Cross or Wait? Predicting Pedestrian Interaction Outcomes at Unsignalized Crossings

Chi Zhang*, Amir Hossein Kalantari†, Yue Yang†, Zhongjun Ni‡,
Gustav Markkula†, Natasha Merat†, and Christian Berger*
* Department of Computer Science and Engineering, University of Gothenburg, Sweden.
Email: {chi.zhang, christian.berger}@gu.se
† Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK.
‡ Department of Science and Technology, Linköping University, Campus Norrköping, Sweden.

Abstract—Predicting pedestrian behavior when interacting with vehicles is one of the most critical challenges in the field of automated driving. Pedestrian crossing behavior is influenced by various interaction factors, including time to arrival, pedestrian waiting time, the presence of zebra crossing, and the properties and personality traits of both pedestrians and drivers. However, these factors have not been fully explored for use in predicting interaction outcomes. In this paper, we use machine learning to predict pedestrian crossing behavior including pedestrian crossing decision, crossing initiation time (CIT), and crossing duration (CD) when interacting with vehicles at unsignalized crossings. Distributed simulator data are utilized for predicting and analyzing the interaction factors. Compared with the logistic regression baseline model, our proposed neural network model improves the prediction accuracy and F1 score by 4.46% and 3.23%, respectively. Our model also reduces the root mean squared error (RMSE) for CIT and CD by 21.56% and 30.14% compared with the linear regression model. Additionally, we have analyzed the importance of interaction factors, and present the results of models using fewer factors. This provides information for model selection in different scenarios with limited input features.

Index Terms—Pedestrian behavior prediction, machine learning, pedestrian-vehicle interaction, simulator study, automated driving

I. INTRODUCTION

The demand for protecting pedestrians impels the vehicular automation industry to develop automated driving (AD) technologies. One of the most critical and challenging tasks in this domain is understanding and predicting pedestrian behavior during crossing, especially when they are interacting with vehicles at unsignalized crossings. Unsignalized crossings are crossings without signal displays or traffic lights, and can be marked (zebra crossings) or unmarked (non-zebra crossings). The interaction outcome prediction can provide AD systems and drivers with additional information to make safer decisions, thereby preventing vehicle-pedestrian conflicts.

Predicting pedestrian crossing behavior is challenging because pedestrian behavior is influenced by many factors [1]. Many researchers have investigated the objective interaction

factors that impact pedestrian-vehicle interaction such as time to arrival (TTA), waiting time, crossing location, age, and gender [1]–[9], and the subjective interaction factors related to pedestrians' personality traits such as sensation seeking (SS) and social value orientation (SVO) for prediction [5], [10], [11]. These studies have investigated factors that impact interaction and crossing behavior. However, they only focused on analyzing the effects of the interaction from a statistical point of view, but have not evaluated the predictability of the pedestrian crossing models with these factors.

Some studies predicted crossing intention or action using machine learning-based methods and neural networks [12]–[16], but they mainly considered pedestrians' states and environment cues, while neglecting the interaction between pedestrians and vehicles. This may be because these studies used naturalistic data making it hard to study interaction factors without being cautious about potential latent variables that might affect pedestrian behavior. Furthermore, it is usually difficult to distinguish the *original intention* from the *actual action* in naturalistic data, thereby making it intractable to investigate the outcome of the interaction. Moreover, it is impractical to get pedestrians' personality traits such as SS and SVO from naturalistic data. Hence, a simulator study with controlled conditions is a way to understand pedestrian-vehicle interactions and develop predictive models.

To the best of the authors' knowledge, there are no studies on machine learning-based predictive models for pedestrian crossing behavior that focus on pedestrian-vehicle interaction at unsignalized crossings and consider personality traits. To fill this research gap, we use the dataset collected from a distributed simulator to investigate pedestrian-vehicle interaction while crossing and predict pedestrian crossing behavior. In this paper, we aim to develop machine learning-based models to predict whether a pedestrian will cross first or wait when interacting with a vehicle at unsignalized crossings. We consider both objective factors related to crossing conditions and subjective factors related to personalities, and develop predictive models using machine learning methods. The key factors that influence pedestrian crossing decisions are analyzed and identified. The main contributions of this paper are:

1) We use the distributed simulator data collected by Kalantari et al. [5] for prediction and analysis. We build

a baseline predictive model using logistic regression and linear regression with the factors used by Kalantari et al. Compared with previous statistical models, our predictive model provides the baseline information on the predictability of pedestrian interaction outcomes.

2) We develop machine learning interaction models for prediction using the support-vector machine (SVM), random forest (RF), and neural network (NN). We investigate factors including TTA, pedestrian waiting time, being at zebra crossings or not, and both the driver's and the pedestrian's age, gender, SVO slider measure [17], and Arnett inventory of sensation seeking (AISS) [18]. Compared to the baseline model, our proposed crossing decision model improves the accuracy and F1 score by 4.46% and 3.23%. Our model also reduces the mean absolute error (MAE) and root mean squared error (RMSE) by 30.84% and 21.56% for crossing initiation time (CIT) prediction, and by 35.00% and 30.14% for crossing duration (CD) prediction.

3) We analyze interaction factors that influence pedestrian crossing behavior. We do an ablation study to investigate how a model performs when lacking the information of drivers' and pedestrians' age, gender, SVO, and AISS information. This provides guidance on model selection in scenarios when there are only partial input features.

