[[1]](#footnote-1)

A New Modelling Approach for Predicting Vehicle-based Safety Threats

Nicolette Formosa, Mohammed Quddus, Stephen Ison and Andrew Timmis

*Abstract*—Existing autonomous driving systems of an intelligent vehicle such as advanced driver assistant systems (ADAS) assess and quantify the level of potential threat surrounding the ego-vehicle. However, they may not be able to plan the best response to unexpected dangerous situations and do not have the ability to cope with uncertainties since not all vehicles can always keep a safe gap from preceding vehicles and drive at a desired velocity. Previous research has not taken such uncertainties into account, it is, therefore, necessary to develop models which are not restricted by the predefined movement patterns of a vehicle. In fact, existing systems are based on a model that estimates the threat level based only on one factor: Time-To-Collision (TTC). This approach is limited since it cannot handle all scenarios and ignores all uncertainties. To overcome these limitations, this paper utilised deep learning to develop a range of models that rely on a group of factors to reliably estimate the threat level and predict conflicts under uncertainty using ‘looming’. Comparative analyses were undertaken by incorporating new varying input factors to each model (e.g., surrogate safety measures, vehicle kinematics, macroscopic traffic data). Real-world experiments demonstrated that adding new factors increases the sensitivity of the models. Results also indicated that the models that consider looming provide low false alarm rate extending their application for a wider spectrum of traffic scenarios. This is paramount for ADAS as uncertainties are inherent in the deployment of Connected and Autonomous Vehicles (CAVs) in a mixed traffic stream.

*Index Terms*— Advanced Driver Assistant Systems, Deep Learning, Road Safety, Safety Surrogate Measures, Traffic Conflicts.

# I. INTRODUCTION

O

ver the next decade our economy and society will truly be revolutionized by autonomous, connected, intelligent systems and services, which can learn, adapt, take decisions and act independently of human interventions. Many innovations in vehicular technology are being developed with the aim of reducing the human from the driving loop as much as possible such as the rapidly growing safety applications - Advanced Driver Assistant Systems (ADAS). Developing on-board robust ADAS is crucial since they are the main building blocks for the development and deployment of Connected and Autonomous Vehicles (CAVs). An example of a safety ADAS is to estimate the threat level surrounding the ego-vehicle and predict vehicular conflicts in real-time. Nevertheless, if the alerts are poorly designed, they might result in creating more distractions to drivers, delay their responses, and reduce road safety [1], [2].

Developing ADAS is not a trivial task, as state-of-the-art technologies are required to handle a number of uncertainties such as the immediate future behavior of other traffic participants, the measurements of the ego-vehicle performing the computations and the estimates of the environment (e.g., static obstacle, other traffic participants, lane boundaries). These uncertainties cannot always be described from known physical or mathematical laws, so they are inferred from collected measurements. Pervasive sensors collect significant amounts of data, drawing large streams of raw data. Nowadays, non-intrusive detection sensors are implemented on vehicles such as video, radar, and laser sensors. These sensors are paramount for an effective automotive safety system since they process all the information in real-time to follow the most current vehicle-to-vehicle kinematic conditions [3].

Based on the information collected from the sensors and the models implemented within the safety system, the threat level can be estimated. The main challenge lies in identifying an optimal method to quantify and assess this threat. Significant efforts have been put and various assessment measures have been developed in the literature [4]. Surrogate Safety Measures (SSMs) are considered as an appropriate assessment criterion to describe the safety status of the road network and evaluate risk critical events. They make use of microscopic traffic parameters such as vehicle speed, speed differential, lateral and longitudinal accelerations, time headway and space headway [5]. A pre-defined threshold is set for each SSM and if this measurement value is exceeded, an alert of a possible traffic conflict, e.g., a near miss is issued to the driver [5], [6].

Although a range of SSMs can be employed, many existing ADAS use a single factor as a cue for decision-making [7], e.g. the ﻿Mobileye® C2-270 safety system, uses only Time-To-Collision (TTC) [8]. Using only a single factor for every traffic situation can be limiting due to its limitations and the estimates are only relative to the vehicles ahead of an ego-vehicle driving in the same lane [9], [10] resulting in inconsistencies in the decision process [6]. However, to ensure a high level of ADAS safety, it is desirable that the model is robust in measuring uncertainty and is not limited by the vehicle’s predefined movement patterns [11], [12]. ‘*Looming*’ is one method to estimate whether any surrounding objects are in a conflict course with any point of the ego-vehicle [9], [13]. This method allows the prediction models to be robust by introducing a way to include uncertainty while deciding whether there is a significant threat to the ego-vehicle. It extends the algorithm’s ability to work for a wider spectrum of traffic scenarios [9]. However, conclusions are still drawn dependent on the pre-defined SSM threshold value and therefore, the selection of a value is crucial. Another problem is that this value varies significantly across existing studies [6], [14], [15] and the reasons for this are not entirely well documented.

Nevertheless, existing ADAS implemented in vehicles detect objects only in their close proximity, due to the limited range and the field of view of the sensor used [16]. In fact, a missed alert can take place if a safety threat is initiated outside the range of the sensor. One way to overcome this limitation is by adopting a more accurate in-vehicle sensing system. However, this results in a significant increase in the overall cost while it is still not able to provide any information about the dynamic surroundings of the ego-vehicle and traffic states (e.g., macroscopic traffic data). Consequently, the conflict risk may still not be estimated correctly. Therefore, ﻿the utilisation of highly disaggregated traffic data should be taken into consideration [17], [18].

In existing ADAS, there is no consensus on which factors contribute towards increasing the systems’ sensitivity in predicting vehicle threats. The overarching complexity in building such a model lies in the management, integration and use of off-line and real-time data which are large, disaggregate and heterodox in many ways such as uncertainty, sampling frequency, completeness, structure, consistency, and accuracy. This data necessitates a modelling technique, which is capable to handle large, highly disaggregated, and complex data [19]. While earlier modelling techniques have focused on non-parametric methods (e.g. [20]–[22]), recently, machine learning has gained its popularity in traffic safety. However, these techniques still cannot perform the analysis required for big data with complex characteristics and have not addressed the issues emerging when the data is highly imbalanced. Therefore, the results are inclined to be influenced by the large classes (e.g., ‘safe’ traffic dynamics) while neglecting the smaller ones (e.g., vehicle threats). Deep learning has gained its popularity in traffic safety, particularly due to the success of Deep Neural Network (DNN) [17]. They are also able to elicit complex and non-linear patterns to classify and predict while catering for big data [23]. This has catalysed the motivation to investigate the application of DNN to develop a new modelling approach to predict vehicle-based traffic conflicts.

Multiple traffic conflict prediction models were developed by varying their input factors obtained from vehicle-based sensor data (which is microscopic) and infrastructure-based sensor data (which is macroscopic). Comparative analyses amongst the models were undertaken to determine which factors contribute towards increasing the systems’ sensitivity in predicting traffic conflicts. The primary contributions of the paper are as follows:

* Developing a new modelling approach to predict vehicle-based safety threats under uncertainties.
* Conducting a sensitivity analysis to examine how a combination of factors relating to surrogate safety measures, vehicle kinematics and macroscopic traffic data influences the models and their uncertainties.
* Performing real-world experiments to demonstrate that incorporating ‘looming’ provides low false alarm rates.

