Article

Surrogate Safety Measures Prediction at Multiple Timescales in V2P Conflicts Based on Gated Recurrent Unit

Matteo Miani [1,*], Matteo Dunnhofer [2], Christian Micheloni [2], Andrea Marini [3,4] and Nicola Baldo [1]

[1] Polytechnic Department of Engineering and Architecture (DPIA), University of Udine, Via del Cotonificio 114, 33100 Udine, Italy; nicola.baldo@uniud.it

[2] Department of Mathematics, Computer Science and Physics (DMIF), University of Udine, Via delle Scienze 206, 33100 Udine, Italy; matteo.dunnhofer@uniud.it (M.D.); christian.micheloni@uniud.it (C.M.)

[3] Department of Languages, Literatures, Communication, Education and Society (DILL), University of Udine, Via Margreth 3, 33100 Udine, Italy; andrea.marini@uniud.it

[4] Claudiana—Landesfachhochschule für Gesundheitsberufe, I-39100 Bolzano, Italy

[*] Correspondence: matteo.miani@phd.units.it

Abstract: Improving pedestrian safety at urban intersections requires intelligent systems that should not only understand the actual vehicle–pedestrian (V2P) interaction state but also proactively anticipate the event's future severity pattern. This paper presents a Gated Recurrent Unit-based system that aims to predict, up to 3 s ahead in time, the severity level of V2P encounters, depending on the current scene representation drawn from on-board radars' data. A car-driving simulator experiment has been designed to collect sequential mobility features on a cohort of 65 licensed university students who faced different V2P conflicts on a planned urban route. To accurately describe the pedestrian safety condition during the encounter process, a combination of surrogate safety indicators, namely TAdv (Time Advantage) and T_2 (Nearness of the Encroachment), are considered for modeling. Due to the nature of these indicators, multiple recurrent neural networks are trained to separately predict T_2 continuous values and TAdv categories. Afterwards, their predictions are exploited to label serious conflict interactions. As a comparison, an additional Gated Recurrent Unit (GRU) neural network is developed to directly predict the severity level of inner-city encounters. The latter neural model reaches the best performance on the test set, scoring a recall value of 0.899. Based on selected threshold values, the presented models can be used to label pedestrians near accident events and to enhance existing intelligent driving systems.

Keywords: ADAS; traffic safety; surrogate safety measures; driver behavior; Gated Recurrent Units; driving simulator

1. Introduction

Modern innovations in car-sensing devices, along with the development of deep learning techniques and recognition algorithms, has allowed engineers and researchers worldwide to implement increasingly reliable Advanced Driver Assistance (ADAS) and Automated Driving (ADS) Systems, which are expected to lead toward an improvement in road users' safety levels (reducing the severity of injuries and/or preventing fatalities on roads), driving efficiency (e.g., vehicle fuel consumption), and, consequently, the sustainability of transportation infrastructures. The capabilities of modern cars to detect surrounding objects, represent traffic situations, and adapt their dynamic state have been enhanced to the point that some car manufacturers have first-ever presented driving automation systems capable of performing the entire dynamic driving task in a sustained manner under specific operating conditions.

Nevertheless, deaths of vulnerable road users (VRUs) continue to make for a significant percentage of all road fatalities worldwide [1], and, consequently, more attentive

actions are called for attaining the road infrastructure's sustainability in terms of protecting VRUs. Although active safety systems are not considered driving automation as they provide momentary, not sustained, vehicle control intervention during potentially hazardous situations [2], the European New Car Assessment Program (Euro-NCAP), the leading NCAP in the world [3], has recently restated the key role of Automatic Emergency Braking (AEB) systems in preventing accidents that involve cars and VRUs [4], and it has also promoted the further development of intervention-type ADAS to increase the ADS's overall driving automation capabilities. AEB systems are vehicles' active safety systems that assist the driver to avoid potential collisions or mitigate the severity of unavoidable impacts [5]. These systems exploit sensors and recognition algorithms to detect vehicles, pedestrians, and other objects in the road environment that interact (or could interact) with the moving vehicle. If certain safety-critical thresholds for the interaction are exceeded and the driver does not take any evasive action, AEB systems take active control of the brakes and, in some cases, other vehicles' subsystems (steering wheel, throttle, suspension, etc.), thereby reducing fatalities, severity of injuries, and social costs [5,6]. However, AEB systems generally consist of three subsystems or levels, namely "perception", "decision making", and "execution", which can have quite different performance depending on the vendor/supplier that has developed it and the sensor technology used to acquire the scene data. For these reasons and to ensure adequate functioning in a wide range of traffic scenarios, AEB systems are under continuous development. In addition, some researchers have found that if vehicles were capable of understanding and anticipating the intentions (or trajectories) of drivers and nearby road users one second in advance, most traffic accidents could be prevented [7]. In fact, such a prediction can be exploited to make appropriate driving decisions in advance, such as adjusting a vehicle's trajectory to avoid a car-to-pedestrian accident or adapting assistance systems to the driver's intentions (i.e., to determine, in the current situation representation, whether and when to initiate or abort an intervention, avoiding unnecessary driving interference).

Most mass-market ADAS systems do not have a medium-term predictive capability of the road users' intentions [8], as they are designed to act reactively in high-risk situations (particularly AEB systems). Machine learning (ML) techniques are emerging in the ADAS development field as the main approach to motion prediction [9]. In the outlined context, ML models must learn from inputs that are time-series, e.g., the current and past positions of the traffic participants, and produce outputs that are future sequences. In recent years, Long Short-Term Memory (LSTM) [10] and Gated Recurrent Units (GRU) [11] architectures, variants of the more general Recurrent Neural Networks (RNNs), have shown excellent efficiency in time-series prediction tasks: these models can capture the sequences' dynamics by extracting relevant information from an arbitrarily long context window and retaining a state of that information [12]. Since RNNs variants are currently the preferred option for sequential data modeling, relevant LSTM- or GRU-based literature studies are presented hereafter.

Existing techniques for learning driver or other road user behavior sequences from the set of features acquired by the vehicle sensor system can be divided between two methods: classification and regression. Classification problems concern the identification of movement intention labels, which are also called "behavior primitives" [13]: these classes segment complex driving behavior into a sequence of basic elements, such as lane keeping, left/right lane change, left/right turn, go straight, or speed maintenance, braking, and stopping. In the latter context, Khairdoost et al. [14] implemented a deep learning system that can anticipate (by 3.6 seconds on average) driver maneuvers (left/right lane change, left/right turn, and go straight), exploiting the driver's gaze and head position as well as vehicle dynamics data. Differently, regression problems are concerned with predicting the future positions of cars [8], cyclists [15], and pedestrians [16] surrounding the ego-vehicle (i.e., the vehicle, also called the "subject vehicle" or "vehicle under test", whose behavior is of primary interest in the traffic scenario), by a general understanding of their movement dynamics. Among recent RNNs applications in problems relevant to

autonomous vehicle movement within urban settings, Huang et al. [17] encoded temporal and spatial interactions between pedestrians in a crowded space by combining an LSTM and a Graph Attention Network (GAT) to obtain "socially" plausible trajectories.

Although several literature studies have attempted to evaluate the trajectories and intentions of road users, AEB systems require a Risk Assessment Model (RAM, core element of the perception level) that can objectively and quickly capture the risk level of the encounter process between the ego-vehicle and a VRU. In fact, an RAM that can fully understand the relationships between behavior and risk is essential to adequately judge the AEB system's intervention timing and thus prevent collisions [18]. Risk (or severity) level is intended as the potential of an elementary traffic event to become an accident [19]. Specifically for vehicle–pedestrian (V2P) encounters in inner-city traffic (i.e., the simultaneous arrival of a driver and a pedestrian at the crosswalk or in a specific limited area), such process is a traffic event characterized by a continuous interaction over time and space between the two road users [20] The pedestrian's decision to enter the zebra crossing depends on the perceived speed and distance of the approaching vehicle; concurrently, the driver evaluates whether to grant or deny the priority to the pedestrian, based on the estimated arrival time at the crosswalk. Since the two traffic participants may enter a collision course during the encounter process, such a conflict has the potential to end up in a collision. For example, the latter would occur if the driver's attention levels, his/her ability to control the vehicle, or the vehicle's dynamic state were not adequate for a safe stopping behavior [21].

