvre coordination (or

cooperative manoeuvring) may assist the vehicles to quickly adapt their driving based on the dynamics of surrounding objects, avoid confusions about driving intentions, and facilitate the coordination of manoeuvres with other cooperative vehicles [173]. Recently, [174] extended the V2V communication concept by adding the possibility for the infrastructure to support cooperative manoeuvring using V2I communications. They have also investigated the effect of V2X-based manoeuvre coordinates on the traffic. In [175], they proposed a manoeuvre coordination message (MCM) that can be used in cooperative manoeuvres context with/without road infrastructure support. Their approach does not imply infrastructures to coordinate the vehicle manoeuvres. However, it can instead provide complementary traffic information or notifications that vehicles can use to coordinate their manoeuvres (e.g., speed advice for a smoother coordination of the manoeuvres). Their approach requires all CAVs a constant broadcast of MCM with their planned trajectories, and thus CAVs can detect the need to coordinate a manoeuvre without having to predict future trajectories of other cooperative vehicles. In highly interactive driving situations, the CAV requires coordinate manoeuvres of multiple vehicles via V2V distributed process, which needs a pairwise and sequential coordination of other vehicles' intentions. By increasing the number of surrounding vehicles, this process will be time consuming to coordinate all surrounding vehicles, and hence can affect the traffic flow and safety. However, in mixed traffic scenarios where conventional, connected and AV coexist, the infrastructure-based manoeuvre coordination processes would be less reliable and, in some cases, not available. In such situations, [174] proposed to use a fusion-based approaches, which combines information about the driving condition provided by the cooperative awareness message (CAM) and information from car on-board sensors [176].

To the best of our knowledge, not much works developed in a cooperative manoeuvre context have been integrating the on-board sensors and CAM information into a multi-agent learning prediction mechanism. This can provide a reliable information to predict the intention and thus improve the AV decision-making and its understanding of the vehicles' interactions.

c) Few studies have focused on emphasising the qualitative performance-wise of the predictor, by examining two aspects essential to better understand the vehicle on-road behaviour, covering a decoupled analysis of the standard trajectory prediction error, and its physical realism. In the following we will:(i)-attempt to discuss the analysis of standard trajectory prediction errors (e.g., RMSE/MAE); (ii)-examine few studies that discuss the physical feasibility of trajectory prediction on a multi-agent basis.

i) Standard trajectory prediction errors like the RMSE do not capture physical realism of the prediction outcomes due to the presence of unobserved measures of deviations which can lead to a skewed estimate of the error, and thus cannot paint an out-and-out factual picture of the vehicle predicted behaviour under practical operation conditions. In fact, the vehicle's behaviour in the lateral and longitudinal directions reveal discrete interpretation of the prediction errors and thus a very specific explanation can be given about the vehicle's behaviour during operation in each direction. In particular, the traditional interpretability of prediction error in a highway driving scenario can be high longitudinally or be relaxed to several metres. Whereas the lateral requirement become more stringent or is overly constrained and more critical. However, essentially the connector ramps must be found within a reasonable tolerance as a factor of safety on highways. So, there is a more stringent longitudinal requirement based on vehicle operation, which shows an unconventional requisite. This predictability is constrained by distinctive driving situations where vehicle may operate in, and by other key components like road geometry standards (lane width and curvature) and permissible vehicle dimensions. Therefore, the prediction performance analysis solely providing a multi-horizon Gaussian ellipsoidal approximation scheme can be less intuitive and at sometimes skewed when applied to different scenarios. For example, let us contemplate a case study scenario where the approach has achieved an average RMSE of 3m over a prediction horizon of 4s on freeway operation during the inference stage. This case study can be interpreted in two divisions. First, reiterating the performance-wise of the vehicle behaviour prediction achieved an RMSE of 3m over a rather extended prediction horizon of 4s. Second, the system is effective to hold a low value of an average RMSE over a long term. This result can compete with those of recent studies (e.g., [42], [67], [148], [177], [178]) and even have an edge with high workability in a highway driving scenario. Nevertheless, in terms of risk assessment, an RMSE close to 3m on a lateral direction is critical, which means that the vehicle is barely maintained in the lane and its future locations might result off-road when considering a US highway with a wide of 3.6m. So, reporting the average RMSE during tests cannot be informative about the lateral (side-by-side) and longitudinal

(forward-backward) behaviours of the vehicle during typical operation. The average RMSE calculated upon the Euclidean distance would absorb the influence of lateral RMSE as it is of an order of magnitude smaller than that of the longitudinal RMSE. Henceforth, it is convective to perceive that the RMSE measure cannot clearly validate the observed variability and this metric can misrepresent the actuality due to its nature of skewing the centrality of the data points under the influence of anomalies and deviance of two externalities of the data in this context. The aforementioned constraints showed fundamental factors determinant of the error values, where the prediction error bounds in the lateral and longitudinal direction can guarantee a realistic predicted behaviour when operating in different operational scenarios, where the aim is to maintain knowledge that the vehicle is within its lane over different horizons of prediction and to determine the physical realism of the prediction. Recently [179] studied safety-critical location systems by developing an in-depth analysis of the prediction requirements for AVs. Their study yielded prediction constraints called protection level, which restrict possible prediction errors within a safe mode by considering different parameters of the road (e.g., the road width and curvature), and those of the vehicles (e.g., the vehicle length and width) to ensure realistic and allowable prediction of bounding boxes on diverse operating scenarios. We believe this is the first work that examines the localisation demands of the AV level3+ and determines a relevant integrity risk study at the localisation level. Their study gives insights into prediction error constraints, which could be useful for vehicle behaviour prediction.

ii) Maintaining a safety critical dynamic of the predicted vehicle behaviour over a long period of time is essential to achieve a tangible realistic prediction. In this context, multi-agent learning systems have become prevalent to learn vehicle dynamics (feasible steering and acceleration) from multiple trajectory data available during training. However, the performance-wise of these methods, can be skewed and tend to violate the dynamic feasibility of the long-term predicted trajectories. Although being a powerful toolkit, pure DNN based methods provide insubstantial theoretical testament regarding the physical dynamics and feasibility of the predicted intents and trajectories. These methods only predict the centre positions of a vehicle or independently predict position and heading which may still result in infeasible discrepancies between the two, and as a result may poorly capture the vehicle's occupancy in the road space. Several approaches have attempted to address the physical feasibility limitation of conventional multi-agent learning systems. They built goal-conditioned multi-agent forecasting systems, which are more effective for longer horizon forecasting task [31], [125]. These approaches adopt an endpoint conditioned prediction strategy, which condition the motion forecasting task on goal destinations. Likewise, [39] introduced trajectory anchors to guarantee a more structured prediction, but without validating on a dynamic feasibility of the predicted trajectory because of arbitrary predicted offsets. Hybrid predictors - mix the DNNs and structured robotics techniques - have been very recently developed [180], [181], [182], [183] with the aim to validate the physical realism of predicted intents and trajectories. These approaches leverage algorithms based on present goals and can provide a more structured prediction of the long-term behaviour.