Therefore, the factors adopted not only increase sensitivity, but also consider all the interactions between the surrounding vehicles and uncertainties in vehicle states that are inherent in any real-world deployment. This enhances the model’s robustness and applicability to a wider spectrum of scenarios. These novel additions can be implemented in ADAS to predict vehicle-based safety threats and provide a great potential to increase driver compliance to ADAS as it lowers both the false alarms and missed alerts. ﻿

The rest of the paper is organised as follows: firstly, state-of-the-art literature on how the threat level is assessed in existing safety models is presented. The methodology of how the traffic conflict prediction models were developed to estimate the safety threat and how looming is incorporated is then presented. This is then followed by the presentation of the data used. Results from each model is presented and discussed. Finally, the last section summarises the main conclusions of the study.

# II. Literature Review

Existing quantification of potential traffic conflicts can be separated based on how the threat level is assessed in vehicle-based safety models [10] either by using physical model-based methods (learning from physical insight and models) or data-driven based methods (e.g. neural networks) [24], [25]. To foster proactive vehicle-based safety, a traffic conflict is utilised as a measure of crash nearness [26]. Traffic conflicts are characterised by a set of indicators; a spatio-temporal proximity indicator and an evasive action response that takes place so as to avoid a potential accident [27], [28].

## Physical model-based methods

Physical methods can be further divided into four categories based on their decision-making process. In some cases, the decision-making process is dependent on SSMs while in other cases the threat level can be estimated using a probabilistic approach or by optimising a problem. There are also situations where all possible occupancies of ego-vehicle as well as other traffic co-participants are known (e.g., lane change) and so the existence of a potential threat is known beforehand. In these cases, formal methods are applied.

*Safety Surrogate Measures (SSMs)*

SSMs are the most widely used method to provide insight on the identification of safety problems in real-world traffic environments. ﻿Existing SSMs assess whether risky interactions are present by detecting critical actions, such as severe braking manoeuvres and jerking before a dangerous interaction. These dangerous interactions are identified when the threshold value of a SSM is exceeded. SSMs are typically perception- or kinematic-based [4], [8], [16]. Perceptual algorithms are generally made up of time-based threat metrics. They are defined as the remaining time between a preceding and the ego-vehicle before an impact takes place. The most used metric for judging risk situations is TTC [29], [30]. TTC is computed as:

|  |  |
| --- | --- |
|  | (1) |

where and are the distance and the velocity difference between preceding and ego-vehicle respectively. However, TTC is based on several assumptions. To overcome some assumptions, the Modified-TTC (MTTC) [31] was introduced. MTTC is calculated based on a combination of relative velocities and accelerations in various car-following scenarios. When considering intersections, the Post Encroachment Time (PET) metric is the most suitable as there is no need to measure distance and velocity [32], [33]. A kinematic-based approach estimates the threats by calculating the difference between the required stopping distances of two consecutive vehicles, with full deceleration rate [16]. The calculated minimum distance is then compared to the current space headway e.g. Proportion of Stopping Distance (PSD) [34], [35], Deceleration Rate to Avoid the Crash (DRAC) [34]–[36] and Criticality Index Function (CIF) [37]. Literature reveals that kinematic metrics are more sensitive than time because of the deceleration rate, road conditions, speed and spacing headway [38].

*Optimisation Methods*

Dynamic optimisation is becoming a standard for decision-making processes. Model Predictive Control (MPC) is a prevalent optimisation framework in the literature [39]. It comprises an optimisation problem aiming to minimise an objective cost function while satisfying constraints [40]. An MPC-based model obtains the threat level by planning vehicle trajectories, estimating the threat introduced by this path and regulates the inputs to keep the threat below a given threshold. Anderson et. al. [41] used an optimal control framework to avoid collisions by relying on the required minimum front wheel angle. ﻿ However, ﻿their study did not adopt a driver model and failed to capture the predicted driver’s behaviors. Nevertheless, when using optimisation methods computational complications can occur when dealing with multi-vehicle collisions [24].

*Probabilistic Methods*

The underpinning idea behind probabilistic methods is the ability to determine with a certain level of confidence whether a situation can evolve into being a threat while coping with uncertainties. When applied to automation, the major uncertainty is the unknown future behaviors of traffic participants. Eidehall et al. [42] linked particular actions to a set of high-level manoeuvres (e.g. lane change). A probability distribution is assigned to a set of manoeuvres and each object is simulated to estimate the probability ﻿of a collision. Another probabilistic method is the Dynamic Bayesian Network (DBN). An estimate of the risk when approaching an intersection was obtained via DBN by Armand et al. [43]. Katrakazas et. al [44] also used DBN to identify risky situations but combined network with vehicle-level data. One disadvantage of probabilistic methods is that they cannot mathematically prove whether a planned manoeuvre is collision-free [45].

*Formal Methods*

The idea behind formal methods is that mathematical analysis can add to the reliability and robustness of a model and can be divided into logic-based and set-based. Logic-based focus mainly on the design of the model. These approaches formalise a requirement by converting it into logical sentences. A cooperative ADAS was designed based purely on logic by Damm et al. [46] to complete a collision-free driving task, while Rizaldi et al. [47] formalised traffic rules. In set-based methods, a class of acceptable performance are formalised to verify safety of vehicles. A hybrid collision avoidance system is designed for intersections made up of a scheduler assigning time slots to vehicles was developed by Kowshik et al. [48]. However, to determine whether the presence of a conflict can evolve to a safety threat using these methods is challenging to design, mainly due to the complicated requirements such as safety, sequential tasks and restrictiveness arguments [24]. ﻿

## Data-driven methods

Data-driven methods have the unique capability to learn from examples without being programmed to do so. ﻿Generally, prediction methods for threat assessment can be divided into two major categories: statistical techniques and Artificial Intelligence (AI) [19]. However in this study, AI techniques are the necessary choice over statistics because of their effectiveness to handle complex, nonlinear, highly imbalanced dataset [19]. Several studies have employed various techniques and this section considers mainly ﻿neural network models that have led to the development of Deep Learning [17], [49].

The main idea behind AI algorithms involves the analysis of rare patterns from several factors (e.g., speed, acceleration, distance) to predict whether the future state is safe. These can map nonlinear functions with no predefined assumptions and have demonstrated to produce human-competitive results [50]. Studies identified in the literature have made use of case-control designs when developing crash prediction models by using logistic regression models [51]–[53]. Support vector machines were also introduced to formulise the real-time crash risk prediction model [21]. However, when analysing complex real-time datasets, deep learning models are more appropriate than conventional machine learning models. DNNs were adopted by Formosa et al. [17] to classify and predict traffic conflicts. However, the main focus of [17] was to present a centralised digital architecture to unify the data and the vehicle’s movements were limited to one-dimension. Moreover, the input factors adopted for the DNN model to predict traffic conflicts were not explored to determine which factor majorly contributes to conflict predictions.

In conclusion, several methods are introduced in the literature to develop an algorithm to enhance safety by predicting traffic conflicts. Yet, there is still potential for further improvements as each method has different capabilities and weaknesses. In this paper data-driven methods using a deep learning approach are adopted for predicting vehicle-based safety threats because they capture complex nonlinear behaviors. It would otherwise be difficult to capture when using classical engineering principles [24]. Moreover, no consensus is reached as to what is needed as input factors to achieve optimal results. Therefore, this study develops multiple models with varying input factors to compare and analyse their sensitivity. Since SSMs are computationally cheap to design [24], input factors include perceptual- and/or kinematic-based SSMs. In other models, SSMs are incorporated with traffic variables within a hybrid model [17].