II. RELATED WORK

A. Studies on Pedestrian-Vehicle Interaction

Studies have investigated factors that impact pedestrian-vehicle interactions [1], [19]. These interaction factors include time-related factors such as time gap or time to arrival (TTA) [2], [3], [5] and waiting time [5]–[7], the vehicle's dynamics including velocity and deceleration [20], crossing places such as whether at zebra crossings or not [5], [8], [9], and demographics such as age and gender [1]–[3], [5], [8].

In addition to these objective factors, psychological subjective factors that reflect drivers' and pedestrians' individual personality traits also influence their decision-making during the interaction. Arnett [18] defined SS as the need for varied and novel experiences, and the inclination to take risks for such experiences. The author presented a new scale namely the Arnett inventory of sensation seeking (AISS) that is related to risk behavior. SS has been found to be related to pedestrian crossing behavior [5], [10], [21]. Individuals with higher SS are more likely to take risks in traffic. SVO is used to describe how much an individual values the welfare of others compared to their own, and the individual's preference for distributing resources in social exchanges [17], [22], [23], and can be measured by SVO slider measure [17]. SVO is found to be related to a person's decision making [5], [11], [24]. Individuals with higher SVO values are more likely to be altruistic and yield or wait while interacting in traffic, while those with lower SVO values tend to behave more aggressively.

However, most of these studies are from a statistical view, while less attention has been paid to predictive models to evaluate the accuracy that can be achieved using these factors.

To contribute to the AD industry, only understanding the influencing factors of interaction outcomes is not sufficient. In this paper, we build predictive models to provide additional information to AD systems.

B. Pedestrian Crossing Behavior Prediction

Many existing learning-based models on pedestrian crossing behavior prediction are based on naturalistic data [12]–[16]. These studies mainly focused on the features of the pedestrians and the environment, but have not considered the interaction between pedestrians and vehicles explicitly. One possible reason is that these studies used naturalistic data collected under uncontrolled conditions. Since there are too many latent factors that can simultaneously affect pedestrian behavior and pedestrian-vehicle interaction in naturalistic data, it is difficult to investigate the interactions. Furthermore, without a controlled environment, it is hard to distinguish the *intention* and the *actual action* in naturalistic data, as labeling the intention of a pedestrian is a challenging task. However, without knowing the intention, it is hard to analyze the outcome of the interaction. Rasouli et al. [25] addressed and labelled intention by asking multiple annotation participants to observe the video of pedestrians and label the crossing intention, and then took the average, which is very time-consuming and costly. Besides, the naturalistic data contain noise making it harder to identify the factors that influence interactions.

Moreover, although some of the studies using naturalistic data considered pedestrian-vehicle interaction [20], [26], [27], it is hard to get the human factor parameters and personality traits such as SVO slider measure and AISS without the help of questionnaires. Völz et al. [20], [26] predicted pedestrian crossing behavior at zebra crossing considering the relative position and relative velocity between a pedestrian and a vehicle. Zhang et al. [27] proposed a predictive model at the zebra crossing, considering factors including TTA, the speed and position of both the vehicle and the pedestrian, and the age and gender of the pedestrian. Jayaraman et al. [28] used simulator data for analyzing and predicting the crossing behavior considering waiting time, but they only focused on interaction with AV approaching at different speeds. These studies have not included personality traits such as SVO and AISS due to the scarcity in such information, and they also have not investigated the influence of the presence of zebra crossing. Therefore, to overcome the challenges of using naturalistic data for crossing behavior prediction, this paper is based on a distributed simulator study (DSS). We focus on pedestrian-vehicle interaction and consider personality traits and the presence of zebra crossing.

III. METHODOLOGY

A. Distributed Simulator Data Collection

In this section, we briefly introduce the dataset. We use the same data collected by Kalantari et al. [5] in their distributed simulator study (DSS).

The DSS was conducted by connecting two high-fidelity simulators: the University of Leeds Driving Simulator (UoLDS)

Fig. 1. Illustration for the distributed simulator study (DSS). (a) A pedestrian from the driver's view: the pink bubbles are the body markers representing the pedestrian. (b) An interaction example: a pedestrian is crossing the zebra and interacting with the vehicle to their right.

Fig. 2. Top view of the zebra (left) and non-zebra crossing (right) with the designated standpoints (blue × markers). The first marker shows the pedestrian's standing point, and the second shows the curb of the virtual road. The grey rectangles are visual obstructions (bus stops).

was connected to the Highly Immersive Kinematic Experimental Research (HIKER) pedestrian lab. UoLDS is a motion-based driving simulator having eight degree-of-freedom which is housed in a 4 m-diameter spherical projection dome. The field-of-view of the projection system is 300°. HIKER is a CAVE-based simulator with dimensions equal to 9×4 m and eight Barco F90 4k projectors used to project virtual scenes at 120 Hz to its floor and walls. Fourteen body markers were attached to the pedestrian's body representing them as pink spheres to the driver as shown in Fig. 1a. The vehicle was also observable as an entity to the pedestrian as shown in Fig. 1b.