VIII. Conclusion

Predominantly, this review paper aims to contribute to knowledge around three key facets encircling the significance and need for an accurate motion prediction framework for AVs. First, the salient innovations made in motion prediction and on-road behaviour of the surrounding traffic in high-level automation (SAE level 3+) is investigated. Second, a distinct classification and definition of the technical terminology used in the state-of-the-art to describe motion prediction is formulated. Third, the contemporary model-based prediction and data-driven solutions around other road users' motion prediction is reviewed. The performance of model-based prediction frameworks has shown effectiveness over a constrained prediction horizon that is less than 1 second. However, this prediction range of these frameworks are too narrow to allow the AV in anticipating potential hazards, or even seem to be unrealistically inadequate time duration for the human driver to regain control of the vehicle. At SAE level 3 (conditional automation), human drivers must be prepared to regain control of the vehicle in the event of a system failure in emergency situations. For instance, the recent event of Tesla's AV stopping at a traffic light, suddenly sped forward hitting an ongoing cyclist who later died. In such a situation, human-driver should have enough time to trigger system takeover request and hereby regain the vehicle control. Therefore, reliable prediction over a longer prediction horizon is highly required. Data-driven motion prediction approaches have sparked significant innovation in recent years and have shown promising efficiency, especially in complex and highly interactive driving situations. These approaches are escalating rapidly and actively engaging researchers currently. Most data-driven prediction approaches rely on deep learning techniques and have drawn heavily on multi-agent interaction and Social Robots paradigms to provide prediction with a social consistency. Interaction is a driving factor of how a

machine can learn various encounters in the environment around it. Interaction sometimes can be very case specific and has high level of ambiguity and uncertainties. This high endogeneity needs to be predicted with minimum error to avoid hazardous and unforeseen events. However, model-based methods cannot effectively predict this interaction, as of today there is a shortage in an efficient model predicting social interactions. While the approaches of modelling social interactions are inherently different with fluctuating prediction performance. For instance, approaches that implicitly design social interactions are at a low-level of interactions, with lower performance than approaches that explicitly design social interactions, generally based on graph theory, are defined at a higher level of interactions. These social interaction approaches (low-level and high-level interactions), besides considering interactions and operating under heterogeneous traffic conditions, other equally important factors, like the road context (traffic rules and environment conditions), can positively affect the prediction response over a long prediction horizon that generally varies between 5s and 10s. Nonetheless, the real-world constraints, such as sensor impairments, limited observability of road scene and other uncertainties have been partially dealt with by the existing approaches and is still a serious concern in motion prediction task.

REFERENCES

[1] *Global Status Report on Road Safety 2018: Summary*, World Health Org., Geneva, Switzerland, 2018.

[2] M. Bradley, "Cost of road trauma in Australia," Austral. Automobile Assoc., Canberra, ACT, Australia, Summary Rep., Sep. 2017.

[3] "Queensland road crash weekly report," Transp. Main Roads, Brisbane, QLD, Australia, Tech. Rep. 1230, 2021. [Online]. Available: https://www.tmr.qld.gov.au

[4] *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*, vol. 2, Nat. Highway Traffic Safety Admin., Washington, DC, USA, 2015, pp. 1–2.

[5] K. Merfeld, M.-P. Wilhelms, and S. Henkel, "Being driven autonomously—A qualitative study to elicit consumers' overarching motivational structures," *Transp. Res. C, Emerg. Technol.*, vol. 107, pp. 229–247, Oct. 2019, doi: 10.1016/j.trc.2019.08.007.

[6] P. Wu, "A mixed methods approach to technology acceptance research," *J. Assoc. Inf. Syst.*, vol. 13, no. 3, pp. 1–16, 2012.

[7] *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, Standard J3016_201806, SAE Standard, 2018.

[8] I. Nastjuk, B. Herrenkind, M. Marrone, A. B. Brendel, and L. M. Kolbe, "What drives the acceptance of autonomous driving? An investigation of acceptance factors from an end-user's perspective," *Technol. Forecasting Social Change*, vol. 161, Dec. 2020, Art. no. 120319, doi: 10.1016/j.techfore.2020.120319.

[9] S. A. Beiker, "Legal aspects of autonomous driving," *Santa Clara Law Rev.*, vol. 52, p. 1145, Jan. 2012.

[10] R. Shanke *et al.*, "Autonomous cars: Self-driving the new auto industry paradigm," Morgan Stanley, New York, NY, USA, Morgan Stanley Res.'s Blue Paper, 2013.

[11] M. Bertoncello and D. Wee, "Ten ways autonomous driving could redefine the automotive world," McKinsey Company, New York, NY, USA, Jun. 2015. [Online]. Available: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/ten-ways-autonomous-driving-could-redefine-the-automotive-world

[12] A. Eugensson, M. Brännström, D. Frasher, M. Rothoff, S. Solyom, and A. Robertsson, "Environmental, safety legal and societal implications of autonomous driving systems," in *Proc. Int. Tech. Conf. Enhanced Saf. Vehicles (ESV)*. Seoul, South Korea, vol. 334, 2013, pp. 1–15.

[13] R. Shabanpour, S. N. D. Mousavi, N. Golshani, J. Auld, and A. Mohammadian, "Consumer preferences of electric and automated vehicles," in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2017, pp. 716–720, doi: 10.1109/MTITS.2017.8005606.

[14] S. Dibaj, J. P. Zmud, A. Golroo, I. N. Sener, M. Habibian, and M. Hasani, "Towards an understanding of the travel behavior impact of autonomous vehicles," *Transp. Res. Proc.*, vol. 25, no. 1, pp. 2504–2523, 2017, doi: 10.1016/j.trpro.2017.05.281.

[15] A. Brown, J. Gonder, and B. Repac, "An analysis of possible energy impacts of automated vehicle," in *Road Vehicle Automation*, G. Meyer and S. Beiker, Eds. Berlin, Germany: Springer, 2014, pp. 137–153.

[16] F. Favarò, S. Eurich, and N. Nader, "Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations," *Accident Anal. Prevention*, vol. 110, pp. 136–148, Dec. 2018, doi: 10.1016/j.aap.2017.11.001.

[17] D. Gruyer, O. Orfila, S. Glaser, A. Hedhli, N. Hautière, and A. Rakotonirainy, "Are connected and automated vehicles the silver bullet for future transportation challenges? Benefits and weaknesses on safety, consumption, and traffic congestion," *Frontiers Sustain. Cities*, vol. 2, p. 63, Jan. 2021.

[18] B. Brown, "The social life of autonomous cars," *Computer*, vol. 50, no. 2, pp. 92–96, Feb. 2017, doi: 10.1109/MC.2017.59.

