Influence of road environmental elements on pedestrian and cyclist road crossing behaviour

Yuichi Saito ○, Fuma Kochi○, Makoto Itoh○, Takesato Fushima○, Takashi Sugano○ and Yasunori Yamamoto○

○Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan; ○Master’s Program in Risk and Resilience Engineering, University of Tsukuba, Tsukuba, Japan; ○Mazda Motor Corporation, Hiroshima, Japan

ABSTRACT
The pedestrian and cyclist-related accident fatality rate is higher than that of other traffic accidents. One of the pedestrian behaviours that leads to traffic accidents is the act of moving rapidly onto the road from a blind spot without warning. Expert drivers practice hazard-anticipatory driving and will naturally seek to reduce uncertainty by attempting to fit their current driving context into a pre-existing category. Risk management is the process of identifying hazards and assessing and controlling risks to attain safety. The purpose of this study was to evaluate the influence that driving context-altering road environmental elements exert on the road-crossing behaviour of pedestrians and cyclists. Thus, this study attempted to identify covert hazards (obscured pedestrians and cyclists). A logistic regression analysis was employed along with data from the near-miss incident database, in which approximately 140,000 near-crash-relevant events were registered in 2017. By using the logistic regression analysis along with the annotations recorded in the database, we constructed a predictive model to identify covert hazards. The study demonstrated the feasibility of using a set of environmental elements that shape the driving context to construct a predictive model that identifies covert hazards.

1. Introduction
Erratic pedestrian and cyclist behaviour associated with moving into the road from blind spots is one of the factors that contribute to the occurrence of accidents and near-misses. Accidents involving either pedestrians (36.5%) or cyclists (13.4%) [1] are some of the most common types of accidents that lead to traffic fatalities. Pedestrians and cyclists experience an increased accident risk when crossing roads, owing to the shared road space [2]. Among the fatal pedestrian- and cyclist-related accidents, 66.5% occurred when they crossed the road. Prevention of these accidents is vital for decreasing future traffic accident fatalities.

Urban areas that are crowded with residential and commercial buildings and stores contain many blind spots in which the driver’s field of view is poor. One of the pedestrian and cyclist behaviours that leads to traffic accidents is the act of moving rapidly onto the road from a blind spot without warning. While the driver’s environment leads to (i) cognitive errors, (ii) decision errors and aggressive behaviours, and (iii) performance errors [3, 4] on the part of the driver, pedestrians can also cause accidents. For example, distracted pedestrians [5] or risk-takers may neglect to look for an approaching car while crossing a road. This may be due to circumstances such as talking with friends and walking the dog, which can cause distraction. Additionally, using a cell phone or camera or reading a magazine can cause inattentive blindness.

Car manufacturers have developed advanced driver assistance systems (ADAS) to prevent traffic accidents pertaining to pedestrian and cyclist behaviour. The function of an ADAS includes enhancing driver perception, drawing attention to potential risks, setting off warning signals, and performing safety control [6]. For example, night vision/pedestrian detection, forward collision warning, and forward collision mitigation/avoidance systems have already been put into practical use. Night vision/pedestrian detection systems display an object photographed by an infrared camera on the display. When a forward collision warning system detects a pedestrian in front of the vehicle, it provides a warning (alert) to the driver to allow them to take evasive action. If evasive actions are not taken, the system activates autonomous emergency braking to avoid a crash. However, these ADASs support the driver only after they have detected a pedestrian or vehicle. Thus, these systems fail to address potential risks. When a pedestrian or a cyclist initiates a road crossing from a blind spot, the conventional forward collision mitigation/avoidance systems may not be able to avoid a crash due to the limited and short time margin. In such situations, expert drivers will naturally seek to reduce uncertainty by attempting to fit their current...
driving context into a pre-existing category [7]. Experienced drivers perform preventive driving, reducing their speed in advance based on knowledge and experience. Our research objective is the quantification of the hazard-anticipatory knowledge acquired by drivers.

Related works that have investigated the factors that influence vehicle-pedestrian crashes fall into two categories:

1. Microscopic spatial analysis [8,9] is carried out at specific locations in order to identify the relevant factors and solve safety issues.

2. Macroscopic spatial analysis [10–15] is focused on zonal-level traffic accidents at various levels of the entirety of the specified area and captures spatial trends and safety issues.

Chen and Zhou [15] explored the relationship between vehicle and pedestrian crash frequency vs. risk and various built environment factors such as road network and land use, intersection type, stop sign density, zonal speed limit, and school density. However, most of the studies that have investigated macroscopic spatial analysis cannot be accurately tracked in terms of driver behaviour when an accident occurs.