Sixty-four participant pairs (32 drivers at age: mean $(M) = 31.53$, range $(R) = 21 - 50$, standard deviation $(SD) = 1.72$; paired with 32 pedestrians at age: $M = 25.09$, $R = 19 - 34$, $SD = 0.87$) interacted with each other under different scenarios. The traffic scenarios relied on five different TTAs (3, 4, 5, 6 and 7 s) and four crossing locations (two zebra and two non-zebra crossings as shown in Fig. 2). This led to 20 conditions repeated in two separate experimental blocks, resulting in 40 randomized trials per participant pair. An 890 m two-way urban road with traffic on both lanes was created in Unity 3D to apply these settings.

Both participants were told that they were with the assumption that they were in a hurry (e.g., late for a work meeting) while they wanted to pass safely as well. They were also reminded that the priority at zebra crossings was for pedestrians. The pedestrian was asked to stand at a point on the HIKER's floor where their vision was obstructed by a bus stop on the right side making them unable to see when the subject vehicle was approaching but they could see that vehicles were going both ways. Whenever they heard an auditory tone, they needed to step up to a second point where they could see the subject vehicle, and then they could decide whether to cross or not. Drivers were told that they needed to watch the speed limit (30 mph) and they could see that sometimes pedestrians are coming from behind one of these bus stops. Likewise, they could either drive on or slow down and yield to the pedestrians.

B. Input and Output Variables

We aim to predict the "cross or wait" decisions of all pedestrians, and for those crossing cases, we are also concerned about their crossing initiation time (CIT) and crossing duration (CD). The candidate variables to be used as prediction input features are listed in Table I. Kalantari et al. [5] used the

participant pair as a feature. However, as we usually cannot get the same participant pair in a real driving scenario, i.e., the same driver meets the same pedestrian, this information is not practical to be used for prediction. So we do not use participant pair as our input feature. The interaction outcomes are listed in Table II.

TABLE I
CANDIDATE VARIABLES USED FOR PREDICTION (MODEL INPUTS).

Variable [Unit]	Description (type)
T_a [s]	Time to arrival (TTA), which was set to 3, 4, 5, 6, and 7 s. TTA represents the time for the vehicle to approach the center of the crossing (continuous)
T_w [s]	Waiting time ($M = 52.71$, $R = 13.8 - 106.98$, $SD = 19.04$), the time that the pedestrian needed to wait in each trial before prompting to cross by auditory tone. In the actual environment, it denotes the pedestrians' duration of stay towards the crossing area due to different reasons such as working with their smartphones (continuous)
L	Crossing location type, including two categories: zebra and non-zebra (categorical)
A_d, A_p	Age for both pedestrians and drivers. (discrete)
G_d, G_p	Gender for both pedestrians and drivers (categorical)
SVO_d, SVO_p [degree]	SVO slider measure for both pedestrians and drivers, calculated from the SVO slider measure questionnaire [17]. SVO_d: $M = 53.17$, $R = 45.00 - 78.38$, $SD = 8.35$, SVO_p: $M = 53.67$, $R = 43.92 - 75.26$, $SD = 7.82$ (continuous)
$AISS_d$, $AISS_p$	AISS for both pedestrians and drivers, calculated from the 20-item AISS questionnaire [18]. $AISS_d$: $M = 53.78$, $R = 43.00 - 69.00$, $SD = 6.70$, $AISS_p$: $M = 50.47$, $R = 27.00 - 61.00$, $SD = 7.17$ (continuous)

TABLE II
VARIABLES OF PEDESTRIAN INTERACTION OUTCOMES (MODEL OUTPUTS).

Variable [Unit]	Description (type)
P	Cross or wait, pedestrians' crossing decision as outcome of the interaction, 1 for crossing, 0 for waiting (binary categorical)
CIT [s]	Crossing initiation time, from the time the pedestrians were prompted to cross by the auditory tone to the time they started crossing the road (continuous)
CD [s]	Crossing duration, the time pedestrians started crossing to the time they reached the central hatch (continuous)

Three feature sets are used in prediction, as listed below.
- **Baseline:** features used in Kalantari et al.'s work [5], including T_a, T_w, L, A_p, G_p, ΔSVO, $\Delta AISS$, $ParticipantPair$, where $\Delta SVO = SVO_p - SVO_d$, $\Delta AISS = AISS_p - AISS_d$.
- **Ours:** all input features listed in Table I, including T_a, T_w, L, A_d, A_p, G_d, G_p, SVO_d, SVO_p, $AISS_d$, $AISS_p$.
- **Ours with ΔSVO, $\Delta AISS$:** in addition to the listed input features, the delta information of SVO and AISS is added. The features include T_a, T_w, L, A_d, A_p, G_d, G_p, SVO_d, SVO_p, $AISS_d$, $AISS_p$, ΔSVO, $\Delta AISS$.

C. Predictive Models

The following machine learning models are used for predicting pedestrian-vehicle interaction.

a) Linear Models: Logistic regression and linear regression were used by Kalantari et al. [5] for analysis. We use them as a baseline for prediction.

b) Support-Vector Machine (SVM): SVM aims to find a hyperplane in the feature space and can be used for classification. Here we use a linear kernel for classification, so the SVM is also considered as a linear-based model.

c) Random Forest (RF): RF is an ensemble learning method for classification and regression. RF constructs a large number of decision trees. It outputs the most selected label for the classification task and outputs the average prediction for the regression task. The number of estimators is set to 100. The maximum depth is set to five to avoid over-fitting.

d) Neural Networks (NNs): NNs are based on a collection of artificial nodes and can be used for both classification and regression. NNs usually contain several node layers, including an input layer, an output layer, and one or several hidden layers. As the number of our model's input features is small, here we use the multilayer perceptron (MLP), which is a fully connected feedforward NN. Since the input variables are in different types and with different scales, normalization is required to scale the input variables in the same range. To avoid over-fitting of MLP models, we used two hidden layers with node sizes of 16 and four, respectively.