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

[20] A. Boubezoul, A. Koita, and D. Daucher, "Vehicle trajectories classification using support vectors machines for failure trajectory prediction," in *Proc. Int. Conf. Adv. Comput. Tools Eng. Appl.*, Jul. 2009, pp. 486–491, doi: 10.1109/ACTEA.2009.5227873.

[21] 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, doi: 10.1109/ACCESS.2020.2982170.

[22] F. Wirthmüller, M. Klimke, J. Schlechtriemen, J. Hipp, and M. Reichert, "Predicting the time until a vehicle changes the lane using LSTM-based recurrent neural networks," *IEEE Robot. Autom. Lett.*, vol. 6, no. 2, pp. 2357–2364, Apr. 2021, doi: 10.1109/LRA.2021.3058930.

[23] F. Wirthmüller, J. Schlechtriemen, J. Hipp, and M. Reichert, "Teaching vehicles to anticipate: A systematic study on probabilistic behavior prediction using large data sets," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 1–16, Nov. 2020, doi: 10.1109/TITS.2020.3002070.

[24] N. Oliver and A. Pentland, "Graphical models for driver behavior recognition in a smartcar," in *Proc. IEEE Intell. Vehicles Symp.*, Oct. 2000, pp. 7–12, doi: 10.1109/IVS.2000.898310.

[25] R. J. Rummel, "Perceiving and behaving," in *Understanding Conflict and War: The Conflict Helix*. Beverly Hills, CA, USA: Sage, 1976.

[26] T. M. Newcomb, R. H. Turner, and P. E. Converse, *Social Psychology: The Study of Human Interaction*. New York, NY, USA: Holt, Reinhart and Winston, 1965, p. 149.

[27] W. Ju, "The design of implicit interactions," *Synth. Lectures Hum.-Centered Informat.*, vol. 8, no. 2, pp. 1–93, 2015.

[28] 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.

[29] H. R. M. Pelikan, "Why autonomous driving is so hard: The social dimension of traffic," in *Proc. Companion ACM/IEEE Int. Conf. Human-Robot Interact.*, New York, NY, USA, Mar. 2021, pp. 81–85, doi: 10.1145/3434074.3447133.

[30] X. Mo, Y. Xing, and C. Lv, "Heterogeneous edge-enhanced graph attention network for multi-agent trajectory prediction," 2021, *arXiv:2106.07161*.

[31] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr, and M. Chandraker, "DESIRE: Distant future prediction in dynamic scenes with interacting agents," 2017, *arXiv:1704.04394*.

[32] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Attention based vehicle trajectory prediction," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 1, pp. 175–185, May 2021.

[33] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision*. Cham, Switzerland: Springer, 2021.

[34] M. Henaff, A. Canziani, and Y. LeCun, "Model-predictive policy learning with uncertainty regularization for driving in dense traffic," 2019, *arXiv:1901.02705*.

[35] H. Song, W. Ding, Y. Chen, S. Shen, M. Yu Wang, and Q. Chen, "PiP: Planning-informed trajectory prediction for autonomous driving," 2020, *arXiv:2003.11476*.

[36] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Driver behavior classification at intersections and validation on large naturalistic data set," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 724–736, Jun. 2012.

[37] T. Gindele, S. Brechtel, and R. Dillmann, "A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.*, Nov. 2010, pp. 1625–1631, doi: 10.1109/ITSC.2010.5625262.

[38] I. Dagli and D. Reichardt, "Motivation-based approach to behavior prediction," in *Proc. Intell. Vehicle Symp.*, vol. 1, Jun. 2002, pp. 227–233, doi: 10.1109/IVS.2002.1187956.

[39] B. Varadarajan *et al.*, "MultiPath++: Efficient information fusion and trajectory aggregation for behavior prediction," 2021, *arXiv:2111.14973*.

[40] R. Roriz, A. Campos, S. Pinto, and T. Gomes, "DIOR: A hardware-assisted weather denoising solution for LiDAR point clouds," *IEEE Sensors J.*, vol. 22, no. 2, pp. 1621–1628, Jan. 2021, doi: 10.1109/JSEN.2021.3133873.

[41] S. Mozaffari, O. Y. Al-Jarrah, M. Dianati, P. Jennings, and A. Mouzakitis, "Deep learning-based vehicle behavior prediction for autonomous driving applications: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 1–15, Jan. 2020, doi: 10.1109/TITS.2020.3012034.

[42] F. Altché and A. de La Fortelle, "An LSTM network for highway trajectory prediction," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 353–359, doi: 10.1109/ITSC.2017.8317913.

[43] S. Carrasco, D. Fernández Llorca, and M. Ángel Sotelo, "SCOUT: Socially-COnsistent and UndersTandable graph attention network for trajectory prediction of vehicles and VRUs," 2021, *arXiv:2102.06361*.

[44] Y. Charlie Tang and R. Salakhutdinov, "Multiple futures prediction," 2019, *arXiv:1911.00997*.

[45] A. S. Mohammed, A. Amamou, F. K. Ayevide, S. Kelouwani, K. Agbossou, and N. Zioui, "The perception system of intelligent ground vehicles in all weather conditions: A systematic literature review," *Sensors*, vol. 20, no. 22, p. 6532, 2020, doi: 10.3390/s20226532.

[46] G. Adaimi and A. Alahi. (2018). *Learning Nuisances to Track Pedestrians in Autonomous Vehicles*. [Online]. Available: http://infoscience.epfl.ch/record/255623

[47] Y. Li, P. Duthon, M. Colomb, and J. Ibanez-Guzman, "What happens for a ToF LiDAR in fog?" *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 6670–6681, Nov. 2021, doi: 10.1109/TITS.2020.2998077.

[48] R. Heinzler, F. Piewak, P. Schindler, and W. Stork, "CNN-based LiDAR point cloud de-noising in adverse weather," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2514–2521, Apr. 2020, doi: 10.1109/LRA.2020.2972865.

[49] J. Halkias and J. Colyar, "NGSIM interstate 80 freeway dataset," U.S. Federal Highway Admin., Washington, DC, USA, Tech. Rep. FHWA-HRT-06–137, 2006.

[50] 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)*, Dec. 2018, pp. 2118–2125, doi: 10.1109/ITSC.2018.8569552.

[51] B. Mersch, T. Höllen, K. Zhao, C. Stachniss, and R. Roscher, "Maneuver-based trajectory prediction for self-driving cars using spatio–temporal convolutional networks," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2021, pp. 4888–4895, doi: 10.1109/IROS51168.2021.9636875.

[52] C. Dong, Y. Zhang, and J. M. Dolan, "Lane-change social behavior generator for autonomous driving car by non-parametric regression in reproducing kernel Hilbert space," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 4489–4494, doi: 10.1109/IROS.2017.8206316.

[53] C. Hubmann, J. Schulz, M. Becker, D. Althoff, and C. Stiller, "Automated driving in uncertain environments: Planning with interaction and uncertain maneuver prediction," *IEEE Trans. Intell. Vehicles*, vol. 3, no. 1, pp. 5–17, Mar. 2018, doi: 10.1109/IVS.2018.8500697.