Today, powerful sensor technology is available (e.g. cameras, light detection and ranging (LIDAR), and radar [16]), and drive recorders [17] can also be useful for quantifying hazard-anticipatory knowledge. Since 2004, the Smart Mobility Research Center (SMRC) at the Tokyo University of Agriculture and Technology has managed the near-miss incident database. The recorded number of near-miss events and traffic accidents exceeded 140,000 in 2017. The near-miss incident data were recorded by drive recorders on taxis. When the longitudinal deceleration exceeded $-0.45G$, the drive data were automatically recorded 10s before the trigger and 5s after the trigger. This database includes vehicle and driver information. In addition, qualitative data (annotations) describing environmental elements that altered the driving context, such as area type, road type, intersection type, and traffic density, were added by rating experts. Previous studies [18–20] have proposed the potential risk estimation method, which is based on modelling the knowledge and experience contained in the near-miss database. These investigations focused on analysing the near-miss event formation process and proposed a machine learning approach to predict the time margin allowed for drivers to perform evasive action to avoid a crash, which represents the severity level of a near-miss.

The purpose of this study is to evaluate the influence that driving context-altering environmental elements exert on the road-crossing behaviour of pedestrians and cyclists. Whereas previous studies have attempted to estimate the potential risk (i.e. severity), the present study attempts to identify covert hazards (concealed pedestrians and cyclists). A near-miss incident database is used to achieve this purpose. In this study, we propose the following hypothesis:

- The environmental elements that alter the driving context can determine whether a pedestrian or a cyclist will initiate a road-crossing from a blind spot.

Predicting or knowing the hazard information that could determine whether a pedestrian or a cyclist will move rapidly onto the road from a blind spot without warning can be useful in terms of assisting drivers to avoid near-miss incidents. Thus, the contribution of this study is in obtaining the hazard-anticipatory information related to whether a pedestrian or a cyclist will move rapidly onto the road from a blind spot without warning. The remainder of this paper is organized as follows: In Section 2, the methodology is described, which includes data extraction, pre-processed annotation, and data analysis methods. In Section 3, the results obtained from the analysis are presented. In Section 4, a discussion is provided regarding the abovementioned hypothesis.

2. Methodology

2.1. Data extraction

We searched and extracted the specified data by using the management tool provided in the near-miss incident database. In this study, the near-miss incident data were extracted in the following manner:

- Near-miss incidents that were extracted were separated into severity levels that included low, medium, and high. These incident severity levels were defined by the time margin allowed for drivers to perform evasive action to avoid a crash, and the rating procedure was based on a subjective method that depended on the experience of rating experts.
- Data related to pedestrian and cyclist behaviour were selected.
- Data recorded between 1 January 2011 and 1 January 2015 were selected.

The number of data points extracted using this procedure is shown in Table 1. Our target scenarios involved a pedestrian or cyclist initiating a road-crossing from a blind spot at an unsignalized intersection. In the database, the information related to blind spots was not annotated, except for recorded video format information. Therefore, the following information was manually extracted while referring to recorded video footage:

- Whether a pedestrian or a cyclist will move rapidly onto the road from a blind spot. Data pertaining...
to pedestrians was marked with the symbol “0,” and data pertaining to cyclists was marked with the symbol “1.”

- Elements that create a blind spot. These were annotated and categorized into four different groups: Building walls, parked vehicles, motionless vehicles, and moving vehicles.
- Whether or not signalized intersections were involved. This study focused on investigating scenarios that did not involve a signalized intersection. Near miss data at signalized intersections were thus excluded.
- Whether or not the road on which the vehicle was travelling was a priority road. This study focused on investigating the erratic behaviour of pedestrians and cyclists. The events in which pedestrian-and cyclist-travelled roads were at priority were excluded.

The extracted data as well as the classification of the blind spot elements are presented in Tables 2 and 3. The percentage (67.4%) of events in which cyclists initiated a road-crossing from a blind spot created by a building wall was relatively high compared to pedestrians (35.6%).