Many researchers have addressed the issue of risk assessment for pedestrian–vehicle interactions both for current and future encounter states (for the latter, on the basis of predicted trajectories) [7]. However, the application of traffic safety indicators, also referred to as Traffic Conflict Techniques (TCTs), has been very successful as a proactive surrogate approach (i.e., complementary to accident statistical analysis) for traffic event safety assessment, due to its efficiency and short analysis time [22]. There are various continuous or discrete TCTs for V2P conflicts [23]. Nevertheless, Laureshyn et al. [19] identified and developed a set of safety indicators to continuously describe the severity level of the encounter process and, thus, to relate "individual interactions to the general safety situation" of the event. In fact, a single indicator is not sufficient to accurately classify interaction patterns into severity categories (i.e., the RAM's purpose), as it cannot fully reflect the current safety situation [23]. Differently, a combination of at least two indicators should be considered to properly identify pedestrian "near-accident" [24] situations (i.e., the traffic conflicts between safe passages and collisions), which are the most relevant for pedestrian-AEB systems (PAEB), using appropriate threshold values on the selected TCTs. Among the indicators analyzed by Laureshyn et al. [19], Table 1 provides a detailed description of those most relevant to the discussion that follows: Time to Collision (TTC), Nearness of the Encroachment (T_2), Post-Encroachment Time (PET), and Time Advantage (TAdv).

Upon selecting safety indicators to classify near-accident events, the study by Kathuria and Vedagiri [25] showed that TTC and PET profiles are of equal importance to effectively categorize pedestrian–vehicle interactions at unsignalized intersections into severity levels when neither pedestrian nor vehicle takes an evasive action. Moreover, Zhang et al. [23] trained a GRU neural network to predict near-accident events at signalized intersections using PET and TTC indicators generated from videos captured by fixed cameras. The latter work is also of particular interest, as it represents the first attempt (although it was not intended for the development of PAEB systems) to implement a model capable of directly predicting the current severity level of V2P interactions, which were described by three categories: "serious conflict", "slight conflict", and "safe". However, as described previously, the encounter process may have crash potential even though the two road users are not on a collision course [19]. Conversely, the TTC calculation requires both users to be on a collision course and, thus, limits the events to be consider in safety analysis. These findings reveal that V2P encounters are extremely complicated and that,

to improve the reliability of RAMs, safety indicators capable of describing the whole encounter process as a continuous interplay between vehicle and pedestrian should be considered.

The solution to this problem is offered by the supplementary indicators presented by Laureshyn et al. [19]: namely, the Nearness of the Encroachment (T_2) and the Time Advantage (TAdv), which broaden the concepts of TTC and PET, respectively, to situations in which the two road users are not on a collision course (Table 1). In addition, Borsos et al. [26] recently performed a comparison of collision course indicators with indicators that include crossing course interactions at signalized intersections, demonstrating that TTC and T_2 are transferable for crash probability estimation, with stricter threshold values for T_2.

Table 1. Definition of TCTs.

Indicator	Definition	Type	Remarks
Time to Collision (TTC)	The time it would take for two road users on a collision course to collide if they maintained their current trajectory and relative speed.	A set of values continually calculated over time	Among all the pedestrian–vehicle front-end contact points in a collision, the one leading to the lowest TTC value should be selected. The lower the TTC, the higher the risk.
Nearness of the Encroachment (T_2)	The expected time that it takes for the second road user to arrive at the conflict zone.	A set of values continually calculated over time	At the moment of transfer from crossing to collision course, T_2 provides a smooth transition between the two situations and equals the TTC.
Post-Encroachment Time (PET)	The time that would elapse between the passage of the first and the second user through the same conflict zone.	A discrete value	For an encounter, the TAdv has a single value that can be measured directly.
Time Advantage (TAdv)	At any time, it represents the expected PET value if the road users maintained their current trajectory and relative speed.	A set of values continually calculated over time	Values above 2–3 s indicate that a user has a temporal advantage over his opponent in a competition over the same spatial zone and is likely to pass first.

This paper presents a GRU-based system that predicts, up to 3 s ahead in time, the severity level of V2P encounters in inner-city traffic (i.e., encounters between a car and a pedestrian on a pedestrian crossing), depending on the current scene representation drawn from on-board radars' data. A car driving simulator experiment has been designed to collect sequential mobility features on a cohort of 65 licensed university students and generate T_2 and TAdv indicators for accurately classifying pedestrian safety conditions during the whole encounter process. Based on selected threshold values, the presented model could be used to label in advance pedestrians near accident events and to enhance existing PAEB systems. So, this might be a relevant contribution to improving the transportation safety.

Compared to the existing literature that has mainly focused on predicting the trajectories (or intentions) of the ego-vehicle and/or other surrounding road users [9], the developed approach differs not only by modeling safety parameters directly related to the

V2P interaction severity levels but also by the fact that the multi-step-ahead prediction depends on a low-dimensional representation of the current situation: the calibrated system depends only on six parameters related to driver mobility features and traffic scene properties. Indeed, Ortiz et al. [13] proved that there is no need to employ the state of the actuators as features for predicting future behavior: the authors have actually predicted with simple learning algorithms (i.e., multi-layer perceptron neural networks) the braking behavior of drivers approaching a traffic light with very good accuracy at time scales up to 3 seconds, using as input features only the ego-vehicle speed, state, and distance to the nearby traffic light. In addition, the approach presented in this paper has some aspects in common with the research of Zhang et al. [23], but the authors predict the instantaneous severity level (based on TTC and PET) of vehicle–pedestrian interactions whose dynamics are captured by fixed cameras placed at signalized intersections, whereas the aim of the current study is a multi-step-ahead prediction on a running vehicle.

2. Data Collection

To meet the purpose of the study, it would require the collection of mobility data on several vehicle–pedestrian encounters in real-world urban scenarios. In addition, these data must allow for the proper evaluation of interaction safety indicators (especially during an online system application) and be easily acquired by on-board sensors (i.e., cameras, pedal potentiometers, IMU, GNSS, or millimeter-wave radars). In particular, the analysis of the relevant literature [9] allowed us to identify three main groups of features for studying the problem under analysis: the actuators and steering wheel states, the information about the car dynamics, the pedestrian speed, and the direction vector. To collect this information while keeping drivers safe, a cohort of 65 licensed university students was recruited to participate in a driving simulation experiment at the Road Laboratory of the Polytechnic Department of Engineering and Architecture (DPIA) of the University of Udine.

The use of advanced driving simulators for the analysis of driver behaviors and the development of AEB systems is an accepted and widely recognized practice [27], since the driver's performance observed in driving simulation shows the same patterns as real-world driving (relative validity) [28]. Saito et al. [29] designed and validated with a driving simulator a PAEB system that controls vehicle subsystems (brake and accelerator) in potentially dangerous or uncertain situations to decelerate the vehicle and maintain a safe driving speed. Hou et al. [30], studying the braking behaviors of drivers in typical vehicle-to-bicycle conflicts with a driving simulator, proposed a method to improve the timing and braking phases of bicyclist-AEB systems. Bella et al. [27] recently used a fixed-base driving simulator to evaluate the functionality and effectiveness of two types of ADAS that provide the driver with an audible alarm and a visual alarm to detect a pedestrian crossing into and outside the crosswalk.

2.1. The Car-Driving Simulator Experiment

The car-driving simulator at the Road Laboratory (product name AutoSim 1000-M) has been already validated and successfully used for studying the drivers' braking behavior affected by cognitive distractions [21]. The simulator cockpit is made with real car parts (e.g., dashboard, steering wheel, pedals, gear lever, handbrake, driver seat, seatbelt) of an Italian city car. These are important components of the equipment to give more realistic sensations to the driver during the simulation experiment, along with the steering force feedback, the engine sound, and the two-degree-of-freedom motion base system that reproduces the vehicle's roll and pitch. Three 43-inch LCD screens allow the road scenario to be visually reproduced, showing a 180° view.