# Methodology

This section explains the modelling approach for predicting vehicle-based safety threats by incorporating new factors under uncertainty. A set of equations are developed to identify whether any surrounding objects are on a conflict course with any point of the ego-vehicle. Factors influencing traffic conflicts are presented together along with their estimation procedure to develop and train a prediction model. This section continues by presenting the performance metrics adopted for comparison purposes between the different models based on the optimal thresholds.

## Uncertainty Modelling

Generally, when estimating the conflict risk from a vehicle-based safety system, the estimates are usually computed for a specific path [54] such that the movement of vehicles are considered as one-dimensional and the kinematic variables are considered as scalars [55]. However, vehicle movements may not always follow the lanes of the road because they might be in the process of changing lanes (either discretionary or mandatory) or in some locations such as merging and diverging areas and a car park or an industrial area where lanes are not present [9], [13]. In these situations, conflict estimation still needs to be carried out and the driver needs to be aware of non-standard driving and additional vehicles other than the leading vehicle can pose a threat to the ego-vehicle. Therefore, estimating the safety threat in a 2D space based on the relative motion between vehicles is crucial for safety models.

*Looming*

The looming process adopted in this study allows the developed models to be robust in measuring uncertainty, especially for vehicles that travel in a non-standard pattern [9], [13]. This process determines whether an object (e.g., a stationary object or a vehicle) is on a conflict course with any points of the vehicle’s 2D footprint. In fact, there can exist situations where the object is on a conflict course with certain points of the vehicle’s 2D footprint but not with other points. Therefore, four loom points (i.e., the front right and left extremes of ego-vehicle’s 2D footprint and the furthest left and right points on the detected object) must be considered when determining if a conflict course exists. The looming process is introduced briefly in this section but the readers are referred to [9], [13] for a comprehensive review of this process.

The algorithm identifies surrounding vehicles irrespective of their size. A buffer area surrounding each detected vehicle (presented in Fig. 1) is introduced to include the vehicle’s footprint and a recovery space surrounding it. The addition of the buffer stems from the fact that the drivers tend to keep a certain safety distance from other objects [54]. If omitted, drivers may perceive it as an invasion from other objects [54]. By adding this extra space in the algorithm, it makes drivers more compliant to ADAS as any decisions stemming from the models are consistent with driver perceptions.

The algorithm works by estimating which vehicle might cause a conflict risk to the ego-vehicle based on a loom angle, bearing angle, and yaw rate. Loom angle is the angle between the furthest left and right points (of target vehicle plus buffer) from the loom points on the ego-vehicle. The bearing angle () is defined as the azimuth angle of the line that connects the velocity vector at the ego-vehicle's loom point and the closest point of the target vehicle including the buffer area. The yaw rate () is the change of the heading angle of the ego-vehicle to determine its direction.

To determine if a conflict course exists with the ego-vehicle, the measurements of the buffer area and the estimation of specific measurements for every object detected are required as presented in Fig. 1. These include loom angle (obtained from andwhich refer to the angle between the velocity vector of the ego-vehicle and the furthest left and right of the preceding vehicle respectively), bearing angle, and distance .

Diagram

Description automatically generated with medium confidence**Fig. 1.** Illustration of angles , and to estimate the conflict risk.

Additional parameters such as length and width of the object detected, lateral and longitudinal displacements, relative lateral and longitudinal speeds and relative longitudinal acceleration between an object and the ego-vehicle are also required for the conflict course estimation. From Fig. 1., the side of the ego-vehicle where the conflicting object is detected is essential. This is because the equations for each angle changes if a conflicting object is present on the left- or right-hand side. Considering that the preceding vehicle is on the left-hand side of the ego-vehicle the following equations are derived using trigonometry:

|  |  |
| --- | --- |
|  | (2) |
|  | (3) |
|  | (4) |
|  | (5) |

where n refers to vehicle ID, t refers to specific timestamp, and are the widths of the ego-vehicle and the preceding vehicle respectively, is the length of the preceding vehicle, and represent the longitudinal and lateral displacement between both vehicles respectively and and represent the buffer area surrounding the preceding vehicle which depends on the ego-vehicle's speed. The loom angle is estimated by subtracting from :

|  |  |
| --- | --- |
|  | (6) |

If the vehicle is on the right-hand side of the ego-vehicle, is the same as and is the same as . A total of 18 conflict combination course are considered to determine whether a conflict course exists. A systematic approach in which a combination of loom angle (), bearing angle (), and yaw rate () were increased, decreased, or kept constant was employed. From the different combinations, a traffic conflict can occur if the distance between the preceding vehicle and the ego-vehicle decreases. This is because as the distance decreases, the loom angle also increases. Therefore, no safety threat is present if the loom angle decreases or does not change. Nevertheless, there is still no guarantee for a conflict if the loom angle increases. This highlights the importance of considering the combination of different angles. All combinations[[2]](#footnote-2)1 are presented in the footnote. However, only four conflict courses (combinations of the , and ) show that a conflict course takes place, and these are presented in Fig. 2. Fig. 2 also clearly visualises and explains the cases on the straight and curved roads and how these are estimated.

## Input factors to DNN models

The development of the DNN requires both the output (i.e., traffic conflicts) and the input (influencing factors). Studies were conducted to determine which influencing factors best represent traffic conflicts [26], [27], [56], each providing different conclusions about their effectiveness. Therefore, this study develops 12 models with varying input factors to compare their prediction ability and determine which factors are beneficial for traffic conflict predictions. These include the traditional safety models using TTC only, perceptual-based, kinematic-based, all SSMs, macroscopic traffic data, vehicle’s dynamic state data and hybrid approach. These models are also compared when using traditional estimation and using looming.

*(i) Safety Surrogate Measures*

SSMs are good indicators to assess the conflict risk and suitable tools to identify and predict unsafe circumstances [33], [36], [57], [58]. They can show the states between safety and a conflict and therefore, can show the potential conflict causality and mechanism. In this study, a total of six SSMs are considered as input factors, three are perception-based (TTC, MTTC, PET) and three kinematic-based are employed (PSD, DRAC, CIF).

*(ii) Macroscopic Traffic Data*

Although the association between macroscopic information from infrastructure sensors (e.g. Inductive Loop Detectors (ILD)) and conflict risk has been explored in some studies, it is rarely incorporated in safety models [5].Traffic variables can act as a proxy for the traffic condition indicating any hazardous traffic conditions which can result in conflicts. The traffic variables include speed, flow, density, occupancy and headway [59]. The speed variance between lanes can also be estimated

Diagram

Description automatically generated

**Fig. 2.** Summary of potential traffic conflict courses based on the loom angle, bearing angle and the ego-vehicle yaw rate showing loom points on vehicle.

from this data. All these six input factors together with their means and standard deviation (SD) across all lanes are adopted. These were estimated either specific by lane or averaged over a road section.

Recent literature shows that traffic variables from ILD do influence the performance of traffic conflict predictions [17]. Nevertheless, literature shows that using macroscopic data only in conflict prediction, is not beneficial since the conflict-risk is estimated at the aggregated level for each road section and the variability of each driver on the road is not captured.