### 2.2. Environmental elements that alter the driving context

The annotations that expressed environmental elements had been added to the recorded data in the database by SMRC ratings experts. However, annotations were not added to all data. For example, annotations were not added to low severity incidents involving near-miss events. In addition, the criteria for classification of annotations were not publicly available. Therefore, in this study, the criteria for the classification of the annotations were redefined as shown in Table 4. Based on these criteria, annotations were attached to all of the extracted data (total 1423 events). A total of 13 parameters were used. These contextual properties could be classified into three groups: (a) static context properties, (b) dynamic context properties, and (c) other context properties, as shown in Table 4. The 12 parameters excluding $Y_{\text{gap}}$ were qualitative properties. The lateral distance to the object causing the blind-spot, $Y_{\text{gap}}$, was one factor that could contribute to a reduction in the driver’s field of view. The value of the lateral distance for each event was estimated using a distance measurement tool provided in the database. Although it is not easy for a vehicle system equipped with sensor technology to estimate the lateral distance to an object that causes a blind spot in real-time, the lateral distance, $Y_{\text{gap}}$, was treated as though it could be estimated in this study. The sidewalk type, as shown in Table 4, was defined based on the degree of separation between the car driving corridor and the footpath. SIDE1 was defined as the condition in which the car driving corridor and the footpath were not completely separated. SIDE2 was defined as the condition in which the car driving corridor and the footpath were separated by a white line. SIDE3 was defined as the condition in which the car driving corridor and the footpath were clearly separated by curbs, and SIDE4 was defined as the condition in which the car driving corridor and the footpath were clearly separated by guardrails. The definitions for the density of parked vehicles, pedestrians, traffic, and leading vehicle are also shown in Table 4. With the exception of $Y_{\text{gap}}$, the data used in this study were transformed into dummy variables (symbol “1” or “0”). Thus, a total of 38 categories were used.

### 2.3. Model description

The purpose of this analysis was to investigate the associations between pedestrian or cyclist behaviour and possible contributory variables (environmental elements) by using the data recorded in the near-miss incident database. This study used a logistic regression analysis to achieve this purpose. The goal of the analysis using logistic regression was to find the best fit parameters through the use of the model-building technique. In the logistic regression model, a latent variable can be described using the following expression:

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p,$$

where $x_j$ is the value of the $j$th explanatory variable, $\beta_j$ is the corresponding coefficient for $j = 1, 2, 3, \ldots, p$, and $p$ is the number of explanatory variables. With the latent
variable, the specific form of the logistic regression model was formulated as follows:

\[ \pi(x) = \frac{\exp(g(x))}{1 + \exp(g(x))}, \quad (2) \]

where \( \pi(x) \) is the conditional probability of a positive outcome. The logistic regression model can also be expressed as a logit transformation:

\[ g(x) = \ln \left[ \frac{\pi(x)}{1 + \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p. \quad (3) \]

The maximum likelihood method was then employed to measure the associations. The logistic regression model determines the coefficients that produce the observed outcome (a scenario in which either a pedestrian or a cyclist moves rapidly onto the road from a blind spot). Thus, the best estimate \( \beta \) can be obtained. In the logistic regression model, the influence that the contributory variables exert on the outcome can be revealed by the odds ratio:

\[ OR = \exp(\beta_i), \quad (4) \]

with a 95% confidence interval (CL). If the odds ratio is 1, it denotes that the environmental element does not have an impact on the outcome. Therefore, if the upper limit of the odds ratio (Upper CL) is less than 1, it denotes that the environmental element has an impact on the outcome. This implies that the environmental element leads to a scenario in which a pedestrian moves rapidly onto the road from a blind spot. However, if the lower limit of the odds ratio (Lower CL) is greater than 1, it also denotes that the environmental element has an impact on the outcome. This implies that the environmental element leads to a scenario in which a cyclist moves rapidly onto the road from a blind spot.

In this study, using the stepwise method, the explanatory variables could be chosen based on the value of the index used for evaluating the goodness of fit in the Akaike Information Criterion (AIC). Its effectiveness in describing the associations between pedestrian or cyclist behaviour and possible contributory variables (environmental elements) could be assessed and revealed with numerical overall measures using the Hosmer-Lemeshow statistic.

\subsection{2.4. Data description}

The dataset used in this analysis was derived from a sample in which a pedestrian or cyclist initiated a road-crossing from a blind spot created by a building wall. To facilitate the interpretation of the results of the analysis, the number of data points was made uniform in both cases. Among the 267 scenarios in which pedestrians moved rapidly onto the road from blind spots created by building walls (Table 2), 250 scenarios were randomly selected. Among the 459 scenarios in which cyclists moved rapidly onto the road from blind spots created by building walls (Table 3), 250 scenarios were randomly selected. The explanatory variables could be continuous or discrete, as can be seen in Table 4.