In the Lab's AutoSim 1000-M driving simulator [21], it was possible to record a dataset for the GRU models' development by observing test participants' behavior toward two planned traffic encounters: a boy/girl entering a crosswalk from the curb. The simulated scenes (Figure 1) were set in a typical urban environment on a course that could take

about 15 minutes to be completed, considering the 50 km/h speed limit. On the urban course, participants experienced many traffic light intersections, tight turns (90°), short straight streets, and occasional crossings of pedestrians, some of which occurred outside the crosswalks. The encounters that were used to record data (and compute surrogate safety indicators for each participant in an offline fashion, following the procedure described in Appendix A of the study by Laureshyn et al. [19]) were set on a four-lane road (two lanes in each direction), placing traffic signs and markings that met European standards (Figure 1).

| (**a**) | (**b**) |

Figure 1. Frontal view of simulated scenarios: (**a**) boy crossing; (**b**) girl crossing.

The scenarios were designed as follows: (1) the participant arrives at a red traffic light approximately 200 m from the crosswalk, on which he/she has a clear view; (2) as the participant starts driving, the pedestrian (initially hidden) walks at a 90° angle toward the road and then stops at the edge of the curb; (3) at the moment the vehicle, based on its current speed, is about 3 s from the crosswalk, the pedestrian enters the zebra crossing and maintains a speed of 1.4 m/s. These conditions, consistent with relevant literature [31], require the participant to stop, giving priority to the pedestrian. In this way, the data collected refer to stopping maneuvers that avoided a collision with the pedestrian.

2.2. Partecipant Statistics and Experimental Procedure

The students recruited for the experiment were between the ages of 20 and 30, had a valid driver's license, and had driven at least 5000 km in the past year. All participants were properly trained in the use of the simulator actuators before the experimental driving and were able to test their abilities on a simulated suburban course that took approximately 5 minutes. Before the urban driving simulation, each participant filled in a questionnaire aimed at collecting his/her personal data (e.g., age, driving experience). Conversely, at the end of the simulation test, a second questionnaire about the discomfort perceived while driving was completed by each participant to identify and remove from the cohort those drivers who had experienced excessive annoyance. The simulation test procedure just presented has been drawn from relevant literature that provided for its validation [21,27,31]. The recruited cohort includes 19 females and 41 males, with a mean age of 24.0 (SD 3.43) and 24.1 (SD 1.91) years and a mean non-verbal intelligence quotient (IQ) [32] of 33.4 (SD 2.50) and 33.9 (SD 1.66), respectively. Therefore, the sample of males and females can be considered roughly balanced for age and IQ. However, the original cohort had an additional five participants who were excluded from the analyses, as they could not complete the experiment due to excessive discomfort.

It is worth pointing out that participation in the experiment was voluntary, there was no monetary reward, and all participants gave informed consent after being instructed on

the simulated driving test procedure. Conversely, they were not informed about the objectives of the research. Finally, the study has been conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Local Ethics Committee of Udine's University (Progetto_Guida).

2.3. Problem Formulation

Formally, the severity prediction system has to perform a mapping between the situation representation at the current driving time t, i.e., the real-time scene properties and vehicle status described by a set of sensor-observed features, and the expected severity level of the vehicle–pedestrian encounter, which is defined by T2 and TAdv, for times t+1s, t+2s, and t+3s. In practice, a learning algorithm is trained to accurately predict the surrogate safety indicators at instants $t-2$s, $t-1$s, and t using the vehicle sensing system observations available at instant t-3s. Once the expected convergence between predicted outputs and ground-truth targets is reached, the trained model represents the desired medium-term prediction system and can be used to make forward-in-time predictions based on the current scene representation. It is assumed that the input features can all be acquired simultaneously and at a regular time interval given by the lowest sampling frequency among those of the involved sensors, since the goal is to perform real-time prediction on running vehicles. However, considering that the current study is based on a driving simulator experiment, the calculation of surrogate safety indicators for each participant and the training of the learning algorithm were both performed in an offline fashion (please refer to Appendix A of the study by Laureshyn et al. [19]) at the end of the driving experiments.

2.4. Low-Dimensional Input Representation

The coordinate system established for the vehicle sensing system is shown in Figure 2. The following features form the six components of the learning model input vector and describe the V2P interaction process at the current time t for each participant in the experimental data set: driver's behavior primitive $A(t)$, current $T_2(t)$ value, ego-vehicle's speed $v_v(t)$, pedestrian's speed $v_p(t)$, and position vector $(r_0(t), \phi(t))$.
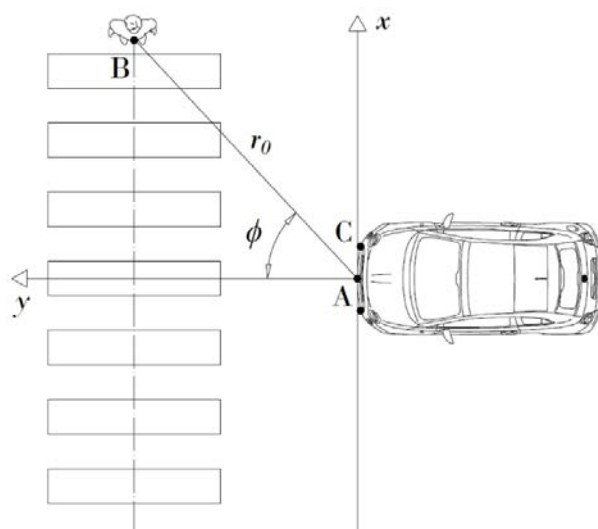

Figure 2. Ego-vehicle coordinate system.

Behavior primitives are the set of elementary behaviors into which the driver's approaching maneuver (in the longitudinal dimension of the event) can be segmented based only on speed, gas pedal, and brake pedal information [13]. These data, easily acquired on modern vehicles (e.g., by means of pedal potentiometers, IMU, GNSS), are processed to define a categorical variable whose values, in the case of V2P conflicts, are 0 for

"stopped behavior", 1 for "braking behavior", and 2 for "maintaining speed behavior". Based on the study by Ortiz et al. [13], the parts of the stream wherein the vehicle speed v_v is less than 2 km/h (about 0.56 m/s) are labeled "stopped". The "braking" behavior begins at the moment the driver releases the gas pedal completely, since in urban traffic, this condition represents the beginning of slowing down in response to an event. All other moments in the stream that do not belong to the two illustrated categories are labeled as "maintaining speed". The categorical variable of behavior primitives represents important information for the GRU model, as it allows the sequence of mobility features to be segmented according to the current driver behavior. Similarly, the current value of T_2, which can be calculated directly using the other features as shown by Laureshyn et al. [19], is of high importance, since neurosciences indicates that the human brain relies closely on judgments of TTC (and, consequently, of T_2, since the two indicators are transferable [26]) to perform coordinated action [33].

Position vector components are the azimuth angle ϕ of the pedestrian with respect to the travel direction and the distance r_0 between the ego-vehicle (point A or C) and the pedestrian (point B, please see Figure 2). Specifically, v_p, r_0, and ϕ are the pedestrian state data acquired by millimeter-wave radars mounted on the vehicle front end (points A and C, Figure 2). In fact, since the detection of the pedestrian presence in the scene is usually performed using robust frame classification algorithms on vehicle camera videos [34], the status of the detected pedestrian can be easily acquired by radars: in this study, we assumed data acquisition equipment consisting of a long-distance millimeter-wave radar installed at the center of the vehicle front end and a mid-range radar on either side of the vehicle front, with maximum detection distances of 100 m and 50 m and azimuth angles of ±10° and ±45°, respectively, to ensure the optimal scene coverage [18]. Compared to video sensors that require object tracking and perspective transformation after object detection to generate the trajectory profiles of VRUs in time-series [23], radars allow the direct and continuous acquisition of pedestrian mobility features [18]. However, the use of cameras is essential for pedestrian detection in the traffic scene; in this regard, for the further online analysis of the proposed system, the use of the state-of-the-art Mask R-CNN (Region-based Convolutional Neural Network) [35] is recommended to ensure high performance of the automated object detection process.

Therefore, time sequences of participants' maneuvers begin with the detection of the pedestrian's status from the long-distance radar, approximately 100 m from the crosswalk, assuming the simultaneous recognition of the pedestrian's presence by the ego-vehicle detection model. Time sequences end when the first road user (the pedestrian) leaves the conflict zone or the second road user (the driver) comes to a complete stop, and none of the TCTs can be calculated (since the collision is no longer possible). Furthermore, although the sampling rate of radars (e.g., the current 77 GHz band millimeter-wave radar) is 20 Hz, we assumed that the encounter process is recorded at 10 Hz to account for limitations of other on-board equipment and raw data processing times.