D. Evaluation Metrics

The prediction of the crossing decision is a classification problem, so we use prediction accuracy (ACC) and F1 score for evaluation. The evaluation functions are shown below, where P and N denote the numbers of positives and negatives, respectively. TP, TN, FP, FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively.

$$ACC = \frac{TP + TN}{P + N} \tag{1}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{2}$$

The predictions of CIT and CD are regression problems, so we use mean absolute error (MAE) and root mean squared error (RMSE) for evaluation. The evaluation functions are as follows, where y_i denotes the groundtruth for the ith trail, and $\hat{y}_i$ denotes the corresponding prediction, n denotes the number of trails.

$$MAE = \frac{\Sigma|\hat{y}_i - y_i|}{n} \tag{3}$$

$$RMSE = \sqrt{\frac{\Sigma(\hat{y}_i - y_i)^2}{n}} \tag{4}$$

E. Model Implementation Details

There were a total of 1279 trials collected, where 836 trials were crossing cases. To better evaluate models, we use five-fold cross-validation. We randomly divide the dataset into five sets, and each time we use one set for testing and the rest for training. There is no overlap between the training and test sets. We report the average performance over test sets.

The information of categorical features (e.g., crossing location type and gender) is represented by categories rather than in numeric formats. We apply the commonly used one-hot encoding to encode non-ordinal categorical variables.

IV. RESULTS AND DISCUSSIONS

A. Crossing Decision

The prediction accuracy and F1 score of the pedestrian crossing decision are shown in Table III. Using our proposed MLP model, the accuracy and F1 score are improved by 4.46% and 3.23% compared with logistic regression based on the effective input features stated by Kalantari et al. [5].

TABLE III
PREDICTION ACCURACY AND F1 SCORE OF CROSSING DECISION

Model	Zebra		Non-zebra		Total	
(features)	ACC	F1	ACC	F1	ACC	F1
LR(baseline) [5]	91.39%	95.32%	80.16%	75.81%	85.77%	89.24%
LR (ours)	91.39%	95.30%	80.47%	76.01%	85.93%	89.32%
SVM (ours)	91.24%	95.22%	81.09%	77.04%	86.16%	89.55%
RF (ours)	91.24%	95.33%	88.44%	85.93%	89.84%	92.44%
MLP (ours)	91.55%	95.28%	**88.91%**	**86.63%**	**90.23%**	**92.47%**

We investigate the prediction results of zebra and non-zebra crossing (Table III). The prediction results on zebra crossing cases are better than non-zebra crossing cases, showing that zebra crossing cases are more predictable. All models perform similarly in zebra crossing cases. For non-zebra crossing cases, the non-linear models (RF and MLP) perform significantly better than the linear models (LR and SVM). Compared to the LR model with baseline features, our proposed MLP model improved the accuracy and F1 score by 8.75% and 10.82% respectively. This implies that the non-zebra crossing cases are more complex because of non-linearity and can be better handled by our proposed non-linear models.

The prediction accuracy and F1 score versus TTA for different models are shown in Fig. 3. Smaller TTAs lead to lower prediction accuracy and F1 score, which indicates higher prediction difficulty. One possible reason is a smaller TTA means shorter reaction time and more interaction between pedestrians and drivers. This increases the uncertainty of crossing behavior and makes it harder to predict. Non-linear models (RF and MLP) perform better than linear models (LR and SVM). This implies that the interaction process involves non-linearity that can be better handled by non-linear models.

The seven most important input features in each model are listed in Table IV. As MLP is fully connected with all input features with hidden nodes, we do not list its feature importance here. In all three compared models, crossing location type L (whether crossing at a zebra cross or not) and time to arrival T_a are the most important features. The AISS of pedestrians and the pedestrian waiting time are also important for all three models. In linear models (LR and SVM), the models rely more on objective properties such as age and gender, while in non-linear models (RF), personality traits such as AISS and SVO are more important for prediction. As the non-linear models that rely on AISS and SVO for prediction gain better prediction
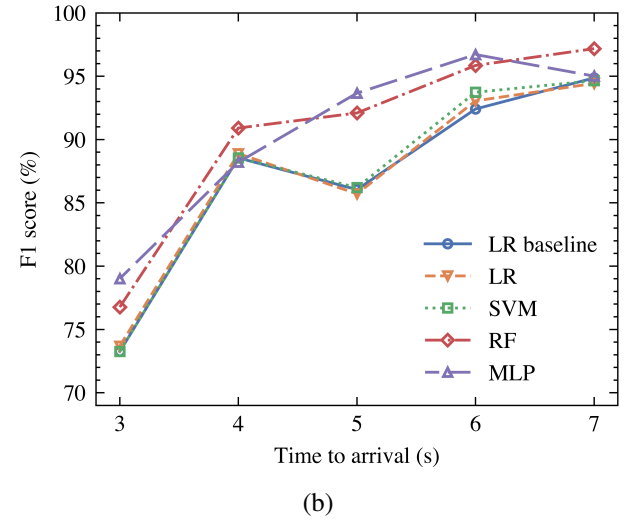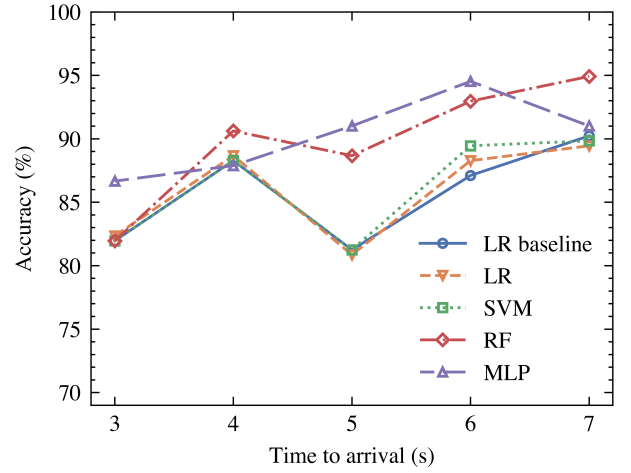