[54] W. Zhan, A. de La Fortelle, Y. Chen, C. Chan, and M. Tomizuka, "Probabilistic prediction from planning perspective: Problem formulation, representation simplification and evaluation metric," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1150–1156, doi: 10.1109/IVS.2018.8500697.

[55] F. Leon and M. Gavrilescu, "A review of tracking, prediction and decision making methods for autonomous driving," 2019, *arXiv:1909.07707*.

[56] M. S. Shirazi and B. T. Morris, "Looking at intersections: A survey of intersection monitoring, behavior and safety analysis of recent studies," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 4–24, Jan. 2017, doi: 10.1109/TITS.2016.2568920.

[57] B. T. Morris and M. Trivedi, "Understanding vehicular traffic behavior from video: A survey of unsupervised approaches," *J. Electron. Imag.*, vol. 22, no. 4, pp. 1–16, 2013, doi: 10.1117/1.JEI.22.4.041113.

[58] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: *Common practices and emerging technologies*," *IEEE Access*, vol. 8, pp. 58443–58469, 2020, doi: 10.1109/ACCESS.2020.2983149.

[59] V. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[60] A. Pentland and A. Liu, "Modeling and prediction of human behavior," *Neural Comput.*, vol. 11, no. 1, pp. 229–242, 1999, doi: 10.1162/089976699300016890.

[61] F. You, R. Zhang, G. Lie, H. Wang, H. Wen, and J. Xu, "Trajectory planning and tracking control for autonomous lane change maneuver based on the cooperative vehicle infrastructure system," *Expert Syst. Appl.*, vol. 42, no. 14, pp. 5932–5946, 2015, doi: 10.1016/j.eswa.2015.03.022.

[62] Z. Li, J. Jiang, and W.-H. Chen, "Automatic lane change maneuver in dynamic environment using model predictive control method," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 2384–2389, doi: 10.1109/IROS45743.2020.9341729.

[63] T. Hulnhagen, I. Dengler, A. Tamke, T. Dang, and G. Breuel, "Maneuver recognition using probabilistic finite-state machines and fuzzy logic," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2010, pp. 65–70, doi: 10.1109/IVS.2010.5548066.

[64] D. Kasper *et al.*, "Object-oriented Bayesian networks for detection of lane change maneuvers," *IEEE Intell. Transp. Syst. Mag.*, vol. 4, no. 3, pp. 19–31, Aug. 2012, doi: 10.1109/MITS.2012.2203229.

[65] D. Shin, S. Yi, K.-M. Park, and M. Park, "An interacting multiple model approach for target intent estimation at urban intersection for application to automated driving vehicle," *Appl. Sci.*, vol. 10, no. 6, p. 2138, 2020, doi: 10.3390/app10062138.

[66] A. Houenou, P. Bonnifait, V. Cherfaoui, and W. Yao, "Vehicle trajectory prediction based on motion model and maneuver recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 4363–4369, doi: 10.1109/IROS.2013.6696982.

[67] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 1549–15498, doi: 10.1109/CVPRW.2018.00196.

[68] A. Benterki, M. Boukhnifer, V. Judalet, and C. Maaoui, "Driving intention prediction and state recognition on highway," in *Proc. 29th Medit. Conf. Control Autom. (MED)*, Jun. 2021, pp. 566–571, doi: 10.1109/MED51440.2021.9480326.

[69] C. Wissing, T. Nattermann, K.-H. Glander, and T. Bertram, "Probabilistic time-to-lane-change prediction on highways," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1452–1457, doi: 10.1109/IVS.2017.7995914.

[70] C. Wissing, T. Nattermann, K.-H. Glander, and T. Bertram, "Trajectory prediction for safety critical maneuvers in automated highway driving," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 131–136, doi: 10.1109/ITSC.2018.8569296.

[71] J. Schlechtriemen, F. Wirthmueller, A. Wedel, G. Breuel, and K.-D. Kuhnert, "When will it change the lane? A probabilistic regression approach for rarely occurring events," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2015, pp. 1373–1379, doi: 10.1109/IVS.2015.7225907.

[72] F. Wirthmüller, J. Schlechtriemen, J. Hipp, and M. Reichert, "Towards incorporating contextual knowledge into the prediction of driving behavior," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–7, doi: 10.1109/ITSC45102.2020.9294665.

[73] 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, Sep. 2019, pp. 839–843, doi: 10.1109/IDAACS.2019.8924448.

[74] J. X. Li, B. Dai, X. Li, X. Xu, and D. Liu, "A dynamic Bayesian network for vehicle maneuver prediction in highway driving scenarios: Framework and verification," *Electronics*, vol. 8, no. 1, p. 40, 2019, doi: 10.3390/electronics8010040.

[75] M. Sefati, J. Chandiramani, K. Kreiskoether, A. Kampker, and S. Baldi, "Towards tactical behaviour planning under uncertainties for automated vehicles in urban scenarios," in *Proc. 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–7, doi: 10.1109/ITSC.2017.8317819.

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

[77] H. Q. Dang, J. Fürnkranz, A. Biedermann, and M. Hoepfl, "Time-to-lane-change prediction with deep learning," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–7, doi: 10.1109/ITSC.2017.8317674.

[78] A. Zyner, S. Worrall, and E. Nebot, "A recurrent neural network solution for predicting driver intention at unsignalized intersections," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1759–1764, Jul. 2018, doi: 10.1109/LRA.2018.2805314.

[79] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Non-local social pooling for vehicle trajectory prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 975–980, doi: 10.1109/IVS.2019.8813829.

[80] A. Benterki, V. Judalet, M. Choubeila, and M. Boukhnifer, "Long-term prediction of vehicle trajectory using recurrent neural networks," in *Proc. 45th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2019, pp. 3817–3822, doi: 10.1109/IECON.2019.8927604.

[81] K. Cho *et al.*, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[82] M. Khakzar, A. Rakotonirainy, A. Bond, and S. G. Dehkordi, "A dual learning model for vehicle trajectory prediction," *IEEE Access*, vol. 8, pp. 21897–21908, 2020, doi: 10.1109/ACCESS.2020.2968618.

[83] Y. Hu, W. Zhan, and M. Tomizuka, "Probabilistic prediction of vehicle semantic intention and motion," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 307–313, doi: 10.1109/IVS.2018.8500419.

[84] P. Felsen, P. Lucey, and S. Ganguly, "Where will they go? Predicting fine-grained adversarial multi-agent motion using conditional variational autoencoders," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 732–747, doi: 10.1007/978-3-030-01252-6_45.

[85] H. M. Le, Y. Yue, P. Carr, and P. Lucey, "Coordinated multi-agent imitation learning," 2017, *arXiv:1703.03121*.