\begin{center}
\begin{table}[h]
\centering
\caption{Definition of environmental elements.}
\begin{tabular}{|l|l|l|l|}
\hline
Property & Environmental elements & Type & Description \\
\hline
Static & \( y_{\text{PPI}} \) & Continuous & Lateral distance to the blind element [m] \\
& AREA & Discrete & Four categories: \textsc{residential}, \textsc{urban}, \textsc{rural}, \textsc{other} \\
& ROAD & Discrete & Three categories: \textsc{one way}, \textsc{both way}, \textsc{other} \\
& SIDEWALK & Discrete & Four categories: \textsc{side1}, \textsc{side2}, \textsc{side3}, \textsc{side4} \\
& INTERSECTION & Discrete & Three categories: 2- or 3-FORKED ROAD, 4- or 5-FORKED ROAD, \textsc{straight} \\
& ROAD WIDTH & Discrete & Four categories: 1 LANE, 2 LANE, 3 LANE, 4 LANE, 5 LANE, 6+ LANE \\
& CROSS WALK & Discrete & Two categories: \textsc{w/o crosswalk}, \textsc{with crosswalk} \\
& Dynamic PARKED VEHICLE (PV) DENSITY & Discrete & Four categories: \textsc{w/o pv}, \textsc{low pv}, \textsc{mid pv}, \textsc{high pv} \\
& PEDESTRIAN DENSITY & Discrete & Three categories: \textsc{low ped}, \textsc{mid ped}, \textsc{high ped} \\
& TRAFFIC DENSITY & Discrete & Three categories: \textsc{low traf}, \textsc{mid traf}, \textsc{high traf} \\
& LEADING VEHICLE (LV) & Discrete & Two categories: \textsc{w/o lv}, \textsc{with lv} \\
& TIME & Discrete & Four categories: \textsc{time6-10}, \textsc{time10-16}, \textsc{time16-20}, \textsc{time20-6} \\
& WEATHER & Discrete & Two categories: \textsc{sunny}, \textsc{rain} \\
\hline
\end{tabular}
\end{table}
\end{center}

\footnotetext{The sidewalk type was defined as degree of a border between the car driving corridor and the footpath. \textsc{side1} was defined as the condition that the car driving corridor and the footpath are not completely separated. \textsc{side2} was defined as the condition that the car driving corridor and the footpath are separated by a white line. \textsc{side3} was defined as the condition that the car driving corridor and the footpath are clearly separated by a ramp, and \textsc{side4} was defined as the condition that the car driving corridor and the footpath are clearly separated by guardrails.}

\footnotetext{The parked vehicle density was defined as the number of parked vehicles present in the approximately 10 s before the near-miss event occurred. The LOW PV was defined as less than two vehicles, MID PV was defined as three to five vehicles, HIGH PV was defined as 6 or more vehicles.}

\footnotetext{The pedestrian density was defined as the number of pedestrians present in the approximately 10 s before the near-miss event occurred. The LOW PED was defined as less than two pedestrians, MID PED was defined as three to nine pedestrians, HIGH PED was defined as 10 or more pedestrians.}

\footnotetext{The traffic density was defined as the number of traffic present in the approximately 10 s before the near-miss event occurred. The LOW PED was defined as less than two running vehicles, MID PED was defined as three to nine running vehicles, HIGH PED was defined as 10 or more running vehicles.}

\footnotetext{The leading vehicle (LV) was defined as a vehicle in front of the own vehicle, and the distinction was made based on whether or not there was a LV in front of the own vehicle.}
2.5. Performance measure

The logistic regression model provides a two-class classification based on a threshold. To judge the model’s performance, four possible outcomes were calculated: true positives, true negatives, false positives, and false negatives. True positives were defined as the model correctly determining that a cyclist moved rapidly onto the road from a blind spot. True negatives were defined as the model correctly determining that a pedestrian moved rapidly onto the road from a blind spot. False negatives occurred when the model determined that a pedestrian moved rapidly onto the road from a blind spot, when in fact a cyclist moved rapidly onto the road from a blind spot. False positives occurred when the model determined that a cyclist moved rapidly onto the road from a blind spot, when in fact a pedestrian moved rapidly onto the road from a blind spot. Based on these four possible outcomes, the accuracy, precision, recall, and specificity were then calculated. The area under the curve (AUC) was also calculated based on the receiver operating characteristic (ROC). Regarding the two-class classification, this study employed a threshold value for maximizing accuracy.

$k$-fold cross-validation was conducted to evaluate the generalization performance of the constructed predictive model. In this study, as can be seen in Table 5, $k$ was set to 5. The following procedure was followed for each of the $k$ folds. The model was trained using $k-1$ of the folds as the training data. The constructed model was then validated using the remaining data. It was thus used as test data to calculate performance measures such as accuracy. The average values of the performance measures calculated in the loop of the $k$-fold cross-validation were evaluated.

| Table 5. Number of data in analysis and validation. |
|----------------|----------------|
| Pedestrian | Cyclist |
| Analysis | 200 | 200 |
| Validation | 50 | 50 |

3. Results

The descriptive statistical values for the environmental elements are listed in Table 6. To identify the significant environmental elements that affected pedestrian or cyclist road-crossing behaviour, logistic regression analysis was employed using the R software package (R 4.0.1). The result of the logistic regression analysis is shown in Table 7. This represents one result from a loop of the 5-fold cross-validation. The explanatory variables (environmental elements) were chosen using the stepwise method. The significance of the explanatory variables (environmental elements) was considered using the $p$-values. In this study, we set the alpha levels at $p < .05$ in order to interpret the significant results. Thus, it was found that the chosen variables were statistically significant with the exception of URBAN. This means that the partial regression coefficient was not statistically zero. The Hosmer–Lemeshow test was conducted, and the test revealed that the model was well-calibrated ($p > 0.05$).