(a)

(b)

Fig. 3. The (a) prediction accuracy and (b) F1 score versus time to arrival for logistic regression (LR), support-vector machine (SVM), random forest (RF), and multilayer perceptron (MLP) models.

results, this indicates that the personality traits contribute to the crossing decision prediction in a non-linear way. In Sec. IV-C, we present more results on prediction without these features to show the features' impacts.

TABLE IV
MOST IMPORTANT FEATURES IN THE CROSSING DECISION MODEL

Model	Most Important Features (sorted by importance)
LR	$L, T_a, G_p, AISS_p, A_d, A_p, T_w$
SVM	$L, T_a, G_p, AISS_p, A_d, T_w, SVO_p$
RF	$T_a, L, T_w, AISS_d, SVO_d, AISS_p, A_p$

B. Crossing Initiation Time and Crossing Duration

For crossing cases, we are also interested in the crossing initiation time (CIT) and crossing duration (CD) as interaction outcomes. The unit of MAE and RMSE is seconds. The smaller error represents better performance.
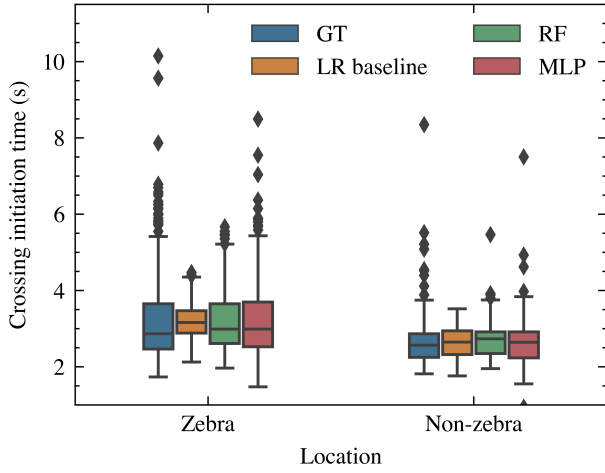
Fig. 4. Box plots of crossing initiation time for groundtruth (GT), linear regression (LR), RF, and MLP models. Outliers are marked as diamond dots.
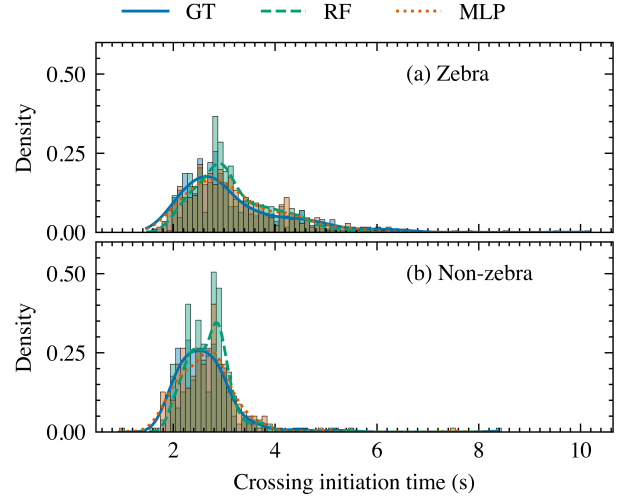

Fig. 5. Distribution of predicted crossing initiation time. The y-axis is the normalized density such that the total area of the histogram in each subplot is equal to 1.

a) Crossing initiation time: The prediction error of CIT is shown in Table V. Features with ΔSVO and $\Delta AISS$ are used as inputs for RF and MLP models. Both RF and MLP show a great improvement compared with the baseline LR model, while RF achieves the best performance, and reduces the MAE and RMSE error by 30.84% and 21.56%, respectively.

TABLE V
PREDICTION ERROR OF CROSSING INITIATION TIME

Model (features)	MAE [s]	RMSE [s]
LR (baseline) [5]	0.618	0.897
RF (ours with ΔSVO, $\Delta AISS$)	**0.428**	**0.704**
MLP (ours with ΔSVO, $\Delta AISS$)	0.500	0.794

The box plots of CIT for the groundtruth and different models at different crossing locations are shown in Fig. 4. Compared with zebra crossing cases, non-zebra crossing cases have shorter CIT with a narrower distribution. This implies that when pedestrians cross at the non-zebra crossing, there are fewer chances for them to hesitate and they have to make a quicker decision. The prediction of the LR baseline model is more concentrated, while the MLP model is more distributed.