2.5. Safety Indicators and Severity Classes Generations

Regarding the model output vector, we decided to split the severity level prediction problem into single-output learning tasks: T_2 continuous values and TAdv categories were the learning target of two distinct GRU models. Although both safety indicators can be calculated continually over time, TAdv does not provide the same smooth transfer between crossing and collision courses as T_2: conversely, it quickly goes to zero and holds that value as long as the two road users remain on a collision course (since, by definition, TAdv cannot take on negative values). After the change from collision to crossing courses due to the driver's braking or slowing down, the TAdv value starts gradually growing. Such a behavior induces singularities in the TAdv pattern during the maneuver (i.e., if TAdv is plotted on a graph as a function of time, it will not make a continuous curve), which are difficult to capture with ML regression techniques, unless overfitting the model.

To avoid running into such problems, TAdv values less than 1s have been labeled as "collision course" and those greater than 1s have been labeled as "crossing course", since this indicator can be interpreted as the minimal delay of the driver that, if applied, will result in a collision course [19]. Conversely, T_2 is defined by a continuous curve over time and provides information about how soon the encroachment will occur. Thus, the TAdv prediction is treated as a binary classification problem, whereas the T_2 prediction represents a regression problem.

After training these models separately, their predictions were used to classify conflict interactions ahead in time, distinguishing between "safe" and "unsafe" processes. These categories have been defined according to the static threshold values reported in the literature [23,26,36,37] and the conditions under which the simulation experiment was performed [31]: when the TAdv class is "collision course" and the T_2 value is less than 3 s over the same time horizon, the interaction is defined as "unsafe"; in all other cases, it is considered "safe". Additional intermediate classes have been avoided (e.g., the case wherein TAdv is reporting "collision course" but the T_2 value is greater than 3 s), since the main interest is to detect actual hazard situations or "serious conflict" iterations ahead of time (i.e., before they happen) [23]. By comparing the models' results with the ground-truth targets, it was possible to evaluate the effectiveness of the proposed system in classifying pedestrian's near-accident events. However, an additional GRU neural network has been developed to directly predict the severity level of V2P encounters to verify which of the proposed models would guarantee the best classification performance.

It is worth recalling that the procedure reported in Appendix A of the study by Laureshyn et al. [19] has been applied for the offline calculation of TCTs. In addition, the dimensions of the simulated vehicle were considered in calculations so that, among all the pedestrian–vehicle front-end contact points in a potential collision, the one leading to the lowest T_2 and TAdv values has been selected. Then, the obtained values were time-translated (by 1, 2, and 3 s backward) to compose the ground-truth target matrix.

Finally, before being inputted to the GRU, each feature (and target) was standardized; i.e., it was subtracted by its minimum value and divided by the distance between its minimum and maximum value computed on the training set, to remain in an acceptable range with respect to the activation functions.

3. Methodology

3.1. Gated Recurrent Unit (GRU)

Due to their extended success in many time-series applications [38–43], RNNs were employed to implement the proposed learning system. In particular, the GRU architecture was selected because of its increased robustness with respect to vanilla RNNs and of its expressive power that matches the more sophisticated LSTM networks, but which is achieved with less training effort.

The working mechanism of GRU networks is now described generally [11,12]. Based on the input feature vector input and previous output, each GRU learning neuron performs different operations using so-called "gate" operators. The "update" gate decides the amount of past information to be forwarded to the future, while the "forget" gate focuses on which part of the past information to forget. In more detail, let us suppose x_t to be the feature vector available at the t step of the time-series, h_t to be the hidden state, and y_t to be the output vector. These vectors are used by the GRU in the following operations:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \tag{1}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1}\right) \tag{2}$$

$$\hat{h}_t = tanh\left(W_c x_t + U_c(r_t \odot h_{t-1})\right) \tag{3}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \tag{4}$$

which are applied for all the temporal states t happening during the series, i.e., from $t = 0$ to $t = T$ with T being the series length (at $t = 0$, h_0 is set to be a zero vector). The before mentioned "update" gate and "forget" gate are implemented by Equations (1) and (2) respectively. W_z, U_z, W_f, U_f, W_c, and U_c are weight vectors that are optimized during the training phase of the GRU, $\sigma(\cdot)$ is an activation function that is implemented as a sigmoid, while $\odot$ is the element-wise product. For a better comprehension, the flow of operations is visualized in Figure 3.
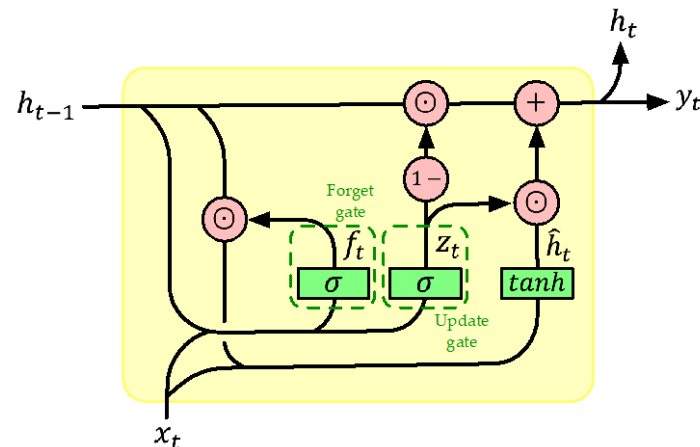

Figure 3. Schematic representation of the operation flow performed by a GRU cell.

These particular operations allow a network to produce more meaningful gradients than vanilla RNNs during the learning phase, ultimately making GRUs acquire enhanced long-term relations between features. To additionally improve the quality and abstractness of such feature relations, the learning model can be organized as stacked layers of GRUs.

In the setting of this work, a GRU composed of two layers each with N neuron cells has been used. The first GRU layer receives in input the feature vector $x_t = \left[A(t), T_2(t), v_v(t), v_p(t), r_0(t), \phi(t)\right]$, which is the sensory information (described in Section 2.4) available at the temporal step t of the time-series representing the driver's maneuver. At the same temporal step, the second GRU layer gets the feature representation outputted by the first GRU layer. After those, a final dense output layer with as many neurons as the outputs was applied to predict the target values. Overall, the role of the GRU layers is to abstract a meaningful representation that summarizes the driver's maneuver up to time-step t. In turn, such higher-level features are exploited by the output layer that produces the predicted future T2 and TAdv states.

3.2. Implementation Details

In this section, the details of the implemented procedure to train the proposed GRU model are given.

A grid search strategy was employed to determine the most important architectural and training hyperparameters such as the number of neurons N of the GRU, the learning rate values, and the learning rate drop factor [44]. The first was tried across the values 64, 128, and 256, the second among 0.05, 0.01, and 0.005, and the third one considering the values 0.40, 0.60, and 0.80 [23]. For each iteration of the grid search, a k-fold cross-validation procedure with five folds was implemented to get the best configuration of the model across different distributions of the employed dataset. In each fold, the full dataset was split subject-wise in training and test sets using a 4:1 ratio, resulting in 68 subjects for training and 17 for testing. For the T2 task, the model has been trained to minimize the

RMSE between its predictions and the ground-truth T_2 targets. For the TAdv task instead, the GRU model was optimized by the minimization of the Binary Cross Entropy loss computed between the model's predictions and TAdv ground-truth categorizations. For all the experiments, the learning procedure has been conducted for 1000 epochs using the Adam optimizer [45]. A weight decay with a factor of 0.0005 was added as a regularization term. The initial learning rate was decayed by the considered values every 50 epochs. A Synthetic Minority Over-Sampling Technique (SMOTE) [46] has been implemented to improve the performance of the model with respect to the test data distribution. This procedure generates new samples of the minority class by interpolating their features. For this work, SMOTE allowed achieving a 1:1 ratio between majority and minority classes that initially was 3:1. Finally, after each epoch, the model was executed on the validation set to assess its generalization capabilities to new maneuvers. The model obtaining the lowest loss function score on such tests, hence the most general one, was retained as the final learned model.