(iii) *Vehicle’s Dynamic State*

Vehicle’s dynamic state data can be used to estimate collision risk. The vehicle's speed and the yaw rate are collected from the vehicle speed sensor, and the gyroscope respectively by scanning the controller area network bus of the vehicle. These two input factors and the rate of change of speed are also employed in this study. However, using only this data results in limiting the conflict prediction accuracy and applicability [5].

## DNN Development

The input factors are fed into the DNN in the form of a vector at 15Hz. Using an optimisation algorithm (e.g., *Adam*), the DNN is trained to learn and adjust the weights within the network to predict and identify patterns from the data which result in a traffic conflict. Every DNN model is finely tuned by adjusting the hyperparameters. The hyperparameters govern how the model fits the data to maximise the model’s predictive accuracy and sensitivity [60], [61].

Several DNN models were trained using *Adam* as an optimiser, by varying number of hidden layers, nodes, initial learning rate value, epoch number and activation functions. A total of 360 DNN models (3 x 3 x 2 x 5 x 4 = 360) for each varying input factor model were tested and shown in Table I.

TABLE I

Hyperparameters tuning for DNN

|  |  |
| --- | --- |
| **Hyperparameter** | **Range** |
| Hidden Layers | 2, 3, 4, 5, 10 |
| No. of nodes | 5, 10, 50, 100 |
| Learning Rate | 0.0001, 0.001, 0.01 |
| Epoch Number | 50, 75, 100 |
| Activation Functions | *tanh,* ReLU |

To evaluate and determine which factors generated the best model, the area under curve (AUC) value is used. This value is used to assess the predictive performance of the model. Each model generates a confusion matrix which provides four values: true positive, false negative, false positive, and true negative [62]. From these values, sensitivity (accuracy of traffic conflict predictions out of all traffic conflict events), accuracy (﻿if both ‘safe traffic dynamics’ and ‘traffic conflict’ cases are predicted correctly) and False Alarm Rates (FAR) (the fraction of safe conditions predicted as threats) can be calculated. To validate the performance of the model, K-fold cross-validation is used.

Nevertheless, selecting an optimal threshold is an essential step. This step selects the cut-off point of the posterior probability to identify a traffic conflict against safe traffic dynamics. The optimal threshold value is obtained using the maximum between-class variance method and based on this threshold value, each model with different input factors is tested to determine which influencing factors obtain the highest sensitivity. This was carried out to avoid crude and subjective judgements in threshold tuning [63].

# Data

The microscopic data was collected using an instrumented vehicle in real-time equipped with (i) PointGrey Grasshopper3 Near InfraRed Camera (ii) ARS 308-21 Continental automotive radar (iii) U-blox NEO M8-L GNSS and an Arduino microcontroller. The Arduino microcontroller was used to collect (a) the vehicle speed and its yaw rate via the bus of the vehicle and (b) the weather data from a temperature, pressure, humidity, precipitation, and anemometer sensors.

These sensors can detect vehicles on a straight and a curved road and provide detailed information of individual vehicles on the road such as speed, acceleration, distance between vehicles, length and widths of vehicles, position, weather conditions at 15 Hz. Real-world data was collected with approximately 19 hours of data comprising around 910,000 observations (850,353 safe traffic dynamics and 59,657 traffic conflicts) from junction 17-23 of the M1 motorway in the UK. A total of 15 trips were conducted at different times from multiple drivers in various traffic conditions to reflect events in the real-world.

Cost-sensitive learning was used to account for data imbalance in which weights based on the imbalance ratio were assigned for misclassifying safe from safety threats [64]. Despite the cost setting based on ratios might be crude, [65] backed the use of such to tackle cost sensitive problems.

The weather data and vehicle dynamic state were collected directly from sensors within the instrumented vehicle. The macroscopic data were collected from ILD along the testbed. Using equations developed in Section III, the edges, and parameters of the vehicles in a conflict course with the ego-vehicle are estimated using data collected from the radar. Based on these values, the SSMs presented in section III were estimated with looming (handling the uncertainties in the data) and the traditional way. It is important to note that looming is incorporated within the model only when estimating SSMs. This is because the estimation of other factors is not affected if another object poses a threat to any point on the ego-vehicle.

A challenge arising from the development of a traffic conflict prediction algorithm was that traffic conflicts data was not readily available. Therefore, a new method had to be developed to automatically extract traffic conflicts using video data collected from an instrumented vehicle. This method had to be independent of the data used as input, otherwise the correlation values would be high, and the results would be biased. This method aimed to identify vehicle-level conflicts (used as an output variable) using image processing techniques and a Faster R-CNN network where vehicles were detected, tracked, and classified accordingly. The lane geometry data was also detected from an in-vehicle perspective. ﻿Traffic conflicts were identified through response to an evasive action, and the temporal and/or spatial proximity [56] and the determining criteria were based on time, distance, speed, braking, acceleration, and deceleration. These conflicts were manually validated and further validated by developed equations. The readers are referred to Formosa et al. [17], [28] for a more thorough description of the traffic conflict extraction.

The relevant input factors for each model were selected via Extra Tree classifiers and Extreme Gradient Boosting. Correlation was also carried out. If a feature is highly correlated with another feature (the Pearson correlation value greater than 0.6), the relative importance between the correlated pairs is evaluated and the less important variable was eliminated. From a series of 32 features, 20 features were included in the final dataset in descending order of importance: PET, PSD, TTC, SpeedVariance, SpeedCar, AccelerationCar, DRAC, MTTC, CIF, Lane, SDFlow, SDSpeed, SDOccupancy, SDDensity, Headway, Yaw, Precipitation, Temperature, WindDirection, WindSpeed. Each variable was standardised over the whole dataset as this speeds up the learning process and leads to a faster convergence [25]. The dataset was then randomly partitioned into 70%-30% for training and testing data respectively as it provides the best trade-off.

# Results

A total of 12 conflict prediction models with varying input factors were developed. The input factors for each model include (1) Vehicle’s Dynamic States - *SpeedCar, AccelerationCar, Yaw*, (2) Macroscopic traffic data - *Speed, Flow, Density, Occupancy, Headway, Speed Variance and their average and standard deviation parameters*, (3) TTC only, (4) Perceptual SSMs – *TTC, MTTC, PET*, (5) Kinematic SSMs – *PSD, DRAC,* *CIF*, (6) All SSMs – *TTC, MTTC, PET, PSD, DRAC, CIF* and (7) Hybrid – *All SSMs, Vehicle’s Dynamic states, Macroscopic Data and Weather data.* Models (3) to (7) were traditionally estimated as well as using looming resulting in 12 models. The hyperparameters for each model were set prior to the learning process and tuned to optimise the learning process. A grid search was conducted and the optimal hyperparameters for each model are presented in Table II

TABLE II

Optimal hyperparameters for each prediction model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input factors** | **Layers** | **No. of nodes** | **Learning**  **Rate** | **Epoch**  **No.** | **Activation**  **Function** |
| Vehicle’s dynamic states | 3 | 10 | 0.01 | 50 | *tanh* |
| Macroscopic | 4 | 10 | 0.001 | 50 | *tanh* |
| TTC only | 2 | 5 | 0.001 | 50 | *tanh* |
| Perceptual SSMs | 3 | 25 | 0.001 | 75 | *tanh* |
| Kinematic SSMs | 3 | 25 | 0.001 | 75 | *tanh* |
| All SSMs | 4 | 50 | 0.001 | 75 | *tanh* |
| Hybrid | 4 | 100 | 0.0001 | 100 | *tanh* |

From Table II it is interesting to note that as the input factors increase, almost all the hyperparameter values increase accordingly while the learning rate decreases. The AUC values from the training data using a 5-fold cross validation, and from the test data for each model are presented in Table III. Comparative analyses between each prediction model were carried out. The factors included in each of the models were varied to determine which ones are the most suitable in providing a high predictive performance.