[86] N. Lee and K. M. Kitani, "Predicting wide receiver trajectories in American football," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2016, pp. 1–9, doi: 10.1109/WACV.2016.7477732.

[87] C. Sun, P. Karlsson, J. Wu, J. B. Tenenbaum, and K. Murphy, "Stochastic prediction of multi-agent interactions from partial observations," 2019, *arXiv:1902.09641*.

[88] E. Zhan *et al.*, "Generative multi-agent behavioral cloning," 2018, *arXiv:1803.07612*.

[89] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 961–971, doi: 10.1109/CVPR.2016.110.

[90] J. Amirian, J.-B. Hayet, and J. Pettre, "Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Jun. 2019, pp. 1–9, doi: 10.1109/CVPRW.2019.00359.

[91] A. Kuefler, J. Morton, T. Wheeler, and M. Kochenderfer, "Imitating driver behavior with generative adversarial networks," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 204–211, doi: 10.1109/IVS.2017.7995721.

[92] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," 2018, *arXiv:1803.10892*.

[93] B. Ivanovic, K. Leung, E. Schmerling, and M. Pavone, "Multimodal deep generative models for trajectory prediction: A conditional variational autoencoder approach," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 295–302, Apr. 2021.

[94] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," 2016, *arXiv:1606.03498*.

[95] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," 2017, *arXiv:1701.04862*.

[96] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 214–223. [Online]. Available: http://proceedings.mlr.press/v70/arjovsky17a.html

[97] N. Nikhil and B. Tran Morris, "Convolutional neural network for trajectory prediction," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, Sep. 2018, pp. 1–11, doi: 10.1007/978-3-030-11015-4_16.

[98] L. Sun, W. Zhan, C.-Y. Chan, and M. Tomizuka, "Behavior planning of autonomous cars with social perception," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 207–213, doi: 10.1109/IVS.2019.8814223.

[99] A. G. Cunningham, E. Galceran, R. M. Eustice, and E. Olson, "MPDM: Multipolicy decision-making in dynamic, uncertain environments for autonomous driving," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2015, pp. 1670–1677, doi: 10.1109/ICRA.2015.7139412.

[100] W. Ding, L. Zhang, J. Chen, and S. Shen, "Safe trajectory generation for complex urban environments using spatio–temporal semantic corridor," *IEEE Robot. Autom. Lett.*, vol. 4, no. 3, pp. 2997–3004, Jun. 2019, doi: 10.1109/LRA.2019.2923954.

[101] S. Lefèvre, C. Laugier, and J. Ibañez-Guzmán, "Exploiting map information for driver intention estimation at road intersections," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2011, pp. 583–588, doi: 10.1109/IVS.2011.5940452.

[102] S. Danielsson, L. Petersson, and A. Eidehall, "Monte Carlo based threat assessment: Analysis and improvements," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2007, pp. 233–238, doi: 10.1109/IVS.2007.4290120.

[103] J. Firl, H. Stübing, S. A. Huss, and C. Stiller, "Predictive maneuver evaluation for enhancement of car-to-x mobility data," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2012, pp. 558–564, doi: 10.1109/IVS.2012.6232217.

[104] J. Mercat, N. El Zoghby, G. Sandou, D. Beauvois, and G. P. Gil, "Kinematic single vehicle trajectory prediction baselines and applications with the NGSIM dataset," 2019, *arXiv:1908.11472*.

[105] Y. Jeong and K. Yi, "Target vehicle motion prediction-based motion planning framework for autonomous driving in uncontrolled intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 1, pp. 168–177, Dec. 2021, doi: 10.1109/TITS.2019.2955721.

[106] S. Dixit *et al.*, "Trajectory planning for autonomous high-speed overtaking in structured environments using robust MPC," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2310–2323, Jun. 2020, doi: 10.1109/TITS.2019.2916354.

[107] Z. Li, J. Jiang, and W.-H. Chen, "Automatic lane merge based on model predictive control," in *Proc. 26th Int. Conf. Autom. Comput. (ICAC)*, Sep. 2021, pp. 1–6, doi: 10.23919/ICAC50006.2021.9594261.

[108] J. Nilsson, M. Brännström, E. Coelingh, and J. Fredriksson, "Longitudinal and lateral control for automated lane change maneuvers," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2015, pp. 1399–1404, doi: 10.1109/ACC.2015.7170929.

[109] J. Suh, B. Kim, and K. Yi, "Stochastic predictive control based motion planning for lane change decision using a vehicle traffic simulator," in *Proc. IEEE Transp. Electrific. Conf. Expo, Asia–Pacific (ITEC Asia-Pacific)*, Jun. 2016, pp. 900–907, doi: 10.1109/ITEC-AP.2016.7513079.

[110] J. Nilsson and J. Sjöberg, "Strategic decision making for automated driving on two-lane, one way roads using model predictive control," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 1253–1258.

[111] J. Nilsson, M. Brännström, J. Fredriksson, and E. Coelingh, "Longitudinal and lateral control for automated yielding maneuvers," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 5, pp. 1404–1414, May 2016.

[112] C. Dong and J. M. Dolan, "Continuous behavioral prediction in lane-change for autonomous driving cars in dynamic environments," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3706–3711, doi: 10.1109/ITSC.2018.8569772.

[113] Y. Xu, T. Zhao, C. Baker, Y. Zhao, and Y. N. Wu, "Learning trajectory prediction with continuous inverse optimal control via Langevin sampling of energy-based models," in *Proc. 18th Int. Conf. Auto. Agents Multiagent Syst. (AAMAS)*, May 2019.

[114] C. Ju, Z. Wang, C. Long, X. Zhang, and D. E. Chang, "Interaction-aware Kalman neural networks for trajectory prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 1793–1800, doi: 10.1109/IVS.2013.6629638.

[115] H. Coskun, F. Achilles, R. Dipietro, N. Navab, and F. Tombari, "Long short-term memory Kalman filters: Recurrent neural estimators for pose regularization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Los Alamitos, CA, USA, Oct. 2017, pp. 5525–5533, doi: 10.1109/ICCV.2017.589.

[116] C. Badue *et al.*, "Self-driving cars: A survey," 2019, *arXiv:1901.04407*.

[117] M. Buehler, K. Iagnemma, and S. Singh, *The 2005 DARPA Grand Challenge: The Great Robot Race*, vol. 36. Springer, 2007.

[118] M. Buehler, K. Iagnemma, and S. Singh, *The DARPA Urban Challenge: Autonomous Vehicles in City Traffic*, vol. 56. New York, NY, USA: Springer, 2009.

[119] A. Zyner, S. Worrall, and E. Nebot, "Naturalistic driver intention and path prediction using recurrent neural networks," 2018, *arXiv:1807.09995*.

[120] 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, doi: 10.1016/j.trc.2020.102615.