4. Results

The aim of this study was to develop a GRU model that could predict, up to 3 s ahead in time, the level of severity of vehicle–pedestrian encounters in inner-city traffic. In this regard, two equivalent approaches have been identified [9], i.e., (1) using multiple GRUs to separately model the supplementary safety indicators (T_2 and TAdv) that allow the interaction severity to be estimated, or (2) using a single Recurrent Neural Network as a sequence classifier to directly label near-accident events, based on relevant mobility features of V2P encounters. It is worth pointing out that although the two approaches considered are equivalent from the standpoint of the final outcome (i.e., the prediction of severity levels ahead in time), Approach (1) would allow more flexibility in labeling the pedestrian's near-accidental events, since TCTs' threshold values that are different from those considered in this study could be selected to classify the risk level of V2P interactions (e.g., using specific threshold values for geographic contexts of system operation, which are derived from studies of local/national driving behaviors) [23]. In what follows, we refer by the acronyms GRU_{T2} and GRU_{TADV} to the GRU models predicting T_2 and TAdv, respectively. Differently, to make the comparison between the two approaches presented previously, the severity classification model resulting from Approach (1) is presented as $M\text{-}GRU_{SL}$, whereas that of Approach (2) is presented as $S\text{-}GRU_{SL}$.

In this section, the generalization capabilities of the trained models are evaluated for each time horizon (i.e., 1 s, 2 s, 3 s ahead) based on commonly used evaluation metrics, which are different for classification (i.e., Accuracy, Precision, Recall, Specificity, False Alarm Rate (FAR), and Area Under the Curve (AUC)) and regression problems (i.e., Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE)), and averaging their scores over the five test folds. These metrics are briefly presented throughout the discussion to make the models' evaluation clearer to the reader; however, more detailed descriptions can be found in [9,23,47]. Moreover, to prevent predictions from reacting with a delay especially for longer time horizons [8], an additional metric to evaluate T_2's time-series regression, which is usually applied in multi-step-ahead prediction problems [47], has been considered, namely the Modified Index of Agreement (md) [48]. In fact, md is able to concurrently consider differences in observed and predicted means and variances, providing a better evaluation of model predictions than traditional metrics.

After comparing the models' performance under different hyperparameter combinations, the most appropriate values for each GRU model have been selected: the initial learning rate is set to 0.001 for the GRU_{TADV} model and 0.005 for both GRU_{T2} and $S\text{-}GRU_{SL}$ models, whereas the learning rate drop factor is 0.8 for all considered models. The unit number N within the GRU memory is 64 for GRU_{T2}, 256 for GRU_{TADV}, and 128 for $S\text{-}GRU_{SL}$. Tables 2 and 3 present the evaluation metrics scores, validated on the five test

folds, distinguishing by model and prediction time horizon (standard deviations on the five folds are shown in brackets). In addition, Figure 4 presents the empirical cumulative distribution probability (ECDP) of the absolute prediction error over each fold for the GRU_{T2} model. Conversely, Figures 5 and 6 show the Receiver Operating Characteristic curve (ROC) for the GRU_{TADV} and $S-GRU_{SL}$ models, respectively. This curve, which plots recall as a function of FAR, is a comprehensive metric to evaluate classification models' performance [23], since the closer the area under the ROC curve (AUC) is to 1, the better the prediction quality. Finally, Table 4 summarizes the performance of the severity classification models, $M-GRU_{SL}$ and $S-GRU_{SL}$, averaging the test results over folds and time horizons.

Table 2. Experimental results (means and standard deviations) for T2 and TAdv models.

GRU_{T2}	Train Set	Test Set					
Horizons	RMSE	MAE	MAPE	RMSE	*md*		
1s	0.282 (0.017)	0.221 (0.024)	4.539 (0.371)	0.327 (0.033)	0.943 (0.003)		
2s	0.386 (0.014)	0.319 (0.029)	6.617 (0.437)	0.465 (0.042)	0.902 (0.007)		
3s	0.497 (0.036)	0.415 (0.034)	9.048 (0.483)	0.587 (0.054)	0.852 (0.007)		
GRU_{TADV}	Train Set	Test Set					
Horizons	AUC	Accuracy	Precision	Recall	Specificity	FAR	AUC
1s	0.997 (0.001)	0.974 (0.003)	0.919 (0.017)	0.914 (0.023)	0.985 (0.004)	0.015 (0.004)	0.997 (0.001)
2s	0.995 (0.001)	0.968 (0.004)	0.892 (0.027)	0.902 (0.020)	0.980 (0.005)	0.020 (0.005)	0.994 (0.001)
3s	0.992 (0.002)	0.962 (0.007)	0.880 (0.029)	0.874 (0.054)	0.978 (0.006)	0.022 (0.006)	0.991 (0.004)

Table 3. Experimental results (means and standard deviations) for severity classification models.

$M-GRU_{SL}$		Test Set					
	Horizons	Accuracy	Precision	Recall	Specificity	FAR	
	1s	0.971 (0.003)	0.879 (0.049)	0.844 (0.068)	0.986 (0.006)	0.014 (0.006)	
	2s	0.967 (0.005)	0.884 (0.052)	0.792 (0.107)	0.987 (0.007)	0.013 (0.007)	
	3s	0.958 (0.009)	0.873 (0.036)	0.700 (0.095)	0.988 (0.002)	0.012 (0.002)	
$S-GRU_{SL}$	Train Set	Test Set					
Horizons	AUC	Accuracy	Precision	Recall	Specificity	FAR	AUC
1s	0.998 (0.001)	0.984 (0.004)	0.932 (0.029)	0.917 (0.042)	0.992 (0.005)	0.008 (0.005)	0.998 (0.001)
2s	0.997 (0.001)	0.981 (0.006)	0.911 (0.043)	0.907 (0.027)	0.989 (0.006)	0.011 (0.006)	0.997 (0.002)
3s	0.996 (0.002)	0.974 (0.006)	0.887 (0.038)	0.873 (0.057)	0.987 (0.006)	0.014 (0.006)	0.995 (0.002)

Table 4. Test results comparison between severity classification models.

Test Results Comparison	Mean Values over Time Windows and Folds (and Standard Deviations)					
	Accuracy	Precision	Recall	Specificity	FAR	AUC
$M-GRU_{SL}$ Performance	0.965 (0.008)	0.879 (0.043)	0.779 (0.105)	0.987 (0.005)	0.013 (0.005)	-
$S-GRU_{SL}$ Performance	0.980 (0.006)	0.910 (0.039)	0.899 (0.045)	0.989 (0.006)	0.011 (0.006)	0.996 (0.002)
Percentage Variation	+1.55%	+3.53%	+15.40%	+0.20%	−15.38%	-

As expected, the quality of predictions increases if the time horizon decreases, no matter which model is considered. This result suggests that there is a strong correlation between the selected mobility features at time t and the desired target at time t+1. In contrast, this correlation becomes weaker for longer time horizons, and as a result, distinguishing the expected target becomes more difficult. For example, considering the GRU_{T2} model (Table 2), all evaluation metrics get slightly worse as the time-step gets longer both in training and testing stages. However, focusing on the test RMSE (column 5, Table 2), i.e., the square root of the mean squared difference between predicted and observed val-

ues, its worsening is characterized by a deviation close to one-tenth of a second: the maximum deviation, equal to 0.138 s, is measured moving from the 1s (RMSE = 0.327 s) to the 2s (RMSE = 0.465 s) prediction time-frame, whereas the prediction quality worsens by 0.122 moving from 2 s to 3 s (RMSE = 0.587 s). Therefore, the results obtained are on the whole satisfactory, as also shown by the other parameters: the test MAE (column 3, Table 2), i.e., the mean absolute error, was on each time scale less than half a second; the 90th percentile of the absolute error (Figure 4), i.e., the value of the prediction error (in absolute value) which is exceeded in no more than 10% of time sequences, is equal to 0.472 s (averaged over the five folds) at the 1 s, 0.696 s at the 2 s, and 0.967 s at the 3 s prediction time-frame; the md parameter is always greater than 0.850 whatever the prediction time horizon, proving the close agreement between the observed and forecast curves.