From the analysis, it was found that there were four significant environmental elements namely ONE WAY, BOTH WAY, MID PED, and W/O LEADING VEHICLE (LV) for the pedestrian road-crossing behaviour. These environmental elements indicated a negative partial regression coefficient. Additionally, the upper limits of the odds ratio (Upper CL) were less than 1. This indicated an increased likelihood of a pedestrian initiating a road-crossing from a blind spot. For the MID PED condition, the results showed that pedestrians were more likely to initiate a road-crossing. As can be seen in Table 6, the number of events in which pedestrians initiated a road-crossing under the MID PED condition was large compared to cyclists (54 pedestrians vs. 23 cyclists). Typically, under these circumstances, cyclist movement behaviours are limited compared to those of pedestrians. For the WITH LV condition, the results also showed that pedestrians were more likely to initiate a road-crossing. The number of events in which pedestrians initiated a road-crossing under the WITH LV condition was large compared to cyclists (57 pedestrians vs. 27 cyclists). The results suggested that pedestrian behaviour could be characterized by quickness and flexibility.

In addition, it was found that there were nine significant environmental elements associated with cyclist road-crossing behaviour. These were namely SIDE1, SIDE2, SIDE3, 2- or 3-FORKED ROAD, 4- or 5-FORKED ROAD, 1 LANE, WITH CROSSWALK, SUNNY, and TIME16-20. These environmental elements indicated a positive partial regression coefficient, and the lower limits of odds ratio (Lower CL) were more than 1. The likelihood of a cyclist initiating a road-crossing from a blind spot was higher than that of a pedestrian initiating a road crossing from a blind spot. For the SIDE1, SIDE2, and SIDE3 conditions, the results showed that cyclists were more likely to initiate a road-crossing. SIDE4 was defined as the condition in which the car driving corridor and the footpath were clearly separated by guardrails. In general, pedestrians have more freedom of movement than cyclists. As can be seen in Table 6, the number of events in which pedestrians initiated a road-crossing under the SIDE4 condition was large compared to cyclists (62 pedestrians vs. 30 cyclists). Thus, the results suggested that SIDE1, SIDE2, and SIDE3 elements could contribute to cyclist road-crossing behaviour. For the 2- or 3-FORKED ROAD and 4- or 5-FORKED ROAD conditions, the results also showed that cyclists were more likely to initiate a road-crossing. The number of
Table 6. Descriptive statistical values for environmental elements (pedestrian: \( n = 250 \), cyclist: \( n = 250 \)).