[121] T. Yang, Z. Nan, H. Zhang, S. Chen, and N. Zheng, "Traffic agent trajectory prediction using social convolution and attention mechanism," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Oct. 2020, pp. 278–283, doi: 10.1109/IV47402.2020.9304645.

[122] C. Breazeal and B. Scassellati, "A context-dependent attention system for a social robot," in *Proc. IJCAI*, 1999, pp. 1146–1153. [Online]. Available: http://ijcai.org/Proceedings/99-2/Papers/068.pdf

[123] B. R. Duffy, C. Rooney, G. M. O'Hare, and R. O'Donoghue, "What is a social robot?" in *Proc. 10th Irish Conf. Artif. Intell. Cogn. Sci.*, Sep. 1999, pp. 1–3.

[124] H. Gweon and R. Saxe, "Developmental cognitive neuroscience of theory of mind," in *Neural Circuit Development and Function in the Brain: Comprehensive Developmental Neuroscience*, J. Rubenstein and P. Rakic, Eds. Amsterdam, The Netherlands: Elsevier, 2013.

[125] N. Rhinehart, R. McAllister, K. Kitani, and S. Levine, "PRECOG: PREdiction conditioned on goals in visual multi-agent settings," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2821–2830, doi: 10.1109/ICCV.2019.00291.

[126] X. Li, X. Ying, and M. C. Chuah, "GRIP: Graph-based interaction-aware trajectory prediction," in *Proc. IEEE Intell. Transp. Syst. Conf. (ITSC)*, Oct. 2019, pp. 3960–3966, doi: 10.1109/ITSC.2019.8917228.

[127] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, and D. Manocha, "TrafficPredict: Trajectory prediction for heterogeneous traffic-agents," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 6120–6127, doi: 10.1609/aaai.v33i01.33016120.

[128] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, pp. 1805–1824, Aug. 2000, doi: 10.1103/PhysRevE.62.1805.

[129] C. Dong, Y. Chen, and J. M. Dolan, "Interactive trajectory prediction for autonomous driving via recurrent meta induction neural network," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 1212–1217, doi: 10.1109/ICRA.2019.8794392.

[130] S. Depeweg, J.-M. Hernandez-Lobato, F. Doshi-Velez, and S. Udluft, "Decomposition of uncertainty in Bayesian deep learning for efficient and risk-sensitive learning," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 1184–1193. [Online]. Available: http://proceedings.mlr.press/v80/depeweg18a.html

[131] J. Ho and S. Ermon, "Generative adversarial imitation learning," 2016, *arXiv:1606.03476*.

[132] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," 2017, *arXiv:1709.04875*.

[133] A. Graves, *Long Short-Term Memory*. Berlin, Germany: Springer, 2012, pp. 37–45, doi: 10.1007/978-3-642-24797-2_4.

[134] X. Li, X. Ying, and M. C. Chuah, "GRIP++: Enhanced graph-based interaction-aware trajectory prediction for autonomous driving," Jul. 2019, *arXiv: 1907.07792*.

[135] H. Jeon, D. Kum, and W. Jeong, "Traffic scene prediction via deep learning: Introduction of multi-channel occupancy grid map as a scene representation," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1496–1501, doi: 10.1109/IVS.2018.8500567.

[136] E. Isufi, F. Gama, and A. Ribeiro, "EdgeNets: Edge varying graph neural networks," 2020, *arXiv:2001.07620*.

[137] J. Li, H. Ma, and M. Tomizuka, "Conditional generative neural system for probabilistic trajectory prediction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 6150–6156, doi: 10.1109/IROS40897.2019.8967822.

[138] N. Rhinehart, K. M. Kitani, and P. Vernaza, "R2P2: A reparameterized pushforward policy for diverse, precise generative path forecasting," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 772–788, doi: 10.1007/978-3-030-01261-8_47.

[139] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 2255–2264, doi: 10.1109/CVPR.2018.00240.

[140] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified LSTM models for trajectory prediction," *IEEE Access*, vol. 7, pp. 38287–38296, 2019, doi: 10.1109/ACCESS.2019.2907000.

[141] M. Schreiber, S. Hoermann, and K. Dietmayer, "Long-term occupancy grid prediction using recurrent neural networks," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 9299–9305, doi: 10.1109/ICRA.2019.8793582.

[142] S. Casas, W. Luo, and R. Urtasun, "Intentnet: Learning to predict intention from raw sensor data," in *Proc. 2nd Conf. Robot Learn.*, vol. 87, A. Billard, A. Dragan, J. Peters, and J. Morimoto, Eds., 29–31, Oct. 2018, pp. 947–956. [Online]. Available: http://proceedings.mlr.press/v87/casas18a.html

[143] S. Hoermann, M. Bach, and K. Dietmayer, "Dynamic occupancy grid prediction for urban autonomous driving: A deep learning approach with fully automatic labeling," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2018, pp. 2056–2063, doi: 10.1109/ICRA.2018.8460874.

[144] F. Diehl, T. Brunner, M. T. Le, and A. Knoll, "Graph neural networks for modelling traffic participant interaction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 695–701, doi: 10.1109/IVS.2019.8814066.

[145] X. Mo, Y. Xing, and C. Lv, "ReCoG: A deep learning framework with heterogeneous graph for interaction-aware trajectory prediction," 2020, *arXiv:2012.05032*.

[146] E. Jo, M. Sunwoo, and M. Lee, "Vehicle trajectory prediction using hierarchical graph neural network for considering interaction among multimodal maneuvers," *Sensors*, vol. 21, no. 16, p. 5354, 2021, doi: 10.3390/s21165354.

[147] X. Mo, Y. Xing, and C. Lv, "Graph and recurrent neural network-based vehicle trajectory prediction for highway driving," in *Proc. IEEE Int. Intell. Transp. Syst. Conf. (ITSC)*, Sep. 2021, pp. 1934–1939, doi: 10.1109/ITSC48978.2021.9564929.

[148] T. Zhao *et al.*, "Multi-agent tensor fusion for contextual trajectory prediction," 2019, *arXiv:1904.04776*.

[149] D. Nuss *et al.*, "A random finite set approach for dynamic occupancy grid maps with real-time application," *Int. J. Robot. Res.*, vol. 37, no. 8, pp. 841–866, Jul. 2018, doi: 10.1177/0278364918775523.

[150] D. Lee, Y. P. Kwon, S. McMains, and J. K. Hedrick, "Convolution neural network-based lane change intention prediction of surrounding vehicles for ACC," in *Proc. IEEE 20th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2017, pp. 1–6, doi: 10.1109/ITSC.2017.8317874.

[151] T. Fu and J. Wang. (2021). *Tongji Road Trajectory Sharing Platform (TJRD TS)*. [Online]. Available: https://www.tjrdts.com

[152] W. Zhan *et al.*, "Interaction dataset: An International, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," 2019, *arXiv:1910.03088*.

[153] J. Bock, R. Krajewski, T. Moers, S. Runde, L. Vater, and L. Eckstein, "The inD dataset: A drone dataset of naturalistic road user trajectories at German intersections," 2019, *arXiv:1911.07602*.