(**a**)   (**b**)   (**c**)
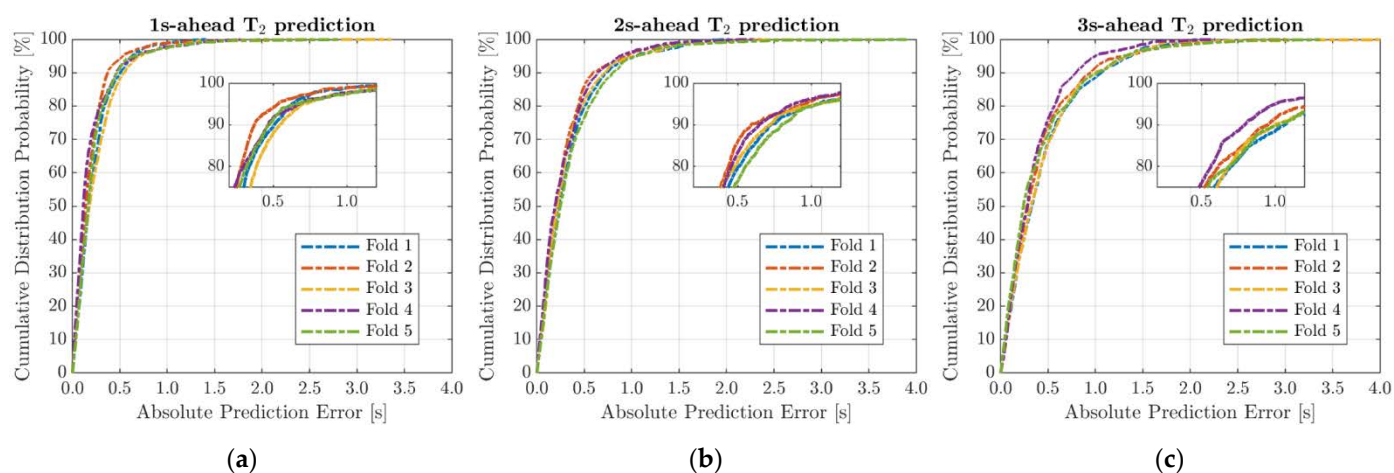
Figure 4. Cumulative distribution probability of absolute prediction errors over the five test folds: (**a**) 1 s; (**b**) 2 s; (**c**) 3 s ahead.
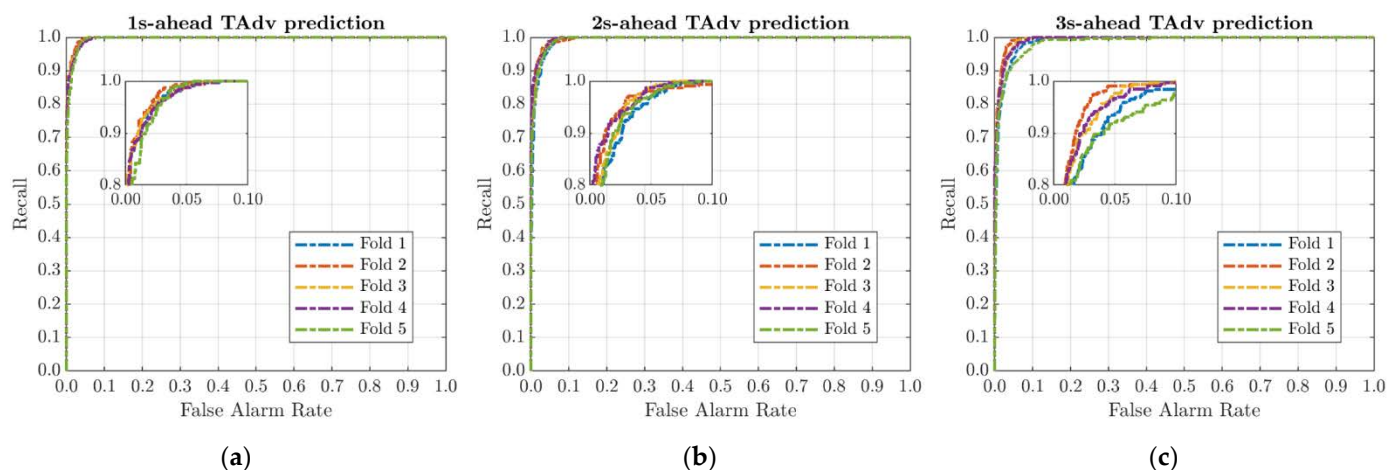

(**a**)   (**b**)   (**c**)

Figure 5. Receiver Operating Characteristic curves of the GRU$_{TADV}$ model for each test fold: (**a**) 1 s; (**b**) 2 s; (**c**) 3 s ahead.
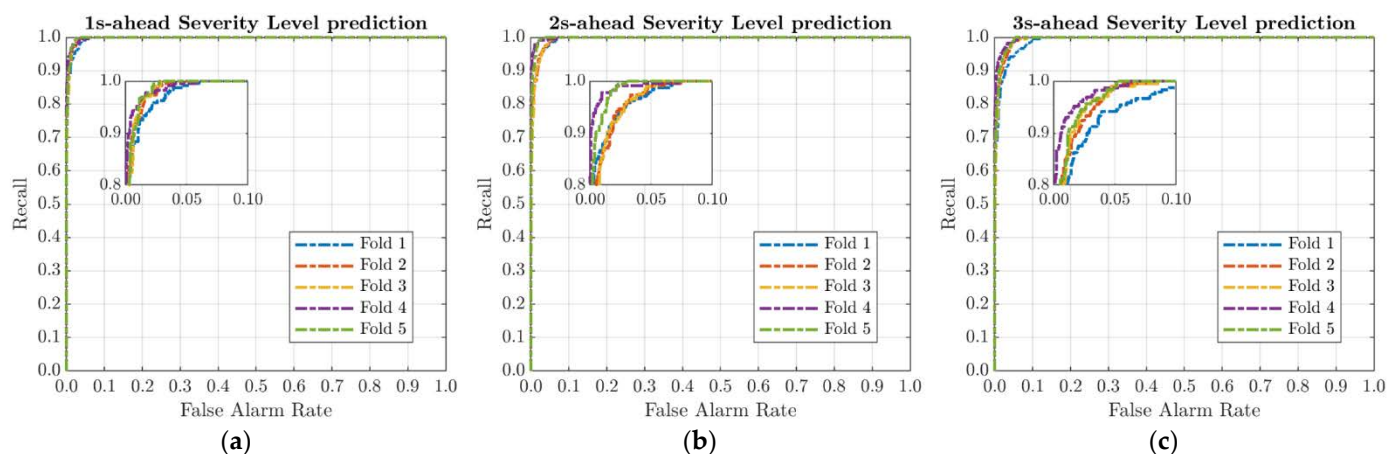
Figure 6. Receiver Operating Characteristic curves of the S-GRU$_{SL}$ model for each test fold: (**a**) 1 s; (**b**) 2 s; (**c**) 3 s ahead.

Regarding the prediction quality of classification models, it is first necessary to remind the reader that in a binary classification problem, samples are labeled as positive and negative so that evaluation metrics can be computed through the confusion matrix. The latter is a representation of the model's classification accuracy, since it makes clear whether the system is mislabeling one class with another: each row of the confusion matrix represents the instances in an actual class, whereas each column represents those in a predicted class. Thus, the number of "true positives" (correctly classified positives), "true negatives" (correctly classified negatives), "false positives" (actual negatives classified incorrectly), and "false negatives" (actual positives classified incorrectly) can be defined. For a better understanding, a schematic representation of the confusion matrix is shown in Figure 7.

Figure 7. Schematic representation of the confusion matrix.

To properly evaluate the TAdv classification model, samples in the "collision course" class have been labeled as positives. Thus, the recall (column 5, Table 2) of the GRU$_{TADV}$ model, representing the proportion of actual positive samples classified correctly, proves that the system is able to predict very accurately (values are greater than 0.870) whether the V2P interaction is moving or not moving on a collision course, whatever the prediction time-frame (despite the slight worsening discussed previously). Similarly, the AUC values (last column of Table 2), greater than 0.995 at each time horizon, confirm the high accuracy of the binary classifier. Differently, Figure 5 shows the effect of data distribution variability over the folds by ROC curves: as the time window gets longer, differences between folds are more evident, i.e., data distribution becomes sparser, and the model loses

slightly in generalization capability. However, the performance is still more than acceptable.