TABLE III

AUC values for each DNN model with varying input factors

|  |  |  |
| --- | --- | --- |
|  | **AUC values** | |
| **Input Factors** | **Training Data** | **Test Data** |
| Vehicle’s dynamic states | 0.629 s.d. (0.011) | 0.623 |
| Macroscopic data | 0.736 s.d. (0.075) | 0.734 |
| **Traditional estimation** | | |
| TTC only | 0.617 s.d. (0.009) | 0.615 |
| Perceptual SSMs | 0.637 s.d. (0.012) | 0.636 |
| Kinematic SSMs | 0.705 s.d. (0.007) | 0.702 |
| All SSMs | 0.738 s.d. (0.011) | 0.736 |
| Hybrid | 0.902 s.d. (0.007) | 0.897 |
| **Estimation of factors with looming** | | |
| TTC only | 0.656 s.d. (0.006) | 0.648 |
| Perceptual SSMs | 0.701 s.d. (0.003) | 0.707 |
| Kinematic SSMs | 0.750 s.d. (0.012) | 0.749 |
| All SSMs | 0.805 s.d. (0.012) | 0.792 |
| Hybrid | 0.947 s.d. (0.001) | 0.941 |

Table III shows how the performance of the models increased with the addition of new factors and introducing looming into a model. It can be observed that the hybrid algorithm provided the highest AUC values. More specifically, integrating the looming into the model resulted in a further 5% increase in the AUC value relative to an equivalent traditional hybrid model. As expected, the model with TTC only provided the lowest overall AUC value. It is also noteworthy to state that the kinematic-based SSMs offered a higher AUC value than the perceptual SSMs as expected from the literature [38]. Two other models were developed using the vehicle’s dynamic states and macroscopic traffic data as factors respectively. These models also offered a higher AUC value than when using TTC only. In summary, it can be concluded that the performance of the hybrid model was increased by around 45% relative to the model with TTC only (with or without looming).

To compare models with different input factors optimal threshold selection for each model was carried out using minimum between-class variance method. The results from each model are presented in Table IV.

TABLE IV

Sensitivity, FAR and accuracy values for each model with varying input factors using traditional estimation and looming

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input factors** | **Optimal threshold** | **Sensitivity** | **FAR** | **Accuracy** |
| Vehicle’s dynamic state | 0.620 | 0.758 | 0.449 | 0.483 |
| Macroscopic | 0.605 | 0.792 | 0.229 | 0.758 |
| **Traditional estimation** | | | | |
| TTC only | 0.578 | 0.289 | 0.096 | 0.857 |
| Perceptual SSMs | 0.640 | 0.262 | 0.034 | 0.913 |
| Kinematic SSMs | 0.543 | 0.392 | 0.044 | 0.926 |
| All SSMs | 0.595 | 0.475 | 0.093 | 0.887 |
| Hybrid | 0.402 | 0.816 | 0.098 | 0.876 |
| **Estimation of factors with looming** | | | | |
| TTC only | 0.590 | 0.331 | 0.064 | 0.869 |
| Perceptual SSMs | 0.639 | 0.383 | 0.047 | 0.905 |
| Kinematic SSMs | 0.462 | 0.498 | 0.048 | 0.932 |
| All SSMs | 0.545 | 0.547 | 0.069 | 0.915 |
| Hybrid | 0.389 | 0.868 | 0.106 | 0.897 |

It can also be observed that when incorporating the looming method, each model obtains consistently higher sensitivity values for a slight increase in FAR. Sensitivity is arguably more important as missing a conflict identification could be more severe. Overall, the hybrid model obtained the best results as the sensitivity is the highest for a FAR acceptable by industry [66]. When considering the sensitivities of the other models, they are relatively low for similar FAR. Nevertheless, the model using macroscopic traffic data compared obtained a higher sensitivity for a 20% FAR. This highlights the importance of adding macroscopic data to the model as it adds information about the dynamic surroundings and consider contextual risk factors. Following this, the vehicle’s dynamic state factors also show their influence on the conflict prediction.

# Discussion

In this research, multiple models with varying input factors were developed using deep learning to predict traffic conflicts from real-time data. Deep learning allows the prediction models to be robust to small inaccuracies and particularities of the network such that they can learn the concept on multiple layers or dimensions of data inputs thereby increasing complexity and intricacy of the networks via a web of interconnected relationships. This approach is more beneficial rather than a model based on assumptions and conservative criteria and is essential for real-world deployment of such systems.

The best traffic conflict prediction results were obtained when the model consists of a hybrid of input factors. While most factors can easily be estimated (i.e., vehicle’s dynamic states) or are readily available (i.e., macroscopic traffic data) to compute SSMs using looming, a high installation sensor cost is required. A cost-effective solution is to use a radar-vision framework which obtains precise measurements and is a cheaper option. Moreover, as the penetration rate of intelligent vehicles increases, this would not incur a large additional cost.

It is also important that models used to predict conflicts do not to rely on one factor as this can be estimated from one sensor, and it is known that sensors are prone to failure. Therefore, it is essential to have multifarious factors from multiple sensors to gather essential dynamic information. This results in a more trustworthy and safe system due to redundancy, an increase in integrity, reliable data outcomes, robustness of the system, sensor failure detection while decreasing the errors related to the sensor systems [67]. Moreover, in particular instances, some factors cannot always offer the optimal timing for conflict warning due to assumptions and the nature of their equation. For example, the TTC equation can result in a missed alert if the gap between vehicles is small and the change in velocity is even smaller. In these instances, the conflict prediction model using TTC only as the input factor does not have full knowledge of the dynamic surroundings. When considering the macroscopic data from the inductive loop detectors as model inputs, this data may be of low quality with missing data and noise. Moreover, the frequency of the data collection from ILD is below 1s. For real-time application, this could be improved even further. A model considering vehicle’s dynamic states only as factors, calculates the risk based only on the vehicle ahead. As a result, to improve safety and accuracy, the hybrid of factors provides more advantages.

Some studies in the literature consider only the vehicles ahead within the same lane of the ego-vehicle [13]. However, not all situations are car-following and not all vehicles obey the traffic rules [9]. Other studies account for lane change events from macroscopic data. This is problematic as conflict-risk is estimated at the aggregated level for each road section and the variability of each driver is ignored. In this study, lane change events are captured by models when incorporating looming. From the developed equations if any point of the ego-vehicle is affected by a lane change event, the model estimates the safety threat even if the threat is not yet in the same lane. Consequently, this novel addition also includes the variability of each driver which is pivotal to improve the accuracy of conflict prediction [5].

When considering the optimal threshold for each model, it can be observed that when using less factors as input, the sensitivity of the systems is relatively low compared to the hybrid model. This is not ideal for a real-time application as the FAR should be kept ‘low’ as much as possible otherwise the drivers would ignore the warning systems. However, an improvement is observed for each model using looming. Moreover, by incorporating looming, the application of the system is extended for a wider spectrum of traffic scenarios.