[154] X. Chen and P. Chaudhari, "MIDAS: Multi-agent interaction-aware decision-making with adaptive strategies for urban autonomous navigation," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2021, pp. 7980–7986, doi: 10.1109/ICRA48506.2021.9561148.

[155] V. Punzo, M. T. Borzacchiello, and B. Ciuffo, "Estimation of vehicle trajectories from observed discrete positions and next-generation simulation program (NGSIM) data," in *Proc. TRB 88th Annu. Meeting*. Washington, DC, USA: U.S. Department of Transportation, Jan. 2009. [Online]. Available: https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm

[156] S. Agarwal, A. Vora, G. Pandey, W. Williams, H. Kourous, and J. McBride, "Ford multi-AV seasonal dataset," *Int. J. Robot. Res.*, vol. 39, no. 12, pp. 1367–1376, 2020, doi: 10.1177/0278364920961451.

[157] M. Cordts *et al.*, "The cityscapes dataset for semantic urban scene understanding," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Los Alamitos, CA, USA, Jun. 2016, pp. 3213–3223, doi: 10.1109/CVPR.2016.350.

[158] H. Caesar *et al.*, "NuScenes: A multimodal dataset for autonomous driving," 2019, *arXiv:1903.11027*.

[159] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361, doi: 10.1109/CVPR.2012.6248074.

[160] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The KITTI dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013, doi: 10.1177/0278364913491297.

[161] J. Fritsch, T. Kühnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2013, pp. 1693–1700, doi: 10.1109/ITSC.2013.6728473.

[162] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3061–3070, doi: 10.1109/CVPR.2015.7298925.

[163] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions," 2019, *arXiv:1812.04767*.

[164] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "Carla: An open urban driving simulator," in *Proc. Conf. Robot Learn.*, vol. 78. S. Levine, and V. Vanhoucke, and K. Goldberg, 2017, pp. 1–16. [Online]. Available: https://proceedings.mlr.press/v78/dosovitskiy17a.html

[165] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis, "A survey on 3D object detection methods for autonomous driving applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 10, pp. 3782–3795, Oct. 2019, doi: 10.1109/TITS.2019.2892405.

[166] S. H. Park *et al.*, "Diverse and admissible trajectory forecasting through multimodal context understanding," 2020, *arXiv:2003.03212*.

[167] B. Ivanovic and M. Pavone, "The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2375–2384, doi: 10.1109/ICCV.2019.00246.

[168] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder–decoder architecture," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1672–1678, doi: 10.1109/IVS.2018.8500658.

[169] C. Choi, S. Malla, A. Patil, and J. H. Choi, "DROGON: A trajectory prediction model based on intention-conditioned behavior reasoning," 2019, *arXiv:1908.00024*.

[170] A. Bender, J. R. Ward, S. Worrall, and E. M. Nebot, "Predicting driver intent from models of naturalistic driving," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015, pp. 1609–1615, doi: 10.1109/ITSC.2015.262.

[171] Z. Li, C. Lu, Y. Yi, and J. Gong, "A hierarchical framework for interactive behaviour prediction of heterogeneous traffic participants based on graph neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 1–13, Jul. 2021, doi: 10.1109/TITS.2021.3090851.

[172] Y. Hu, W. Zhan, and M. Tomizuka, "Scenario-transferable semantic graph reasoning for interaction-aware probabilistic prediction," 2020, *arXiv:2004.03053*.

[173] S. E. Shladover, "Cooperative (rather than autonomous) vehicle-highway automation systems," *IEEE Intell. Transp. Syst. Mag.*, vol. 1, no. 1, pp. 10–19, Jul. 2009, doi: 10.1109/MITS.2009.932716.

[174] A. Correa *et al.*, "Infrastructure support for cooperative maneuvers in connected and automated driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2019, pp. 20–25, doi: 10.1109/IVS.2019.8814044.

[175] A. Correa *et al.*, "On the impact of V2X-based maneuver coordination on the traffic," in *2021 IEEE 93rd Veh. Technol. Conf. (VTC-Spring)*, Apr. 2021, pp. 1–5, doi: 10.1109/VTC2021-Spring51267.2021.9448700.

[176] *Intelligent Transport System (ITS); Vehicular Communications; Basic Set of Applications; Analysis of the Collective—Perception Service (CPS)*, Standard ETSI TR. 103 562, V0.0.15, 2019.

[177] L. Xin, P. Wang, C.-Y. Chan, J. Chen, S. E. Li, and B. Cheng, "Intention-aware long horizon trajectory prediction of surrounding vehicles using dual LSTM networks," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1441–1446, doi: 10.1109/ITSC.2018.8569595.

[178] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1179–1184, doi: 10.1109/IVS.2018.8500493.

[179] T. G. R. Reid *et al.*, "Localization requirements for autonomous vehicles," 2019, *arXiv:1906.01061*.

[180] T. Phan-Minh, E. C. Grigore, F. A. Boulton, O. Beijbom, and E. M. Wolff, "CoverNet: Multimodal behavior prediction using trajectory sets," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 14062–14071, doi: 10.1109/CVPR42600.2020.01408.

[181] H. Cui *et al.*, "Deep kinematic models for kinematically feasible vehicle trajectory predictions," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, May 2020, pp. 10563–10569, doi: 10.1109/ICRA40945.2020.9197560.

[182] H. Girase, J. Hoang, S. Yalamanchi, and M. Marchetti-Bowick, "Physically feasible vehicle trajectory prediction," 2021, *arXiv:2104.14679*.

[183] L. Zhang, P.-H. Su, J. Hoang, G. Clark Haynes, and M. Marchetti-Bowick, "Map-adaptive goal-based trajectory prediction," 2020, *arXiv:2009.04450*.

[184] G. Xie, H. Gao, L. Qian, B. Huang, K. Li, and J. Wang, "Vehicle trajectory prediction by integrating physics- and maneuver-based approaches using interactive multiple models," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5999–6008, Jul. 2018, doi: 10.1109/TIE.2017.2782236.

[185] D. Yang, S. Zheng, C. Wen, P. J. Jin, and B. Ran, "A dynamic lane-changing trajectory planning model for automated vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 228–247, Oct. 2018, doi: 10.1016/j.trc.2018.06.007.

[186] Y. Luo, Y. Xiang, K. Cao, and K. Li, "A dynamic automated lane change maneuver based on vehicle-to-vehicle communication," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 87–102, Jan. 2016, doi: 10.1016/j.trc.2015.11.011.

[187] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," 2017, *arXiv:1705.07874*.

[188] H. Jeon, J. Choi, and D. Kum, "SCALE-Net: Scalable vehicle trajectory prediction network under random number of interacting vehicles via edge-enhanced graph convolutional neural network," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 2095–2102, doi: 10.1109/IROS45743.2020.9341288.