We further compare the distribution of predicted CIT of RF and MLP models with groundtruth. The distribution density is shown in Fig. 5. Although the error of RF is the smallest, the distribution of MLP is closer to the groundtruth, which indicates the neural network model may have better generalizability.

b) Crossing duration: The prediction error of CD is shown in Table VI. Input features are the same as in crossing decision prediction models. Compared with the baseline LR results, RF and MLP gain great improvements. The MLP model achieves the best results, where the MAE and RMSE are 0.282 s and 0.446 s, respectively, showing an improvement of 35.00% and 30.14% compared with the LR model.

The box plots of CD for the groundtruth and different models at different crossing locations are shown in Fig. 6. Pedestrians who crossed the road at zebra crossings took longer time

TABLE VI
PREDICTION ERROR OF CROSSING DURATION

Model	MAE [s]	RMSE [s]
LR (baseline) [5]	0.434	0.638
RF (ours)	0.297	0.458
MLP (ours)	**0.282**	**0.446**

than those who crossed at non-zebra crossings. This could be because people feel safer to cross at a zebra crossing so they can take their time. The prediction of the LR baseline model is more concentrated, while the MLP model is more distributed.

The CD distribution density of the groundtruth and RF and MLP predictions are shown in Fig. 7. The errors of RF and MLP prediction results are close, while the distribution of MLP
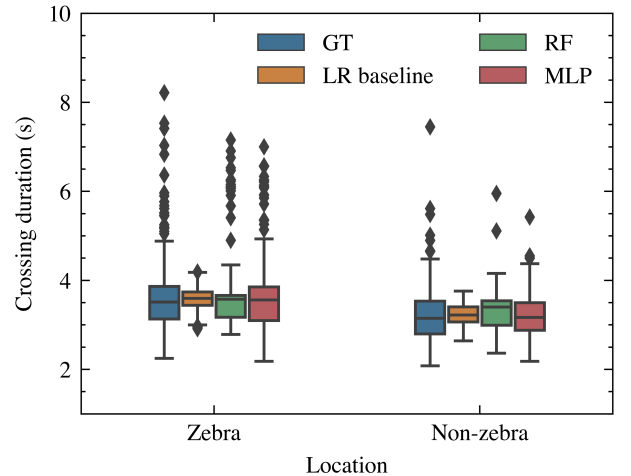

Fig. 6. Box plots of crossing duration for GT, linear regression (LR), RF, and MLP models. Outliers are marked as diamond dots.
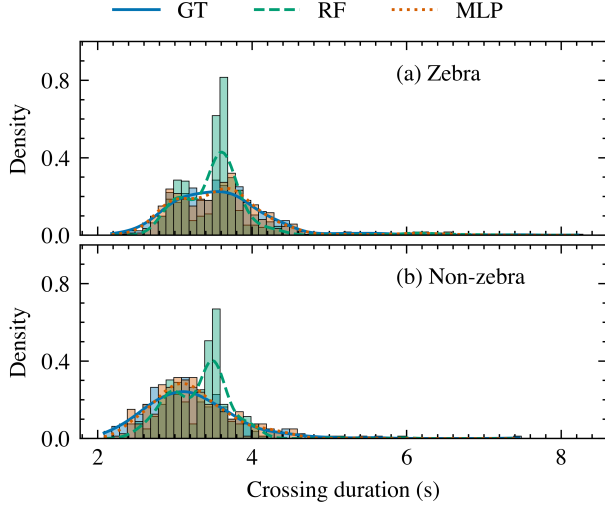
Fig. 7. Distribution of predicted crossing duration. The y-axis is the normalized density such that the total area of the histogram in each subplot is equal to 1.

is shown to be closer to the groundtruth. This indicates that the MLP model is more capable of predicting crossing duration.

C. Ablation Study

In practice, it is usually difficult to obtain all the features we use in our models. Therefore, it is necessary to investigate the performance of the predictive models with only parts of the features. We consider four subsets of features as follows.

Feature subset 1: $T_a, T_w, L, A_d, A_p, G_d, G_p$. Compared with the original feature set, subset 1 removes the features related to the personalities of the drivers and pedestrians, i.e. SVO and AISS information. This feature set fits scenarios without personality traits.

Feature subset 2: T_a, T_w, L, A_d, G_d. Subset 2 removes the age and gender information of pedestrians while keeping this information for drivers. This feature set fits scenarios where the driver's information is known to aid decision-making on a private vehicle for example.

Feature subset 3: T_a, T_w, L, A_p, G_p. Subset 3 removes the age and gender information of drivers while keeping this information for pedestrians. This feature set fits scenarios where the pedestrian's age and gender information is known, for example, for the predictions on a labeled publicly available dataset.

Feature subset 4: T_a, T_w, L. Subset 4 only keeps TTA, waiting time, and crossing location type. This is the basic feature set that the information is easy to get, and the features are important for all predictive models.

a) Crossing decision: The crossing decision prediction results using different subsets of features are shown in Table VII. Surprisingly, with fewer input features, the LR model does not get a significant drop in performance, while the RF and MLP models get much worse results compared with using all features. This shows that the LR model depends mainly on

several important basic features for prediction, while the other two models rely on all features.