It is well known that predictability drops with the increase in prediction time horizon. Predicting short-term conflicts with considerable accuracy and reliability is critically important. For CAVs, modern intelligent transportation systems and vehicular communication it is essential that conflict prediction models utilise disaggregated data and aim to reduce the prediction horizon to less than 1 second [68]. The average human reaction time is 0.8 seconds [69], therefore, to further enhance safety, the prediction horizon in this work aims to predict conflicts within the next 0.2 seconds in the future. In the case of a conflict, the vehicle receives this information and modifies its trajectories and maneuvers accordingly with the latest updated information gathered by sensors.

One particular limitation of this research is the tuning of the hyperparameters because these are selected based on a specific dataset. As a result, if a new dataset is applied to the trained models, the hyperparameters may need to be re-calibrated. However, spatial transferability to similar roads such as motorways is viable. It is also important to note that the developed prediction models rely on the data quality employed, i.e., poor data quality leads to low data utilisation efficiency and serious decision-making mistakes [70]. This paper ensures excellent quality by making use of high capability sensors to collect data and thorough pre-processing and analysis of data to draw valid and applicable conclusions. Another limitation of these models is that deep learning methods can be considered as ‘black-box’ methods. Nevertheless, pre-processing of the data can identify the important input factors and the correlations would help in this regard.

Additionally, while the proposed DNN models provided an improvement in the prediction accuracy, ﻿they can be further enhanced by (i) further adjusting the combination of the number of nodes in the layers where the DNN model is expected to achieve a better performance and (ii) increasing the size of the ﻿dataset the predictability of these models is expected to increase. By increasing the size of the dataset, the chance of overfitting and generalisation error decreases as the models are trained with more examples [23]. By adding more input features the probability of overfitting increases because some features may be redundant or irrelevant, but this can be mitigated by applying feature selection. Further input factors which can be considered include Time-To-React [71], road design (e.g., road inclination) and driver information.

# Conclusion

This paper developed deep learning models that integrate new factors and the uncertainties associated with the estimates of vehicle states using the concept of ‘looming’ to detect the level of safety threat associated with an autonomous driving system. The models took the full advantage of the joint benefits offered by microscopic and macroscopic variables in predicting safety threats in an organically changing operational environment. Conflict detection was based on the measurements in a two-dimensional space that include loom angle, bearing angle and yaw rate of an ego-vehicle with respect to a preceding vehicle irrespective of their size. Real-world experiments confirmed that adding multiple factors improved the performance of the model by about 45% in relation to the model consisting of TTC only. The inclusion of looming into the model further enhanced the performance by 5%. This finding is significant to vehicle manufacturers as most autonomous driving safety systems (e.g., ADAS) rely on only one factor which is TTC. Consequently, the FAR from existing systems is relatively high exposing their deficiencies in system integrity.

From the comparative analysis, the hybrid models developed in this paper show a significant potential to incorporate in existing autonomous driving systems. This is because the model becomes more proactive and reliable with a high sensitivity value (e.g., 87%) for a low FAR (e.g., 10%) at the optimal threshold derived systematically. In the CAV era, the developed hybrid model can be combined with the cloud to provide more accurate traffic conflict predictions using vehicle-to-vehicle and vehicle-to-infrastructure (V2X) communications.

Acknowledgment

The authors take full responsibility for the content of this paper and any errors or omissions.

References

[1] F. Biondi, D. L. Strayer, R. Rossi, M. Gastaldi, and C. Mulatti, “Advanced driver assistance systems: Using multimodal redundant warnings to enhance road safety,” *Appl. Ergon.*, pp. 238–244, 2017.

[2] X. Wang, J. Liu, T. Qiu, C. Mu, C. Chen, and P. Zhou, “A Real-Time Collision Prediction Mechanism with Deep Learning for Intelligent Transportation System,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9497–9508, 2020.

[3] Y. Zhang, E. K. Antonsson, and K. Grote, “A new threat assessment measure for collision avoidance systems,” *IEEE Intell. Transp. Syst. Conf.*, pp. 968–975, 2006.

[4] X. Wang, M. Chen, M. Zhu, and P. Tremont, “Development of a Kinematic-Based Forward Collision Warning Algorithm Using an Advanced Driving Simulator,” *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 9, pp. 2583–2591, 2016.

[5] S. Tak, S. Woo, and H. Yeo, “Study on the framework of hybrid collision warning system using loop detectors and vehicle information,” *Transp. Res. Part C Emerg. Technol.*, vol. 73, pp. 202–218, 2016.

[6] K. Vogel, “A comparison of headway and time to collision as safety indicators,” *Accid. Anal. Prev.*, vol. 35, no. 3, pp. 427–433, 2003.

[7] M. C. Panou, “Intelligent personalized ADAS warnings,” *Eur. Transp. Res. Rev.*, vol. 10, no. 2, 2018.

[8] M. Zhu, X. Wang, and J. Hu, “Impact on car following behavior of a forward collision warning system with headway monitoring,” *Transp. Res. Part C Emerg. Technol.*, vol. 111, pp. 226–244, 2020.

[9] J. Ward, G. Agamennoni, S. Worrall, and E. Nebot, “Vehicle collision probability calculation for general traffic scenarios under uncertainty,” *IEEE Intell. Veh. Symp. Proc.*, no. Iv, pp. 986–992, 2014.

[10] Y. Li, Y. Zheng, B. Morys, S. Pan, J. Wang, and K. Li, “Threat Assessment Techniques in Intelligent Vehicles: A Comparative Survey,” *IEEE Intell. Transp. Syst. Mag.*, 2020.

[11] R. E. Stern *et al.*, “Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments,” *Transp. Res. Part C Emerg. Technol.*, vol. 89, no. April 2017, pp. 205–221, 2018.

[12] M. Makridis *et al.*, “Empirical Study on the Properties of Adaptive Cruise Control Systems and Their Impact on Traffic Flow and String Stability,” *Transp. Res. Rec.*, vol. 2674, no. 4, pp. 471–484, 2020.

[13] J. R. Ward, G. Agamennoni, S. Worrall, A. Bender, and E. Nebot, “Extending Time to Collision for probabilistic reasoning in general traffic scenarios,” *Transp. Res. Part C Emerg. Technol.*, vol. 51, pp. 66–82, 2015.

[14] T. Sayed, M. H. Zaki, and J. Autey, “Automated safety diagnosis of vehicle-bicycle interactions using computer vision analysis,” *Saf. Sci.*, vol. 59, pp. 163–172, 2013.

[15] C. Lu, X. He, H. van Lint, H. Tu, R. Happee, and M. Wang, “Performance evaluation of surrogate measures of safety with naturalistic driving data,” *Accid. Anal. Prev.*, vol. 162, no. May, p. 106403, 2021.

[16] S. Tak, S. Kim, D. Lee, and H. Yeo, “A comparison analysis of surrogate safety measures with car-following perspectives for advanced driver assistance system,” *J. Adv. Transp.*, 2018.

[17] N. Formosa, M. Quddus, S. Ison, M. Abdel-Aty, and J. Yuan, “Predicting real-time traffic conflicts using deep learning,” *Accid. Anal. Prev.*, vol. 136, 2020.