[189] N. Deo, A. Rangesh, and M. M. Trivedi, "How would surround vehicles move? A unified framework for maneuver classification and motion prediction," *IEEE Trans. Intell. Vehicles*, vol. 3, no. 2, pp. 129–140, Jun. 2018, doi: 10.1109/TIV.2018.2804159.

[190] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2016, pp. 549–565.

[191] J. Martinez *et al.*, "On human motion prediction using recurrent neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, 2017, pp. 4674–4683.

[192] S. Pellegrini, A. Ess, and L. Van Gool, "Improving data association by joint modeling of pedestrian trajectories and groupings," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2010, pp. 452–465, doi: 10.1007/978-3-642-15549-9_33.

[193] L. Leal-Taixe, M. Fenzi, A. Kuznetsova, B. Rosenhahn, and S. Savarese, "Learning an image-based motion context for multiple people tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3542–3549.

[194] W. Zhan, L. Sun, D. Wang, Y. Jin, and M. Tomizuka, "Constructing a highly interactive vehicle motion dataset," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Nov. 2019, pp. 6415–6420, doi: 10.1109/IROS40897.2019.8967724.

[195] W. Zhan *et al.*, "Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," 2019, *arXiv:1910.03088*.

[196] S. Pellegrini, A. Ess, K. Schindler, and L. van Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 261–268, doi: 10.1109/ICCV.2009.5459260.

[197] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by example," *Comput. Graph. Forum*, vol. 26, no. 3, pp. 655–664, 2007, doi: 10.1111/j.1467-8659.2007.01089.x.

[198] H. Caesar *et al.*, "NuScenes: A multimodal dataset for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11618–11628, doi: 10.1109/CVPR42600.2020.01164.

[199] W. Luo, B. Yang, and R. Urtasun, "Fast and furious: Real time end-to-end 3D detection, tracking and motion forecasting with a single convolutional net," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 3569–3577, doi: 10.1109/CVPR.2018.00376.

[200] H. Girase *et al.*, "LOKI: Long term and key intentions for trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9803–9812.

[201] X. Huang *et al.*, "The apolloscape dataset for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, Los Alamitos, CA, USA, Jun. 2018, pp. 1067–10676, doi: 10.1109/CVPRW.2018.00141.

[202] W. Ding and S. Shen, "Online vehicle trajectory prediction using policy anticipation network and optimization-based context reasoning," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 9610–9616, doi: 10.1109/ICRA.2019.8793568.

[203] A. Zyner, S. Worrall, J. Ward, and E. Nebot, "Long short term memory for driver intent prediction," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1484–1489, doi: 10.1109/IVS.2017.7995919.

[204] D. J. Phillips, T. A. Wheeler, and M. J. Kochenderfer, "Generalizable intention prediction of human drivers at intersections," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 1665–1670, doi: 10.1109/IVS.2017.7995948.

[205] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," 2014, *arXiv:1412.3555*.

[206] N. Djuric *et al.*, "Uncertainty-aware short-term motion prediction of traffic actors for autonomous driving," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Los Alamitos, CA, USA, Mar. 2020, pp. 2084–2093, doi: 10.1109/WACV45572.2020.9093332.

[207] H. Cui *et al.*, "Multimodal trajectory predictions for autonomous driving using deep convolutional networks," in *Proc. Int. Conf. Robot. Automat. (ICRA)*, May 2019, pp. 2090–2096, doi: 10.1109/ICRA.2019.8793868.

[208] L. Thiede and P. Brahma, "Analyzing the variety loss in the context of probabilistic trajectory prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9953–9962, doi: 10.1109/ICCV.2019.01005.

Djamel Eddine Benrachou received the Ph.D. degree in 2018 for his study on "Hypovigilance detection and assistance to vehicle drivers." He is currently a Ph.D. Research Scholar with the Centre for Accident Research and Road Safety—Queensland (CARRS-Q), Queensland University of Technology (QUT), actively working on social interaction in intention prediction to improve automated vehicle safety. He has previously worked as a Research Associate with the Centre for Development of Advanced Technologies (CDTA)—Algiers, Algeria. His current area of knowledge is social robotics applications to automated vehicles in optimizing motion prediction, social interaction among various road users by developing algorithms using advanced learning tools. His central background is computer science with advanced deep learning techniques in multiple applications.

Sebastien Glaser received the Ph.D. degree in automatic and control in 2004 (defining a driving assistance system in interaction with the driver). He is currently a Professor in intelligent transportation system at CARRS-Q, where he focuses on a safe and sustainable development/deployment of automated driving system in interaction with others road users (drivers, cyclists, and pedestrians). He leads large iMOVE projects on automation with the Queensland Department of Transport and Main Roads, and attracts two Australian Research Council fundings. He worked as a Researcher in the development of connected and automated vehicles (CAV). He was involved in several European Union initiatives (EU FP6, such as SAFESPOT on V2I communication) and in the French National Research Agency (ANR) initiative. Since 2009, he has worked across academic and industrial sectors and held senior researcher positions. He has created, with Dominique Gruyer, CIVITEC, which commercialized the research outputs on virtual environment and simulation (and is now a part of the ESI Group). He has been the Deputy Director and the Director of LIVIC (a research unit of IFSTTAR, the French Institute of Science and Technology for Transport, Spatial Planning, Development and Networks) from 2012 to 2015 and the Project Leader of VEDECOM (Public Private Partnership Research Institute) from 2015 to 2017, developing the autonomous vehicle prototypes. He was involved in French (ANR) and European (FP7 and H2020) initiatives on AV.

Mohammed Elhenawy received the Ph.D. degree from the Transportation Research Laboratory, Virginia Polytechnic Institute and State University. He worked for three years as a Post-Doctoral Researcher with the Virginia Tech Transportation Institute (VTTI), Blacksburg, VA, USA. He is currently a Senior Research Fellow with the Centre for Accident Research and Road Safety—Queensland (CARRS-Q) and a Faculty Member of the Queensland University of Technology. He has authored or coauthored more than 60 ITS-related articles. His research interests include machine learning, statistical learning, game theory, and their application in intelligent transportation systems (ITS) and cooperative intelligent transportation systems (C-ITS).

Andry Rakotonirainy is currently the Director of the Centre for Accident Research Road Safety-Queensland (CARRS-Q), where he is also a Founder of the Intelligent Transport Systems Human Factors Research Program. He has 20 years of research and management experience in computer science and brings advanced expertise in road safety and ITS. He has been proactive in investigating the use of existing and emerging ITS from multiple disciplines. It incorporates disciplines such as computer science, mathematics, human factors, engineering, psychology, and sociology. His research has made extensive use of driving simulators, traffic simulators, and instrumented vehicles for developing system prototypes, assessing cost benefits, understanding human errors, and evaluating system deployment. His research on ITS has received numerous competitive grants and generated extensive interest from road safety stakeholders.