| Property       | Environmental elements | Dart-out object: pedestrian | Dart-out object: cyclist |
|----------------|------------------------|-----------------------------|-------------------------|
| Static         | \( Y_{gap} \)          | \( M = 2.61, SD = 1.07 \)   | \( M = 3.04, SD = 1.36 \) |
| AREA           |                        |                             |                         |
| RESIDENTIAL    | \( n = 69 \)           | \( n = 94 \)                |                         |
| URBAN          | \( n = 135 \)          | \( n = 77 \)                |                         |
| RURAL          | \( n = 46 \)           | \( n = 77 \)                |                         |
| OTHER          | \( n = 0 \)            | \( n = 2 \)                 |                         |
| ROAD           |                        |                             |                         |
| OTHER          | \( n = 0 \)            | \( n = 3 \)                 |                         |
| ONE WAY        | \( n = 61 \)           | \( n = 39 \)                |                         |
| BOTH WAY       | \( n = 189 \)          | \( n = 208 \)               |                         |
| SIDEWALK       |                        |                             |                         |
| SIDE1          | \( n = 36 \)           | \( n = 58 \)                |                         |
| SIDE2          | \( n = 103 \)          | \( n = 114 \)               |                         |
| SIDE3          | \( n = 49 \)           | \( n = 48 \)                |                         |
| SIDE4          | \( n = 62 \)           | \( n = 30 \)                |                         |
| INTERSECTION   |                        |                             |                         |
| 2- or 3-FORKED ROAD | \( n = 48 \)   | \( n = 64 \)                |                         |
| 4- or 5-FORKED ROAD | \( n = 92 \)   | \( n = 141 \)               |                         |
| STRAIGHT       | \( n = 110 \)          | \( n = 45 \)                |                         |
| ROAD WIDTH     |                        |                             |                         |
| 1 LANE         | \( n = 156 \)          | \( n = 175 \)               |                         |
| 2 LANE         | \( n = 68 \)           | \( n = 67 \)                |                         |
| 3 LANE         | \( n = 1 \)            | \( n = 4 \)                 |                         |
| 4 and MORE LANE | \( n = 25 \)       | \( n = 4 \)                 |                         |
| CROSS WALK     |                        |                             |                         |
| W/O CROSSWALK  | \( n = 205 \)          | \( n = 180 \)               |                         |
| WITH CROSSWALK | \( n = 45 \)           | \( n = 70 \)                |                         |
| Dynamic        |                        |                             |                         |
| PARKED VEHICLE (PV) DENSITY | \( n = 152 \)   | \( n = 167 \)               |                         |
| W/O PV         | \( n = 76 \)           | \( n = 74 \)                |                         |
| LOW PV         | \( n = 16 \)           | \( n = 5 \)                 |                         |
| MID PV         | \( n = 6 \)            | \( n = 4 \)                 |                         |
| PEDESTRIAN DENSITY | \( n = 167 \)   | \( n = 214 \)               |                         |
| LOW PED        | \( n = 54 \)           | \( n = 23 \)                |                         |
| HIGH PED       | \( n = 29 \)           | \( n = 13 \)                |                         |
| TRAFFIC DENSITY |                        |                             |                         |
| LOW TRAF       | \( n = 212 \)          | \( n = 227 \)               |                         |
| MID TRAF       | \( n = 33 \)           | \( n = 20 \)                |                         |
| HIGH TRAF      | \( n = 5 \)            | \( n = 3 \)                 |                         |
| LEADING VEHICLE (LV) | \( n = 193 \) | \( n = 223 \)               |                         |
| W/O LV         | \( n = 57 \)           | \( n = 27 \)                |                         |
| WITH LV        |                         |                             |                         |
| Other          |                        |                             |                         |
| TIME           |                        |                             |                         |
| TIME6-10       | \( n = 42 \)           | \( n = 56 \)                |                         |
| TIME10-16      | \( n = 80 \)           | \( n = 78 \)                |                         |
| TIME16-20      | \( n = 46 \)           | \( n = 68 \)                |                         |
| TIME20-60      | \( n = 82 \)           | \( n = 48 \)                |                         |
| WEATHER        |                        |                             |                         |
| SUNNY          | \( n = 218 \)          | \( n = 238 \)               |                         |
| RAIN           | \( n = 32 \)           | \( n = 12 \)                |                         |

events in which pedestrians initiated a road-crossing under the STRAIGHT condition was large compared to cyclists (110 pedestrians vs. 45 cyclists). In contrast, the number of events in which cyclists initiated a road-crossing under the INTERSECTION condition was large compared to pedestrians (2- or 3-FORKED ROAD; 48 pedestrians vs. 64 cyclists, 4- or 5-FORKED ROAD; 92 pedestrians vs. 141 cyclists).

The predictive model used to classify between the two scenarios involving either a pedestrian or a cyclist initiating road-crossing behaviour was constructed using the logistic regression model and incorporated the influence of predictor factors (environmental elements). The four possible outcomes in judging the model’s performance are listed in Table 8. From Table 8, it can be seen that the model demonstrated an accuracy of 67.0%. Table 9 shows the results of 5-fold cross-validation along with a list of the descriptive statistical values for accuracy, precision, recall, and specificity. The model validation demonstrated an accuracy of 69% on average, indicating that the binary logit model possessed moderately accurate predictive power (i.e. the AUC was more than 0.7 on average).

4. Discussion

This study evaluated the influence that driving context-altering environmental elements exert on pedestrian and cyclist road-crossing behaviour. This was done through the construction of a logistic regression model using near-miss incident data. The data included driving data that were recorded automatically 10 s before
The study demonstrated the feasibility of constructing a predictive model to identify covert hazards by using the set of environmental elements that altered the driving context based on the driver's point of view. The results essentially supported this hypothesis.

This study had several limitations. The question of whether the information that was used could indeed be obtained from a vehicle's on-board sensors remains to be addressed. The use of obtainable information is expected from a system's development perspective. The effect that missing information has on prediction accuracy has not yet been examined. Due to the use of annotations, the information used in this study was limited. Thus, the use of image information also has the potential to enable the extraction of more information or features compared to the set of environmental elements that we considered in this study. It is also necessary to consider different methods, such as deep learning techniques, to improve the accuracy of the predictive model.