For the S-GRU$_{SL}$ model and the combined M-GRU$_{SL}$ model, samples within the "unsafe" severity class have been labeled as positives. In the following, since the AUC calculation for a model that results from the combination of two GRU subsystems cannot be performed, the comparison between Approaches (1) and (2) is mainly based on accuracy and recall metrics (columns 3 and 5 of Table 3). Analysis of results on the test set shows an improved ability of the S-GRU$_{SL}$ model to label serious conflict interactions in near-accident events, especially over longer prediction timeframes. The accuracy of the S-GRU$_{SL}$ model, i.e., the proportion of samples (positive and negative) classified correctly among all samples, is always higher than 0.970 on each timescale in contrast to the M-GRU$_{SL}$ model, which only on the shortest time frame (1 s) reaches the value of 0.971. The situation gets even worse for the M-GRU$_{SL}$ model to the advantage of the S-GRU$_{SL}$ model in terms of recall (columns 5, Table 3). In fact, the percentage gain in recall score (+15.40%, fourth column of Table 4) by moving from the M-GRU$_{SL}$ to S-GRU$_{SL}$ model is noteworthy.

These results show that the combined model, due to the nonlinear error propagation from the single models (GRU$_{T2}$ and GRU$_{TADV}$), fails to generalize satisfactorily (as also evidenced by the standard deviations over folds) and performs significantly lower than S-GRU$_{SL}$. This finding is not trivial since, in many multi-step-ahead prediction problems, complex models (e.g., multiple GRUs) have performed better than equivalent simple models, such as a single GRU [9].

In conclusion, despite the good results obtained in individual modeling of the T$_2$ and TAdv safety parameters, the S-GRU$_{SL}$ model achieved the best performance, with an accuracy of 0.980, recall of 0.899 and AUC of 0.996 (averaged over the time windows and folds, Table 4) in predicting near-accidental events of V2P encounters.

5. Conclusions

In this study, it has been shown that a simple Deep Learning system, based on GRU cells, can predict ahead in time changes in the severity level of V2P encounters in inner-city traffic. The proposed system uses a gradient-descent optimizer to learn how to label V2P interaction severity by exploiting individual driving features and traffic scene properties. Such an approach differs from previous studies in the two TCTs considered to classify the pedestrian's near-accident events, namely T$_2$ and TAdv, which can be computed either when the road users are on a crossing or a collision course. The trained model, which directly predicts the severity level class (i.e., "safe" or "unsafe") up to 3 s ahead in time, provides satisfactory and promising results (accuracy = 0.980, recall = 0.899, and AUC = 0.996) for enhancing current PAEB systems. In fact, the proposed system could be applied to warn drivers of the anomalous or hazardous interaction with a pedestrian, to anticipate a braking maneuver, as well as to enhance vehicle deceleration, with the aim of pursuing an improvement in transportation safety, with regard to pedestrians, within the inner-city traffic.

6. Future Research

Future research perspectives opened by the current study include (1) comparing the presented multi-step-ahead prediction system to a physical trajectory prediction system; (2) generalizing the proposed GRU models to different and more complex encounter scenarios; (3) modifying the system for online learning and operation; and (4) validating the prediction system with data acquired during real vehicle–pedestrian encounters.

Indeed, a more thorough baseline for comparing the presented approach with those most widely used in the literature should first be established. Further research efforts are also needed on a wider cohort of drivers, as well as on different drivers' groups (such as young inexperienced drivers, experienced and expert drivers, older drivers, etc.), in different urban environments and driving situations to consider as many behavioral compo-

nents (such as aggressiveness and anxiety) and interaction patterns as possible. Well before the online system validation (i.e., the validation of the calibrated system on data acquired in the real-world scenarios), it will be necessary to integrate the risk assessment model (e.g., the S-GRU$_{SL}$) with an automated object detection system (e.g., the Mask R-CNN) and test their coordinated real-time operation (along with the car sensing system) in collecting reliable mobility data during driving simulations in both simple and complex urban scenarios. At this stage, there may be a need to introduce specific learning subsystems for the evaluation of the more complex scenarios (e.g., the ego-vehicle's concurrent interaction with surrounding vehicles and pedestrians), along with a method to merge the predictions given by each subsystem. The further transition to a real vehicle application will require an additional functional requirement: although the risk assessment model (i.e., the presented prediction system) is essential for the good functional safety of an PAEB system, the key to achieving active pedestrian collision avoidance is to control the vehicle dynamics. For this purpose, it will be mandatory to design the upper- and lower-layer controllers which, after receiving the control signal (i.e., the likely occurrence of a collision) from the RAM, are in charge of outputting the deceleration value required for safe stopping and controlling the vehicle subsystems (i.e., throttle opening and brake line pressure regulation) to realize the control of the actual vehicle deceleration, respectively. The control strategy proposed by Yang et al. [18], based on fuzzy neural network and PID (Proportional Integral Derivative controller) theory, could represent a handy reference for the implementation of control modules. Thereafter, using a properly equipped vehicle, the developed automatic emergency braking pedestrian collision avoidance system will have to be online validated in vehicle–pedestrian test scenarios established by the Euro-NCAP standards [49].

In conclusion, this research represents a valuable contribution to the study, understanding, and knowledge of the interaction processes between road users in an urban environment and, consequently, to improving the sustainability of transportation infrastructures.

References

1. Centers for Disease Control and Prevention, 2020: "Road Traffic Injuries and Deaths—A Global Problem". Available online: https://www.cdc.gov/injury/features/global-road-safety/index.html (accessed on 19 August 2021).
2. Shi, E.; Gasser, T.; Seeck, A.; Auerswald, R. The Principles of Operation Framework: A Comprehensive Classification Concept for Automated Driving Functions. *SAE Intl. J CAV* **2020**, *3*, 27–37. https://doi.org/10.4271/12-03-01-0003.

3. Large, D.; Cieslik, I.; Kovaceva, J.; Bruyas, M.P.; Kunert, M.; Krebs, S.; Arbitmann, M. Improving the effectiveness of active safety systems to significantly reduce accidents with vulnerable road users-the Project PROSPECT (Proactive Safety for Pedestrians and Cyclists). In Proceedings of the 26th Enhanced Safety of Vehicles (ESV) Conference, Eindhoven, The Netherlands, 10–13 June 2019; pp. 1–16.

4. Euro NCAP, 2018: "2020 ROADMAP". Available online: https://www.euroncap.com/en/about-euro-ncap/timeline/ (accessed on 19 August 2021).

5. Rosen, E.; Kallhammer, J.E.; Eriksson, D.; Nentwich, M.; Fredriksson, R.; Smith, K. Pedestrian injury mitigation by autonomous braking. *Accid. Anal. Prev.* **2010**, *42*, 1949–1957. https://doi.org/10.1016/j.aap.2010.05.018.

6. Badea-Romero, A.; Paez, F.J.; Furones, A.; Barrios, J.M.; de-Miguel, J.L. Assessing the benefit of the brake assist system for pedestrian injury mitigation through real-world accident investigations. *Saf. Sci.* **2013**, *53*, 193–201. https://doi.org/10.1016/j.ssci.2012.10.004.

7. Wu, R.; Zheng, X.; Xu, Y.; Wu, W.; Li, G.; Xu, Q.; Nie, Z. Modified Driving Safety Field Based on Trajectory Prediction Model for Pedestrian–Vehicle Collision. *Sustainability* **2019**, *11*, 6254. https://doi.org/10.3390/su11226254.

8. Altché, F.; de La Fortelle, A. An LSTM network for highway trajectory prediction. In Proceedings of the 20th IEEE International Conference on Intelligent Transportation Systems, Yokohama, Japan, 16–19 October 2017; pp. 353–359. https://doi.org/10.1109/ITSC.2017.8317913.

9. Mozaffari, S.; Al-Jarrah, O.Y.; Dianati, M.; Jennings, P.; Mouzakitis, A. Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review. *arXiv preprint* **2020**, arXiv: 1912.11676.

10. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

11. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint* **2014**, arXiv:1412.3555.

12. Lipton, Z.C.; Berkowitz, J.; Elkan, C. A Critical Review of Recurrent Neural Networks for Sequence Learning. *arXiv preprint* **2015**, arXiv:1506.00019.

13. Ortiz, M.G.; Fritsch, J.; Kummert, F.; Gepperth, A. Behavior prediction at multiple time-scales in inner-city scenarios. In Proceedings of the 2011 IEEE Intelligent Vehicles Symposium, Baden-Baden, Germany, 5–9 June 2011; pp. 1068–1073. https://doi.org/10.1109/IVS.2011.5940524.

14. Khairdoost, N.; Shirpour, M.; Bauer, M.A.; Beauchemin, S.S. Real-time driver maneuver prediction using LSTM. *IEEE Trans. Intell. Veh.* **2020**, *5*, 714–724. https://doi.org/10.1109/TIV.2020.3003889.

