A stacked ensemble model for traffic conflict prediction in different road environments with multi-modal sensor data

Bowen Cai[a*], Léah Camarcat[a,b], Nicolette Formosa[b,c], Mohammed Quddus[a,d]
[a] Imperial College London, UK
[b] National Highways England
a. b.cai22@imperial.ac.uk
b. leah.camarcat18@imperial.ac.uk
c. nicolette.formosa@nationalhighways.co.uk
d. mohammed.quddus@imperial.ac.uk
* Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 27, 2024

Keywords: (up to five) Conflict prediction; Safety Surrogate Measure; variable contribution; stacked ensemble learning;

1. INTRODUCTION

Traffic collisions were responsible for 1.19 million deaths worldwide in 2021 (World Health Organization, 2023). In response to the challenge posed by the scarcity of collision data, the concept of traffic conflicts has emerged as a valuable tool for evaluating traffic safety. Theoretical frameworks, including extreme value theory and collision probability analysis, have been employed to validate the relationship between traffic conflicts and collisions. Safety surrogate measures (SSMs) offer a means to assess conflict risk, drawing upon data collected by sensors such as cameras, radars, and lidars installed on various platforms, including drones, roadside units, and vehicles. The extraction of pertinent information from raw data, including vehicle positioning, speed, and acceleration, necessitates sophisticated computer vision and data processing techniques, enabling the computation of diverse indicators essential for comprehensive safety evaluation.

Existing studies have assessed the performance of several risk metrics derived from longitudinal manoeuvres against four evaluation criteria: accuracy, timeliness, robustness and efficiency (e.g. Lu et al. 2021). They demonstrated that not a single measure performs well in predicting a traffic conflict reliably. For instance, they found that metrics performing better in prediction accuracy such as Time-To-Collision (TTC) or single-step Probabilistic Driving Risk Field (S-PDRF) performed weakly in timeliness. A significant research gap in literature is the absence of studies evaluating SSMs performance in different road environments and SSM selection justification for specific context-dependent applications (Arun et al., 2021). For instance, Tak et al. (2020) found that deceleration-based metrics perform better for rear-end conflicts than Time-To-Collision (TTC) and distance-based metrics whereas He et al. (2018) found that modified TTC (MTTC) outperformed TTC and Deceleration Rate to Avoid a Crash (DRAC). Moreover, studies that include vulnerable road users and that consider other road types than signalised intersections are lacking in the literature (Arun et al., 2021). The transferability of common model architectures for conflict prediction across different road scenarios has not been examined and is essential for industrial deployment of predictive models.

The primary objectives of this study are two-fold: (i) developing a new stacked ensemble learning model to predict a traffic conflict accurately and assess the transferability of a common model in different scenarios, and (ii) conducting an exhaustive comparative analysis of various safety surrogate measures, evaluating their effectiveness in identifying conflicts across three diverse road environments: (a) junctions in urban areas, (b) segments on high-speed roadways, and (c) vehicle-based highway data.

The stacked ensemble learning model is trained independently on the three datasets and the stochastic feature selection algorithm is performed during model training to assess the contribution of each indicator to measure the true conflict risk. Obtaining the most contributing features to the prediction enhances the model interpretability and facilitates real-time applications for proactive safety management and this constitutes a substantial contribution to the existing body of knowledge.

2. METHODOLOGY

2.1. Data acquisition and processing

The most challenging aspect in traffic conflict studies is to detect 'actual' traffic conflicts from data captured by sensors including drones, roadside camera, lidar, radar, etc. The traffic dynamics data for both junctions and motorway segments were collected using drones in China, while instrumented vehicles (equipped with radars, cameras and GPS) were employed to gather vehicle-based data on high-speed roadways in both China and the UK. The initial data processing involved utilising computer vision to detect vehicles and determine their real-world positions through spatial coordinate transformation. This facilitated the calculation of velocity and acceleration from which all SSMs were calculated.

Traffic conflicts were discerned through human judgment, whereby the presence of evasive actions, such as braking, accelerating or swerving, indicative of manoeuvres undertaken to avert collisions, served as the basis for detection (Federal Highway Administration (FHWA), 1989). Specifically, 621 conflicts were identified within the urban junction dataset, while 640 conflicts were observed in the infrastructure-based motorway dataset. Additionally, the vehicle-based motorway dataset recorded 193 conflicts over a cumulative duration of 19 hours of driving. To ensure data equilibrium, optimal sensitivity and specificity, an equivalent number of non-conflict instances were randomly selected from each dataset (Yu et al., 2020).