5. Conclusions

Risk management is the process of identifying hazards, assessing risks, and controlling risks to attain safety. The purpose of this study was to evaluate the influence that driving context-altering environmental elements exert on the road-crossing behaviour of pedestrians and cyclists. Thus, this study attempted to identify covert hazards (concealed pedestrians and cyclists). To investigate this, a near-miss incident database was used. Through a logistic regression analysis incorporating the annotations recorded in the database, we constructed a predictive model to identify covert hazards. The study demonstrated the feasibility of constructing a predictive model to identify covert hazards by using a set of environmental elements that altered the driving context based on a driver's point of view. Identifying

### Table 7. Results of Logistic regression analysis.

| Partial regression coefficient | S.E. | P-value | OR (95% CI) |
|-------------------------------|------|---------|-------------|
| URBUN | -0.52 | 0.26 | 0.051 | 0.59 (0.34–1.00) |
| ONE WAY | -4.10 | 0.72 | < 0.001 | 0.01 (0.01–0.06) |
| BOTH WAY | -3.63 | 0.62 | < 0.001 | 0.02 (0.01–0.08) |
| SIDE 1 | 1.21 | 0.45 | 0.007 | 3.36 (1.38–8.15) |
| SIDE 2 | 0.90 | 0.36 | 0.013 | 2.46 (1.20–5.05) |
| SIDE 3 | 0.88 | 0.39 | 0.026 | 2.41 (1.10–5.26) |
| 2- or 3-FORKED ROAD | 1.31 | 0.34 | < 0.001 | 3.72 (1.91–7.27) |
| 4- or 5-FORKED ROAD | 1.01 | 0.30 | < 0.001 | 2.77 (1.51–5.07) |
| 1 LANE | 0.59 | 0.28 | 0.036 | 1.82 (1.03–3.19) |
| WITH CROSSWALK | 1.10 | 0.32 | < 0.001 | 3.01 (1.60–5.67) |
| MID PED | -1.25 | 0.35 | < 0.001 | 0.28 (0.14–0.57) |
| W/O LEADING VEHICLE | -0.93 | 0.33 | 0.004 | 0.39 (0.20–0.74) |
| SUNNY | 1.94 | 0.47 | < 0.001 | 6.97 (2.72–17.8) |
| TIME16-20 | 0.88 | 0.28 | 0.001 | 2.43 (1.39–4.24) |

Note: We set alpha levels at p < .05 for interpreting significant results.

### Table 8. Four possible outcomes.

| Model’s judgment | Pedestrian initiates a road-crossing | Cyclist initiates a road-crossing |
|------------------|------------------------------------|----------------------------------|
| True state       | 36                                 | 14                               |
| Pedestrian initiates a road-crossing | 19 | 31 |
| Cyclist initiates a road-crossing | 36 | 14 |

### Table 9. Performance measure in 5-fold cross-validation.

| Max. | Min. | Mean | SD | n |
|------|------|------|----|---|
| Accuracy | 0.74 | 0.65 | 0.69 | 0.03 | 5 |
| Precision | 0.74 | 0.60 | 0.68 | 0.05 | 5 |
| Recall | 0.92 | 0.58 | 0.74 | 0.14 | 5 |
| Specificity | 0.80 | 0.38 | 0.64 | 0.16 | 5 |
| AUC | 0.78 | 0.64 | 0.71 | 0.05 | 5 |
the contextual information on road environment elements that contribute to the occurrence of pedestrian and cyclist road-crossing behavior would be useful for acquiring hazard-anticipatory knowledge. Further investigation is required that could involve the addition of annotations along with the incorporation of additional training data to improve the prediction accuracy. The proposed future work is as follows:

- In order to improve safety, it would be useful to know from which blind spot elements other road users initiate a road-crossing. A future research challenge is the construction of a model to identify the priority of blind spot elements.
- It would be useful to know the velocity of other road users at a road-crossing. A future research challenge is the construction of a model to identify the velocity of covert hazards.
- A vital part of the development of an ADAS incorporating hazard-anticipatory knowledge is the construction of the framework for context-sensitive hazard anticipation to predict the potential risk of situations based on the driving context and the driver's behavioral state.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Notes on contributors

Yuichi Saito received the BS and MS degrees in mechanical engineering from the Shibaura Institute of Technology in 2008, and 2010, and the Doctor's degree in engineering from University of Tsukuba in 2015. Since 2019, he has been with the University of Tsukuba, where he is an Assistant Professor with the Faculty of Engineering, Information and Systems. His research interests include smart collaborations between human and machine, shared control, and vehicle dynamics.

Makoto Itoh received the BS, MS, and Doctor's degrees in engineering from University of Tsukuba in 1993, 1995, and 1999, respectively. Since 2013, he has been a Professor with the Faculty of Engineering, Information and Systems, University of Tsukuba. His research interests include enhancement of operator's situation awareness, adaptive automation, and building of appropriate trust as well as prevention of over-trust and distrust.

Takesato Fushima received his BS and MS degrees from the Department of Decision Science and Technology, Tokyo Institute of Technology, Japan, in 2012 and 2014, respectively. He joined Mazda Motor Corporation in 2014. Currently, he is engaged in the development of path planning for automated driving at the Technical Research Center. His research interests include driving risk prediction.