15. Huang, Z.; Wang, J.; Pi, L.; Song, X.; Yang, L. LSTM based trajectory prediction model for cyclist utilizing multiple interactions with environment. *Pattern Recognit.* **2021**, *112*, 107800. https://doi.org/10.1016/j.patcog.2020.107800.

16. Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Fei-Fei, L.; Savarese, S. Social LSTM: Human trajectory prediction in crowded spaces. In Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 961–971. https://doi.org/10.1109/CVPR.2016.110.

17. Huang, Y.; Bi, H.; Li, Z.; Mao, T.; Wang, Z. STGAT: Modeling spatial-temporal interactions for human trajectory prediction. In Proceedings of the 17th IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 6271–6280. https://doi.org/10.1109/ICCV.2019.00637.

18. Yang, W.; Zhang, X.; Lei, Q.; Cheng, X. Research on longitudinal active collision avoidance of autonomous emergency braking pedestrian system (AEB-P). *Sensors* **2019**, *19*, 4671. https://doi.org/10.3390/s19214671.

19. Laureshyn, A.; Svensson, Å.; Hydén, C. Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation. *Accid. Anal. Prev.* **2010**, *42*, 1637–1646. https://doi.org/10.1016/j.aap.2010.03.021.

20. Várhelyi, A. Drivers' speed behaviour at a zebra crossing: A case study. *Accid. Anal. Prev.* **1998**, *30*, 731–743. https://doi.org/10.1016/S0001-4575(98)00026-8.

21. Baldo, N.; Marini, A.; Miani, M. Drivers' braking behavior affected by cognitive distractions: An experimental investigation with a virtual car simulator. *Behav. Sci.* **2020**, *10*, 150. https://doi.org/10.3390/bs10100150.

22. Zheng, L.; Ismail, K.; Meng, X.H. Traffic conflict techniques for road safety analysis: Open questions and some insights. *Can. J. Civ. Eng.* **2014**, *41*, 633–641. https://doi.org/10.1139/cjce-2013-0558.

23. Zhang, S.; Abdel-Aty, M.; Wu, Y.; Zheng, O. Modeling pedestrians' near-accident events at signalized intersections using gated recurrent unit (GRU). *Accid. Anal. Prev.* **2020**, *148*, 105844. https://doi.org/10.1016/j.aap.2020.105844.

24. Hydén, C. *The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique*; Bulletin Lund Institute of Technology, Department: Lund, Sweden, 1987.

25. Kathuria, A.; Vedagiri, P. Evaluating pedestrian vehicle interaction dynamics at un-signalized intersections: A proactive approach for safety analysis. *Accid. Anal. Prev.* **2020**, *134*, 105316. https://doi.org/10.1016/j.aap.2019.105316.

26. Borsos, A.; Farah, H.; Laureshyn, A.; Hagenzieker, M. Are collision and crossing course surrogate safety indicators transferable? A probability based approach using extreme value theory. *Accid. Anal. Prev.* **2020**, *143*, 105517. https://doi.org/10.1016/j.aap.2020.105517.

27. Bella, F.; Silvestri, M. Vehicle–pedestrian interactions into and outside of crosswalks: Effects of driver assistance systems. *Transport* **2021**, *36*, 98–109. https://doi.org/10.3846/transport.2021.14739.

28. Zhang, Y.; Guo, Z.; Sun, Z. Driving Simulator Validity of Driving Behavior in Work Zones. *J. Adv. Transp.* **2020**, *2020*, 4629132. https://doi.org/10.1155/2020/4629132.

29. Saito, Y.; Raksincharoensak, P. Shared control in risk predictive braking maneuver for preventing collisions with pedestrians. *IEEE Trans. Intell. Veh.* **2016**, *1*, 314–324. https://doi.org/10.1109/TIV.2017.2700210.

30. Hou, L.; Duan, J.; Wang, W.; Li, R.; Li, G.; Cheng, B. Drivers' Braking Behaviors in Different Motion Patterns of Vehicle-Bicycle Conflicts. *J. Adv. Transp.* **2019**, *2019*, 4023970. https://doi.org/10.1155/2019/4023970.

31. Bella, F.; Borrelli, V.; Silvestri, M.; Nobili, F. Effects on Driver's Behavior of Illegal Pedestrian Crossings. *Adv. Intell. Syst. Comput.* **2019**, *786*, 802–812. https://doi.org/10.1007/978-3-319-93885-1_74.

32. Raven, J.C. *Progressive Matrices: A Perceptual Test of Intelligence*, 1st ed.; H. K. Lewis & Co. Ltd.: London, UK, 1938.

33. Field, D.T.; Wann, J.P. Perceiving time to collision activates the sensorimotor cortex. *Curr. Biol.* **2005**, *15*, 453–458. https://doi.org/10.1016/j.cub.2004.12.081.

34. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

35. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 386–397. https://doi.org/10.1109/TPAMI.2018.2844175.

36. Ni, Y.; Wang, M.; Sun, J.; Li, K. Evaluation of pedestrian safety at intersections: A theoretical framework based on pedestrian-vehicle interaction patterns. *Accid. Anal. Prev.* **2016**, *96*, 118–129. https://doi.org/10.1016/j.aap.2016.07.030.

37. Zheng, L.; Ismail, K.; Sayed, T.; Fatema, T. Bivariate extreme value modeling for road safety estimation. *Accid. Anal. Prev.* **2018**, *120*, 83–91. https://doi.org/10.1016/j.aap.2018.08.004.

38. Mikolov, T.; Karafiat, M.; Burget, L.; Jan, C.; Khudanpur, S. Recurrent neural network based language model. In Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), Makuhari, Chiba, Japan, 26–30 September 2010; pp. 1045–1048.

39. Fu, R.; Zhang, Z.; Li, L. Using LSTM and GRU neural network methods for traffic flow prediction. In Proceedings of the 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC 2016), Wuhan, China, 11–13 November 2016; 7804912, pp. 324–328. https://doi.org/10.1109/YAC.2016.7804912.

40. Cao, J.; Li, Z.; Li, J. Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A* **2019**, *519*, 127–139. https://doi.org/10.1016/j.physa.2018.11.061.

41. Karevan, Z.; Suykens, J.A.K. Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Netw.* **2020**, *125*, 1–9. https://doi.org/10.1016/j.neunet.2019.12.030.

42. Dunnhofer, M.; Martinel, N.; Micheloni, C. Tracking-by-Trackers with a Distilled and Reinforced Model. In *Lecture Notes in Computer Science, Proceedings of the 15th Asian Conference on Computer Vision (ACCV 2020), Kyoto, Japan, 30 November–4 December 2020*; Ishikawa, H., Liu CL., Pajdla, T., Shi, J., Eds.; Springer: Cham, Switzerland, 2020; pp. 631–650. https://doi.org/10.1007/978-3-030-69532-3_38.

43. Dunnhofer, M.; Martinel, N.; Micheloni, C. Weakly-Supervised Domain Adaptation of Deep Regression Trackers via Reinforced Knowledge Distillation. *IEEE Robot. Autom. Lett.* **2021**, *6*, 5016–5023. https://doi.org/10.1109/LRA.2021.3070816.

44. MathWorks, 2021: "trainingOptions—Options for Training Deep Learning Neural Network". Available online: https://www.mathworks.com/help/deeplearning/ref/trainingoptions.html (accessed on 19 August 2021).

45. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. *arXiv preprint* **2014**, arXiv:1412.6980.

46. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. Smote: Synthetic minority over-sampling technique. *J. Artif. Int. Res.* **2002**, *16*, 321–357. https://doi.org/10.1613/jair.953.

47. Papacharalampous, G.; Tyralis, H.; Koutsoyiannis, D. Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stoch. Environ. Res. Risk Assess.* **2019**, *33*, 481–514. https://doi.org/10.1007/s00477-018-1638-6.

48. Krause, P.; Boyle, D.P.; Bäse, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **2005**, *5*, 89–97. https://doi.org/10.5194/adgeo-5-89-2005.

49. Euro NCAP, 2018: "Vulnerable Road User (VRU) Protection". Available online: https://www.euroncap.com/en/for-engineers/protocols/vulnerable-road-user-vru-protection/ (accessed on 19 August 2021).