2.2. Selected SSM

Thirteen commonly used SSMs were selected from a total of 28 SSMs in the literature review due to lack of data availability in real scenarios such as mass of vehicles, the released energy that may affect vehicle occupants, and the change in total kinetic energy before and after the collision. The selected SSMs involve time-based, distance-based and deceleration-based measuring metrics. Time-based metrics include Time-To-Collision (TTC), modified TTC (MTTC), Post-Encroachment Time (PET), Time integrated TTC (TIT), Time Exposed TTC (TET), Headway (H) and Gap Time (GT). Selected distance-based metrics are headway, Proportional Stopping Distance (PSD) and Difference of Stopping distance and Safety distance (DSS). Deceleration-based metrics employed are Deceleration Rate to Avoid a Crash (DRAC), Potential Index for Collision with Urgent Deceleration (PIUCD) and Crash Index (CAI). Every SSM is calculated between two vehicles at an instant in time. Furthermore, the distance and the difference in speed between vehicles were added to the input data for exhaustivity.

2.3. Stacked ensemble learning model

The stacked ensemble learning approach integrates predictions from several machine learning models into a single meta-learner model to make robust prediction and improve the overall model's generalisability (Kalule et al., 2023). The robustness achieved in stacked ensemble models results from leveraging the strengths of different machine learning algorithm as each individual model may excel at capturing different aspects of the dataset (Kalule et al., 2023). The stacked ensemble learning model is composed of two layers and presented in **Equation 1** and **Figure 1**. The first layer has three models running in parallel: (i) Random Forest (RF), (ii) Support Vector Machine Radial (SVM-R), and (iii) Deep Neural Network (DNN) composed of three hidden layers. The results generated by these three models serve as inputs to the second layer which is a Gradient Boosting Machine (GBM) model whose output is the final binary prediction detecting the presence or absence of a conflict.

$$y_{pred} = GBM\ (RF(X), SVMR(X), DNN(X))\ \#(1)$$

These four models and this architecture were selected for their highest performance among 60 combinations of 25 machine learning and ensemble models tested. Each model in the first layer is independently trained on the dataset. The models' outputs create a new training set to train the GBM forming the second layer. Various hyper-parameter fine-tuning techniques were applied across different models. The Adam optimizer was utilized for the DNN, while the Grid Search method was implemented for SVM, RF, and GBM. To mitigate overfitting, an early stop mechanism was employed. Furthermore, . Train set, validation set, and test set are split according to the ratio 6:2:2. The following figure presents the architecture of the proposed stacked ensemble learning model.

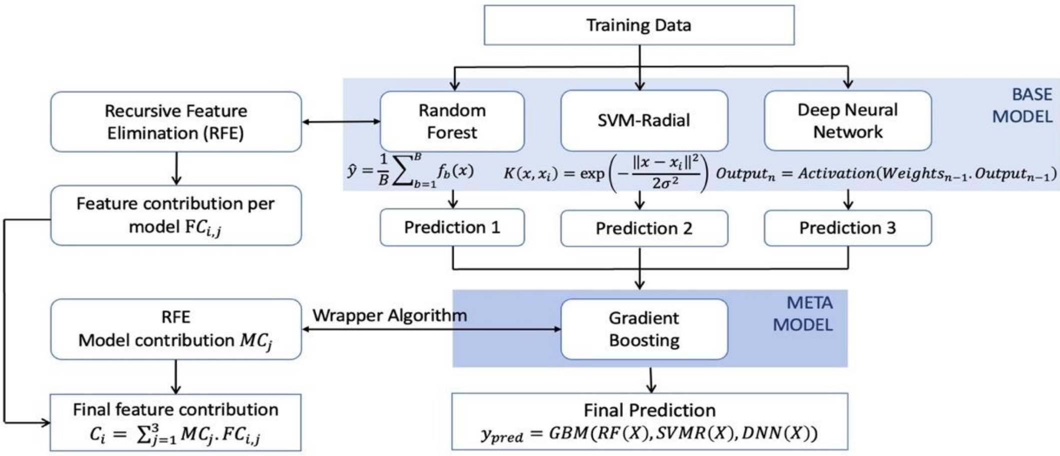

Figure 1. *Stacked ensemble learning model architecture*

2.4. Feature importance assessment methodology

Feature importance evaluation has been performed throughout the stacked model using the Recursive Feature Elimination (RFE) method. Recursive feature elimination is a process where predictors are systematically removed from a full set based on their importance scores, which are calculated through cross-validation (Guyon et al., 2002). Applying RFE on the three models in the first layer outputs feature contribution $FC_{i,j}$ for each feature *i* and model *j*. Secondly, each model contribution MC_j is computed by applying RFE on GBM model in the second layer as the input features are the prediction from each model of the first layer. The final contribution C_i of each feature *i* is computed with a weighted sum as presented in Equation 2.