For subset 1, the MLP model gets the best results with accuracy and F1 score of 88.43% and 91.30%, respectively. The neural networks are able to handle the case when there are many features with non-linear relations. Without SVO and AISS information, the performance of MLP drops, so they are important for the MLP model. For subsets 2 and 3, the RF model gets the best performances, while MLP gets the worst results. This indicates that age and gender features play important roles in the MLP model. As discussed in Sec. IV-A, the age and gender features are not in the top six important features of the RF model. Therefore, even without age and gender information, it can still provide a relatively good prediction compared with the other two models. For subset 4, the RF model gets the best performance, while the LR model achieves competitive results. This indicates that for a simple model with very few input features, the linear model is still strong and powerful compared with the other two non-linear models.

TABLE VII
CROSSING DECISION PREDICTION USING A SUBSET OF THE FEATURES

Model	LR		RF		MLP	
Features	ACC	F1	ACC	F1	ACC	F1
All	85.93%	89.32%	89.84%	92.44%	**90.23%**	**92.47%**
SubSet 1	85.77%	89.24%	87.89%	90.90%	**88.43%**	**91.30%**
SubSet 2	85.38%	88.97%	**86.09%**	**89.54%**	83.35%	87.38%
SubSet 3	85.93%	89.34%	**86.56%**	**89.92%**	83.19%	87.26%
SubSet 4	85.54%	89.15%	**85.77%**	**89.36%**	85.69%	89.35%

b) Crossing initiation time: The prediction results for CIT using subsets of features are shown in Table VIII. With fewer input features, the errors of LR models are increased slightly, and the errors of RF and MLP models are increased to a great extent. This indicates that the LR model relies more on TTA, waiting time, and crossing location type, while the other two non-linear predictive models also depend on personality trait features, age, and gender.

TABLE VIII
CROSSING INITIATION TIME PREDICTION USING A SUBSET OF THE FEATURES

Model	LR		RF		MLP	
Features	MAE	RMSE	MAE	RMSE	MAE	RMSE
All	0.616	0.900	**0.428**	**0.704**	0.500	0.794
Subset 1	0.646	0.945	**0.485**	**0.757**	0.524	0.802
Subset 2	0.665	0.959	**0.542**	**0.817**	0.679	0.999
Subset 3	0.644	0.948	**0.558**	**0.823**	0.618	0.901
Subset 4	0.678	0.970	**0.659**	**0.949**	0.694	0.980

c) Crossing duration: The prediction results for CD using subsets of features are shown in Table IX. With fewer input features, the errors of LR models slightly increase. The other two non-linear models show a great increase in errors with fewer features. MLP performs best with all features and performs worst with subset 4. This indicates that MLP is more suitable for prediction with a large number of input features while less capable for prediction with fewer features.

TABLE IX
CROSSING DURATION PREDICTION USING A SUBSET OF THE FEATURES

Model	LR		RF		MLP	
Features	MAE	RMSE	MAE	RMSE	MAE	RMSE
All	0.428	0.615	0.297	0.458	**0.282**	**0.446**
SubSet 1	0.474	0.668	0.345	0.496	**0.321**	**0.478**
SubSet 2	0.478	0.677	**0.418**	**0.563**	0.454	0.618
SubSet 3	0.472	0.668	**0.418**	**0.613**	0.507	0.708
SubSet 4	**0.476**	**0.677**	0.491	0.685	0.502	0.702

V. CONCLUSION

In this paper, we have proposed predictive models for pedestrian crossing behavior when interacting with vehicles at unsignalized crossings. Our proposed MLP predictive model has improved the prediction accuracy and F1 score of crossing decisions by 4.46% and 3.23% compared to that of the logistic regression baseline. For crossing initiation time prediction, the proposed RF model decreases the RMSE error by 21.56%, and for crossing duration, the MLP model reduces the RMSE error by 30.14%. Furthermore, we have analyzed the importance and influence of the various input features in different models. The presence of zebra crossing, TTA, AISS of pedestrians, and pedestrian waiting time are important features for all models for crossing decision prediction. Additionally, we have applied an ablation study and presented the predictability of crossing behavior with fewer input features. This provides guidance for selecting proper models when there are only limited input features. In future work, more information such as pedestrians' trajectories and postures can be included to support building predictive models.

REFERENCES

[1] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 900–918, 2019.

[2] A. Gorrini, L. Crociani, G. Vizzari, and S. Bandini, "Observation results on pedestrian-vehicle interactions at non-signalized intersections towards simulation," *Transportation research part F: traffic psychology and behaviour*, vol. 59, pp. 269–285, 2018.

[3] J. P. N. Velasco, Y. M. Lee, J. Uttley, A. Solernou, H. Farah, B. van Arem, M. Hagenzieker, and N. Merat, "Will pedestrians cross the road before an automated vehicle? the effect of drivers' attentiveness and presence on pedestrians' road crossing behavior," *Transportation research interdisciplinary perspectives*, vol. 12, p. 100466, 2021.