[18] C. Katrakazas, M. Quddus, and W.-H. Chen, “A simulation study of predicting conflict-prone traffic conditions in real-time,” *96th Annu. Meet. Transp. Res. Board (trb),washingt. D.C., USA*, pp. 1–12, 2017.

[19] M. G. Karlaftis and E. I. Vlahogianni, “Statistical methods versus neural networks in transportation research: Differences, similarities and some insights,” *Transp. Res. Part C Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, 2011.

[20] M. Hossain and Y. Muromachi, “Understanding crash mechanism on urban expressways using high-resolution traffic data,” *Accid. Anal. Prev.*, vol. 57, pp. 17–29, 2013.

[21] R. Yu and M. Abdel-Aty, “Utilizing support vector machine in real-time crash risk evaluation,” *Accid. Anal. Prev.*, vol. 51, pp. 252–259, 2013.

[22] C. Xu, W. Wang, and P. Liu, “A genetic programming model for real-time crash prediction on freeways,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 574–586, 2013.

[23] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, “Deep learning applications and challenges in big data analytics,” *J. Big Data*, vol. 2, no. 1, pp. 1–21, 2015.

[24] J. Dahl, G. R. De Campos, C. Olsson, and J. Fredriksson, “Collision avoidance: A literature review on threat-assessment techniques,” *IEEE Trans. Intell. Veh.*, vol. 4, no. 1, pp. 101–113, 2019.

[25] A. Theofilatos, C. Chen, and C. Antoniou, “Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction,” *Transp. Res. Rec.*, vol. 2673, no. 8, pp. 169–178, 2019.

[26] A. P. Tarko, “A unifying view on traffic conflicts and their connection with crashes,” *Accid. Anal. Prev.*, vol. 158, no. January, p. 106187, 2021.

[27] L. Zheng, K. Ismail, and X. Meng, “Traffic conflict techniques for road safety analysis: open questions and some insights,” *Can. J. Civ. Eng.*, vol. 41, no. 7, pp. 633–641, 2014.

[28] N. Formosa, M. Quddus, A. Papadoulis, and A. Timmis, “Validating a Traffic Conflict Prediction Technique for Motorways Using a Simulation Approach †,” *Sensors*, vol. 22, no. 556, 2022.

[29] Y. Matsui, M. Hitosugi, K. Takahashi, and T. Doi, “Situations of Car-to-Pedestrian Contact,” *Traffic Inj. Prev.*, vol. 14, pp. 73–77, 2013.

[30] X. Qu, Y. Kuang, E. Oh, and S. Jin, “Safety Evaluation for Expressways: A Comparative Study for Macroscopic and Microscopic Indicators,” *Traffic Inj. Prev.*, vol. 15, no. 1, pp. 89–93, 2014.

[31] K. Ozbay, H. Yang, B. Bartin, and S. Mudigonda, “Derivation and Validation of a New Simulation-based Surrogate Safety Measure,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2083, pp. 105–113, 2008.

[32] A. Van der Horst, *A Time-based Analysis of Road User Behaviour in Normal and Critical Encounters*. 1990.

[33] J. Archer, “Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling: A study of urban and suburban intersections,” Royal Institute of Technology, Stockholm, Sweden, 2005.

[34] V. Astarita, G. Guido, A. Vitale, and V. Giofré, “A new microsimulation model for the evaluation of traffic safety performances,” *Eur. Transp. - Trasp. Eur.*, no. 51, pp. 1–16, 2012.

[35] G. Guido and F. Saccomanno, “Comparing safety performance measures obtained from video capture data,” *J. Transp. Eng.*, vol. 137, no. July, pp. 481–491, 2010.

[36] F. Saccomanno, F. Cunto, G. Guido, and A. Vitale, “Comparing Safety at Signalized Intersections and Roundabouts Using Simulated Rear-End Conflicts,” *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2078, pp. 90–95, 2008.

[37] C.-Y. Chan, “Defining safety performance measures of driver-assistance systems for intersection left-turn conflicts,” *Intell. Veh. Symp. IEEE*, pp. 25–30, 2006.

[38] S. Tak, S. Kim, and H. Yeo, “Development of a Deceleration-Based Surrogate Safety Measure for Rear-End Collision Risk,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2435–2445, 2015.

[39] P. Falcone and F. Borrelli, “Low complexity MPC schemes for integrated vehicle dynamics control problems,” *9th Int. Symp. Adv. Veh. Control*, pp. 875–880, 2008.

[40] F. Borrelli, A. Bemporad, and M. Morari, *Predictive control for linear and hybrid systems*. Cambridge: Cambridge University Press, 2017.

[41] S. J. Anderson, S. C. Peters, T. E. Pilutti, and K. Iagnemma, “Experimental study of an optimal-control-based framework for trajectory planning, threat assessment, and semi-autonomous control of passenger vehicles in hazard avoidance scenarios,” *Springer Tracts Adv. Robot.*, vol. 62, pp. 59–68, 2010.

[42] A. Eidehall and L. Petersson, “Statistical threat assessment for general road scenes using Monte Carlo sampling,” *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 137–147, 2008.

[43] A. Armand, D. Filliat, and J. Ibanez-Guzman, “A Bayesian framework for preventive assistance at road intersections,” *IEEE Intell. Veh. Symp. Proc.*, pp. 1128–1134, 2016.

[44] C. Katrakazas, “Developing an advanced collision risk model for autonomous vehicles,” 2017.

[45] M. Althoff and S. Magdici, “Set-based prediction of traffic participants on arbitrary road networks,” *IEEE Trans. Intell. Veh.*, vol. 1, no. 2, pp. 187–202, 2016.

[46] W. Damm, H.-J. Peter, J. Rakow, and B. Westphal, “Can we build it: Formal synthesis of control strategies for cooperative driver assistance systems,” in *Mathematical Structures in Computer Science*, 2013, pp. 676–725.

[47] A. Rizaldi and M. Althoff, “Formalising Traffic Rules for Accountability of Autonomous Vehicles,” *IEEE Conf. Intell. Transp. Syst.*, pp. 1658–1665, 2015.

[48] H. Kowshik, D. Caveney, and P. R. Kumar, “Provable systemwide safety in intelligent intersections,” *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 804–818, 2011.

[49] D.-J. Lin, M.-Y. Chen, H.-S. Chiang, and P. K. Sharma, “Intelligent Traffic Accident Prediction Model for Internet of Vehicles With Deep Learning Approach,” *IEEE Trans. Intell. Transp. Syst.*, 2021.

[50] J. R. Koza, M. A. Keane, M. J. Streeter, W. Mydlowec, J. Yu, and G. Lanza, *Genetic Programming IV: Routine Human-Competitive Machine Intelligence;* 2006.

[51] M. Abdel-Aty, N. Uddin, and A. Pande, “Split Models for Predicting Multivehicle Crashes During High-Speed and Low-Speed Operating Conditions on Freeways,” *Transp. Res. Rec. J. Transp. Res. Board*, 2005.

[52] M. Abdel-Aty, N. Uddin, F. Abdalla, A. Pande, and L. Hsia, “Predicting freeway crashes based on loop detector data using matched case–control logistic regression,” *Transp. Res. Board 1897*, pp. 88–95, 2004.

[53] M. Ahmed and M. Abdel-Aty, “The viability of using automatic vehicle identification data for real-time crash prediction,” *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 459–468, 2012.