Takashi Sugano received his BS degrees from Tokyo University of Science and MS degrees from The University of Tokyo, Japan, in 1993 and 1996, respectively. In 1996, he joined Toshiba Corporation, he worked on the development of control algorithms for space robots in the space development division. He joined Mazda Motor Corporation in 2005. Currently, he is engaged in the development of path planning and vehicle dynamics control for automated driving at the Technical Research Center.

Yasunori Yamamoto graduated from the Department of Electrical and Electronic Engineering, college of Science and Engineering, Ritsumeikan University. After joining Mazda Motor Corporation in 1988, engaged in preventive safety development work, and is currently a Senior principal engineer of Mazda Technical Research Center. He is engaged in research on autonomous driving technology and next-generation HMI.

ORCID
Yuichi Saito http://orcid.org/0000-0002-6264-7075

References
[1] National Police Agency. Statistics of road traffic accident in 2019; 2019.
[2] Rankavat S, Tiwari G. Pedestrians perceptions for utilization of pedestrian facilities – Delhi, India. Transp Res Part F Traffic Psychol Behav. 2016;42:495–499.
[3] Reason J. Human error. New York: Cambridge University Press; 1990.
[4] NHTSA. National motor vehicle crash causation survey: report to congress. National Highway Traffic Safety Administration Technical Report DOT HS8; 2008. (Tech. rep.).
[5] Ortiz NC, Ramnarayan M, Mizenko K. Distraction and road user behavior: an observational pilot study across intersections in Washington, D.C. J Transp Health. 2017;7, Part A:13–22.
[6] Inagaki T. Design of human–machine interactions in light of domain-dependence of human-centered automation. Cogn Technol Work. 2006;8(3):161–167.
[7] Borowsky A, Oron-Gilad T, Parmet Y. The role of driving experience in hazard perception and categorization: a traffic-scene paradigm. World Acad Sci. 2010;4(66):209–305.
[8] Ni Y, Wang M, Sun J, et al. Evaluation of pedestrian safety at intersections: a theoretical framework based
on pedestrian–vehicle interaction patterns. Accid Anal Prev. 2016;96:118–129.

[9] Ortiz NC, Ramnarayan M, Mizenko K. Distraction and road user behavior: an observational pilot study across intersections in Washington, D.C.. J Transp Health. 2017;7, Part A:13–22.

[10] Osama A, Sayed T. Evaluating the impact of connectivity, continuity, and topography of sidewalk network on pedestrian safety. Accid Anal Prev. 2017;107 (Supplement C):117–125.

[11] Wang X, Yang J, Lee C, et al. Macro-level safety analysis of pedestrian crashes in Shanghai, China. Accid Anal Prev. 2016;96(Supplement C):12–21.

[12] Siddiqui C, Abdel-Aty M, Choi K. Macroscopic spatial analysis of pedestrian and bicycle crashes. Accid Anal Prev. 2012;45(Supplement C):382–391.

[13] Abdel-Aty M, Lee J, Siddiqui C, et al. Geographical unit based analysis in the context of transportation safety planning. Transp Res Part A Policy Pract. 2013;49(Supplement C):62–75.

[14] Wang X, Wu X, Abdel-Aty M, et al. Investigation of road network features and safety performance. Accid Anal Prev. 2013;56(Supplement C):22–31.

[15] Chen P, Zhou J. Effects of the built environment on automobile-involved pedestrian crash frequency and risk. J Transp Health. 2016;3(4):448–456.

[16] Siciliano B, Khatib O, editors. Handbook of robotics. New York: Springer; 2008.

[17] McLaughlin SB, Hankey JM. Naturalistics driving study: linking the study data to the roadway information database. Virginia Tech Transportation Institute; 2015. (Tech. rep.).

[18] Raksincharoensak P, Inoue H. Safety cushion: context-sensitive hazard anticipation-objectified driving behavior of experienced and careful drivers for developing context-sensing driving assistance systems. In: Proceedings of the 4th International Symposium on Future Active Safety Technology Toward Zero Traffic Accidents, Nara, Japan; 2017. 6 pages.

[19] Saito Y, Raksincharoensak P, Inoue H, et al. Context-sensitive hazard anticipation based on driver behavior analysis and cause-and-effect chain study. In: Proceedings of the 14th International Symposium on Advanced Vehicle Control, Beijing, China; 2018. 6 pages.

[20] Inoue H, El-Hajj M, Freudenmann T, et al. Validation methodology to establish safe autonomous driving algorithms with a high driver acceptance using a virtual environment. In: Proceedings of the 5th International Symposium on Future Active Safety Technology Towards Zero-Traffic Accident, Blacksburg, VA; 2019. 6 pages.