[4] J. E. Domeyer, J. D. Lee, H. Toyoda, B. Mehler, and B. Reimer, "Interdependence in vehicle-pedestrian encounters and its implications for vehicle automation," *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[5] A. H. Kalantari, Y. Yang, J. G. de Pedro, Y. M. Lee, A. Horrobin, A. Solernou, C. Holmes, N. Merat, and G. Markkula, "Who goes first? a distributed simulator study of vehicle–pedestrian interaction," *Accident Analysis & Prevention*, vol. 186, p. 107050, 2023.

[6] Y. Wang, Q. Su, C. Wang, and C. G. Prato, "Investigating yielding behavior of heterogeneous vehicles at a semi-controlled crosswalk," *Accident Analysis & Prevention*, vol. 161, p. 106381, 2021.

[7] J. E. Domeyer, J. D. Lee, H. Toyoda, B. Mehler, and B. Reimer, "Driver-pedestrian perceptual models demonstrate coupling: Implications for vehicle automation," *IEEE Transactions on Human-Machine Systems*, 2022.

[8] M.-S. Cloutier, U. Lachapelle, A.-A. d'Amours Ouellet, J. Bergeron, S. Lord, and J. Torres, ""outta my way!" individual and environmental correlates of interactions between pedestrians and vehicles during street crossings," *Accident Analysis & Prevention*, vol. 104, pp. 36–45, 2017.

[9] A. Habibovic, V. M. Lundgren, J. Andersson, M. Klingegård, T. Lagström, A. Sirkka, J. Fagerlönn, C. Edgren, R. Fredriksson, S. Krupenia *et al.*, "Communicating intent of automated vehicles to pedestrians," *Frontiers in psychology*, p. 1336, 2018.

[10] T. Rosenbloom, "Sensation seeking and pedestrian crossing compliance," *Social Behavior and Personality: an international journal*, vol. 34, no. 2, pp. 113–122, 2006.

[11] L. Crosato, C. Wei, E. S. Ho, and H. P. Shum, "Human-centric autonomous driving in an av-pedestrian interactive environment using svo," in *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*. IEEE, 2021, pp. 1–6.

[12] Z. Fang and A. M. López, "Is the pedestrian going to cross? answering by 2d pose estimation," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, Conference Proceedings, pp. 1271–1276.

[13] P. Gujjar and R. Vaughan, "Classifying pedestrian actions in advance using predicted video of urban driving scenes," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, Conference Proceedings, pp. 2097–2103.

[14] M. Chaabane, A. Trabelsi, N. Blanchard, and R. Beveridge, "Looking ahead: Anticipating pedestrians crossing with future frames prediction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, Conference Proceedings, pp. 2297–2306.

[15] B. Yang, W. Zhan, P. Wang, C. Chan, Y. Cai, and N. Wang, "Crossing or not? context-based recognition of pedestrian crossing intention in the urban environment," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[16] S. Zhang, M. Abdel-Aty, Y. Wu, and O. Zheng, "Pedestrian crossing intention prediction at red-light using pose estimation," *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[17] R. O. Murphy, K. A. Ackermann, and M. Handgraaf, "Measuring social value orientation," *Judgment and Decision making*, vol. 6, no. 8, pp. 771–781, 2011.

[18] J. Arnett, "Sensation seeking: A new conceptualization and a new scale," *Personality and individual differences*, vol. 16, no. 2, pp. 289–296, 1994.

[19] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, "Do they want to cross? understanding pedestrian intention for behavior prediction," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 1688–1693.

[20] B. Völz, K. Behrendt, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "A data-driven approach for pedestrian intention estimation," in *2016 IEEE 19th international conference on intelligent transportation systems (itsc)*. IEEE, 2016, Conference Proceedings, pp. 2607–2612.

[21] H. Wang, A. Wang, F. Su, and D. C. Schwebel, "The effect of age and sensation seeking on pedestrian crossing safety in a virtual reality street," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 88, pp. 99–110, 2022.

[22] C. G. McClintock and S. T. Allison, "Social value orientation and helping behavior," *Journal of Applied Social Psychology*, vol. 19, no. 4, pp. 353–362, 1989.

[23] R. O. Murphy and K. A. Ackermann, "Social value orientation: Theoretical and measurement issues in the study of social preferences," *Personality and Social Psychology Review*, vol. 18, no. 1, pp. 13–41, 2014.

[24] L. Crosato, H. P. Shum, E. S. Ho, and C. Wei, "Interaction-aware decision-making for automated vehicles using social value orientation," *IEEE Transactions on Intelligent Vehicles*, 2022.

[25] A. Rasouli, I. Kotseruba, T. Kunic, and J. K. Tsotsos, "Pie: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, Conference Proceedings, pp. 6262–6271.

[26] B. Völz, H. Mielenz, G. Agamennoni, and R. Siegwart, "Feature relevance estimation for learning pedestrian behavior at crosswalks," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. IEEE, 2015, pp. 854–860.

[27] H. Zhang, Y. Liu, C. Wang, R. Fu, Q. Sun, and Z. Li, "Research on a pedestrian crossing intention recognition model based on natural observation data," *Sensors*, vol. 20, no. 6, p. 1776, 2020.

[28] S. K. Jayaraman, D. M. Tilbury, X. J. Yang, A. K. Pradhan, and L. P. Robert, "Analysis and prediction of pedestrian crosswalk behavior during automated vehicle interactions," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 6426–6432.