[54] J. Hou, G. F. List, and X. Guo, “New Algorithms for Computing the Time-to-Collision in Freeway Traffic Simulation Models,” vol. 2014, 2014.

[55] M. Brännström, E. Coelingh, and J. Sjöberg, “Model-based threat assessment for avoiding arbitrary vehicle collisions,” *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 658–669, 2010.

[56] A. P. Tarko, *Measuring road safety using surrogate events*. Emerald Publishing Limited, 2020.

[57] A. Tarko, D. G., N. Saunier, T. Sayed, and S. Washington, “White Paper: Surrogate Measures of Safety,” 2009.

[58] A. Sobhani, W. Young, S. Bahrololoom, and M. Sarvi, “Calculating time-to-collision for analysing right turning behaviour at signalised intersections,” *Road Transp. Res.*, vol. 22, no. 3, pp. 49–61, 2013.

[59] C. Wang, M. A. Quddus, and S. G. Ison, “The effect of traffic and road characteristics on road safety: A review and future research direction,” *Saf. Sci.*, vol. 57, pp. 264–275, 2013.

[60] G. Melis, C. Dyer, and P. Blunsom, “On the state of the art of evaluation in neural language models,” *6th Int. Conf. Learn. Represent.*, pp. 1–10, 2018.

[61] F. Hutter and L. Kotthoff, “Hyperparameter Optimization,” in *Automated machine learning*, 2019, pp. 409–411.

[62] N. Aung, W. Zhang, S. Dhelim, and Y. Ai, “Accident prediction system based on hidden Markov model for vehicular Ad-Hoc network in urban environments,” *Inf.*, vol. 9, no. 12, 2018.

[63] K. Yang, R. Yu, X. Wang, M. Quddus, and L. Xue, “How to determine an optimal threshold to classify real-time crash-prone traffic conditions?,” *Accid. Anal. Prev.*, vol. 117, no. April, pp. 250–261, 2018.

[64] A. Fernández, S. García, M. Galar, and R. C. Prati, *Learning from Imbalanced Data Sets*. Springer International Publishing, 2018.

[65] A. Fernández, M. Galar, S. García, B. Krawczyk, R. Prati, and F. Herrera, *Learning from Imbalanced Data*. Springer, 2018.

[66] F. Basso, L. J. Basso, and R. Pezoa, “The importance of flow composition in real-time crash prediction,” *Accid. Anal. Prev.*, vol. 137, no. January, p. 105436, 2020.

[67] C. Premebida, G. Monteiro, U. Nunes, and P. Peixoto, “A Lidar and vision-based approach for pedestrian and vehicle detection and tracking,” *IEEE Conf. Intell. Transp. Syst. Proceedings, ITSC*, pp. 1044–1049, 2007.

[68] C. Katrakazas, M. A. Quddus, and W.-H. Chen, “A new methodology for collision risk assessment of autonomous vehicles,” *Transp. Res. Board 96th Annu. Meet.*, 2017.

[69] M. Green, “‘How Long Does It Take to Stop?’ Methodological analysis of driver perception-brake times,” *Transp. Hum. Factors*, vol. 2, no. 3, pp. 195–216, 2000.

[70] L. Cai and Y. Zhu, “The challenges of data quality and data quality assessment in the big data era,” *Data Sci. J.*, vol. 14, pp. 1–10, 2015.

[71] S. Sontges, M. Koschi, and M. Althoff, “Worst-case Analysis of the Time-To-React Using Reachable Sets,” *IEEE Intell. Veh. Symp. Proc.*, vol. 2018-June, pp. 1891–1897, 2018.

A person smiling for the camera

Description automatically generated with low confidence**Nicolette Formosa** received a bachelor’s degree in Electrical and Electronics from the University of Malta, a master’s degree in Control Engineering from the University of Malta and a Ph.D degree in intelligent transportation systems from Loughborough University, UK in 2021. She is currently a Research Fellow at Loughborough University, UK. Her current research interests include real-time predictions, sensors, connected and autonomous vehicles, big data in transportation and artificial intelligence for active transport safety.

A person in a suit

Description automatically generated with low confidence**Mohammed Quddus** received a BSc degree in Civil Engineering from Bangladesh University of Engineering and Technology in 1998, a master’s degree in transportation engineering from National University of Singapore in 2001, and a Ph.D. degree from Imperial College London in 2006. He joined as a Lecturer with the School of Civil and Building Engineering, Loughborough University, U.K., in 2006 and was promoted to a Professor of Intelligent Transport Systems (ITS) in 2013. In 2021, Prof Quddus moved to Imperial College London as Chair Professor in ITS. He has authored over 200 technical papers in international refereed journals and conference proceedings. His research interests include connected and autonomous vehicles, AI, and statistical modelling. He is an Associate Editor of Transportation Research Part C: Emerging Technologies.

A picture containing text, newspaper, screenshot

Description automatically generated**Stephen Ison** received a degree in Economics from the University of Liverpool in 1978, an MA in Transport Economics from University of Leeds in 1979 and a PhD form Loughborough University in 2000. In 2001, he joined as a Lecturer with the School of Civil and Building Engineering, Loughborough University, U.K., and was a Professor of Transport Policy in 2007. In 2019 he moved to De Montfort University, UK. He has authored or co-authored more than 270 internationally referred journals and conference papers. The focus of his research is to provide evidence for policy initiatives at local, national, and international level. He is a co-editor of the Journal Research in Transportation Business and Management, Elsevier.

**A picture containing text, newspaper, screenshot

Description automatically generatedAndrew Timmis** received an MEng degree in Civil Engineering from the University of Sheffield, UK in 2010 and a PhD in the management and development of sustainable aviation form Sheffield University Management School in 2015. In 2014, he was appointed as a Post-Doctoral Research Fellow in Lifecycle Assessment in the University of Sheffield, UK. In 2016, he joined as a Lecturer with the School of Civil and Building Engineering, Loughborough University, U.K. His research focuses on the sustainable development of aviation and evaluation of new mobility interventions.

1. Manuscript submittedFebruary, 2021. N. Formosa and A. Timmis are with the School of Architecture, Building and Civil Engineering, Loughborough University, Loughborough, LE11 3TU, U.K. (e-mail: [n.formosa@lboro.ac.uk](mailto:n.formosa@lboro.ac.uk); [a.j.timmis@lboro.ac.uk](mailto:a.j.timmis@lboro.ac.uk)). M Quddus is with the Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, U.K. ([m.quddus@imperial.ac.uk](mailto:m.quddus@imperial.ac.uk)). S. Ison is with the Department of Politics, People and Place, De Montfort University, Leicester, LE1 9BH, U.K. (e-mail: [stephen.ison@dmu.ac.uk](mailto:stephen.ison@dmu.ac.uk)). [↑](#footnote-ref-1)
2. 1 **1:** loom angle () constant, bearing angle () constant, yaw rate () , **2:** constant, constant, = 0, **3:** , constant, , **4:** , constant, , **5:** , constant, , **6:** , constant, , **7:** constant, , , **8:** constant, , = 0, **9:** , , , **10:** , , , **11:** , , , **12:** , , , **13:** constant, , , **14:** constant, , = 0, **15:** , , , **16:** , , , **17:** , , , **18:** , , , where implies it is increasing and means that it is decreasing. [↑](#footnote-ref-2)