$$C_i = \sum_{j=1}^{3} MC_j . FC_{i,j} \#(2)$$

3. FINDINGS AND DISCUSSION

Table 1 presents the outcomes for both the stacked ensemble learning model and the stochastic GBM feature selection algorithm. The vehicle-based motorway scenario achieved the highest overall accuracy of 99.3% with a sensitivity rate of 99% for conflict detection. The infrastructure-based motorway and urban junctions scenarios also demonstrated relatively high accuracy, reaching 87.5% and 88% respectively, with a sensitivity rate of 87.9% and 89.6% for correctly classified conflicts. Vehicle-based scenarios demonstrate high classification accuracy. However, predicting conflicts using infrastructure-based sensors poses slightly more difficulty. This is attributed to the higher number of conflicts present in the infrastructure-based data and the higher heterogeneity in driving behaviour captured by the drones compared to the 5 different drivers collecting the vehicle-based data. Sensitivity and specificity precision closely mirror the accuracy of predictions, indicating minimal disparity in false positive and false negative ratios.

Analysis of the ablation study revealed that, Gap Time, DRAC and CAI emerged as the most effective measures both for infrastructure-based urban junctions and motorway segments data. However, disparities were observed between motorway segments (infrastructure) and vehicle-based data: while time-based measures exhibited superiority for the latter, a mix of different measure types was evident for the former. Conversely, the primary influencing factors for vehicle-based conflict detection differed from those in infrastructure-based urban and motorway contexts. Specifically, deceleration rate emerged as more critical in infrastructure-based urban and motorway scenarios, whereas vehicle-based conflicts were more closely associated with time-based SSMs.

Table 1. *Results of the stacked ensemble learning model for each scenario and most contributing variables*

Scenario	Accuracy	Sensitivity	Specificity	Most significant variables ranked by importance
Urban junctions – infrastructure-based	0.88	0.896	0.865	GAP, DRAC, CAI, PSD, Speed
Motorway segments -infrastructure-based	0.875	0.879	0.871	GAP, DRAC, CAI, PSD, Speed
Motorway segments - vehicle-based	0.993	0.99	0.986	TTC, MTTC, PET, CIF, Speed Variances

This paper highlights the significance of recognising the varying efficacy of SSMs across different road environments, directly influencing their utility in real-world applications such as Advanced Driver Assistance Systems (ADAS) or roadside monitoring using smart infrastructure such as sensors developed by Vivacity Labs for promoting smarter, safe and sustainable cities. However, the study faced limitations due to the size of the dataset (1454 conflicts over more than 10 hours of video recording, less than 0.1% of total observations). Future investigations could strengthen the reliability of findings by expanding the scope to larger datasets. Furthermore, enriching the categorisation of road environments while considering diverse levels of rurality could yield deeper insights into SSM performance across various contexts. Subsequent studies can incorporate additional enhancements, including assigning varying weights to individual ensemble models.

4. CONCLUSION

Different road environments possess different geometry and traffic characteristics, while vehicle behaviour and kinematics parameters exhibit significant variations in distinct road types. Safety surrogate measures are widely used safety indicators, albeit their significance and performance in different road environments are rarely investigated by researchers. This paper not only developed a novel stacked ensemble learning model to predict traffic conflicts but also ranked SSM importance in different road scenarios during training, enabling more interpretable uses of black-box algorithms to avoid the trade-off between accuracy and interpretability. The novel stacked ensemble model achieves 98%, 96%, and 86% prediction accuracy for vehicle-based motorway, urban junctions, and infrastructure-based motorway respectively. The output of the model could be employed in devising potential interventions to prevent imminent collisions in real-time.

REFERENCES

Arun, A., Haque, M.M., Bhaskar, A., Washington, S., Sayed, T., 2021. A systematic mapping review of surrogate safety assessment using traffic conflict techniques. Accid Anal Prev 153, 106016. https://doi.org/10.1016/J.AAP.2021.106016

He, Z., Qin, X., Liu, P., Sayed, M.A., 2018. Assessing Surrogate Safety Measures using a Safety Pilot Model Deployment Dataset. Transp Res Rec 2672, 1–11. https://doi.org/10.1177/0361198118790861/ASSET/IMAGES/LARGE/10.1177_0361198118790861-FIG4.JPEG

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. J Stat Softw 28.

Lu, C., He, X., van Lint, H., Tu, H., Happee, R., Wang, M., 2021. Performance evaluation of surrogate measures of safety with naturalistic driving data. Accid Anal Prev 162, 106403. https://doi.org/10.1016/J.AAP.2021.106403

Tak, S., Kim, S., Lee, D., Yeo, H., 2020. A comparison analysis of surrogate safety measures with car-following perspectives for advanced driver assistance system. J Adv Transp 2018. https://doi.org/10.1155/2018/8040815

World Health Organization, 2023. Global status report on road safety.

Yu, R., Wang, Y., Zou, Z., Wang, L., 2020. Convolutional neural networks with refined loss functions for the real-time crash risk analysis. Transp Res Part C Emerg Technol 119, 102740. https://doi.org/10.1016/J.TRC.2020.102740