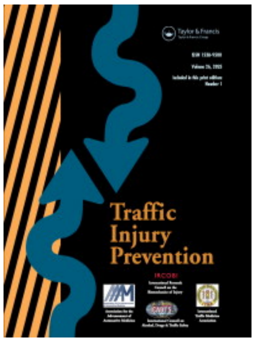

Traffic Injury Prevention

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/gcpi20

Predicting pedestrian-vehicle interaction severity at unsignalized intersections

Kaliprasana Muduli & Indrajit Ghosh

TRAFFIC INJURY PREVENTION
https://doi.org/10.1080/15389588.2024.2404713

Predicting pedestrian-vehicle interaction severity at unsignalized intersections

Kaliprasana Muduli and Indrajit Ghosh

Department of Civil Engineering, Indian Institute of Technology (IIT) Roorkee, Roorkee, India

ABSTRACT

Objectives: This study aims to develop and validate a novel deep-learning model that predicts the severity of pedestrian-vehicle interactions at unsignalized intersections, distinctively integrating Transformer-based models with Multilayer Perceptrons (MLP). This approach leverages advanced feature analysis capabilities, offering a more direct and interpretable method than traditional models.

Methods: High-resolution optical cameras recorded detailed pedestrian and vehicle movements at study sites, with data processed to extract trajectories and convert them into real-world coordinates *via* precise georeferencing. Trained observers categorized interactions into safe passage, critical event, and conflict based on movement patterns, speeds, and accelerations. Fleiss Kappa statistic measured inter-rater agreement to ensure evaluator consistency. This study introduces a novel deep-learning model combining Transformer-based time series data capabilities with the classification strengths of a Multilayer Perceptron (MLP). Unlike traditional models, this approach focuses on feature analysis for greater interpretability. The model, trained on dynamic input variables from trajectory data, employs attention mechanisms to evaluate the significance of each input variable, offering deeper insights into factors influencing interaction severity.

Results: The model demonstrated high performance across different severity categories: safe interactions achieved a precision of 0.78, recall of 0.91, and F1-score of 0.84. In more severe categories like critical events and conflicts, precision and recall were even higher. Overall accuracy stood at 0.87, with both macro and weighted averages for precision, recall, and F1-score also at 0.87. The variable importance analysis, using attention scores from the proposed transformer model, identified 'Vehicle Speed' as the most significant input variable positively influencing severity. Conversely, 'Approaching Angle' and 'Vehicle Distance from Conflict Point' negatively impacted severity. Other significant factors included 'Type of Vehicle', 'Pedestrian Speed', and 'Pedestrian Yaw Rate', highlighting the complex interplay of behavioral and environmental factors in pedestrian-vehicle interactions.

Conclusions: This study introduces a deep-learning model that effectively predicts the severity of pedestrian-vehicle interactions at crosswalks, utilizing a Transformer-MLP hybrid architecture with high precision and recall across severity categories. Key factors influencing severity were identified, paving the way for further enhancements in real-time analysis and broader safety assessments in urban settings.

ARTICLE HISTORY
Received 24 May 2024
Accepted 11 September 2024

KEYWORDS
Pedestrian safety; traffic modeling; deep learning; unsignalized intersections; variable importance

Introduction

Road traffic crashes continue to be a major global concern, causing significant mortality and morbidity among pedestrians, motorcyclists, and cyclists, collectively known as vulnerable road users (VRUs). Recent global data indicate that these groups constitute over half of all traffic-related fatalities (WHO 2023). In India, the situation is particularly alarming, with an annual average of 150,000 fatalities, many occurring at junctions devoid of traffic control measures (MORTH 2022). Research indicates that many pedestrian fatalities in India happen at uncontrolled crossing points (Bansal et al. 2018; Mukherjee and Mitra 2019), underscoring the urgent need for effective safety interventions.

These concerning statistics highlight the necessity for innovative solutions such as predictive models for pedestrian-vehicle interactions. By leveraging data on the dynamics of these interactions, such models can significantly enhance our ability to predict and mitigate the severity of these encounters. Advanced predictive models facilitate the development of integrated pedestrian safety systems that utilize Infrastructure to Vehicle (I2V) and Infrastructure to Pedestrian (I2P) technologies. Through smartphones or wearable devices, these systems offer real-time insights into potential hazards, significantly enhancing pedestrian situational awareness. Historically, studies on pedestrian road-crossing safety have fallen into two categories: crash-based, which analyze previous road crashes to predict future events,

CONTACT Indrajit Ghosh indrafce@iitr.ac.in, indrajit.ghosh@ce.iitr.ac.in Transportation Engineering Group, Department of Civil Engineering, Indian Institute of Technology (IIT) Roorkee, Roorkee 247667, Uttarakhand, India.
Associate editor Alessandro Calvi oversaw the review of this article.
Supplemental data for this article can be accessed online at https://doi.org/10.1080/15389588.2024.2404713.

and conflict-based, which utilize data from non-crash interactions to predict and measure potential conflicts (Shahdah et al. 2014; Formosa et al. 2020; Mohammadian et al. 2021; Nie et al. 2021). Conflict-based studies define traffic conflicts as situations where the paths of road users intersect in such a way that could lead to a crash unless corrective actions are taken (Hydén and Linderholm 1984), providing a proactive means to assess traffic safety.

Real-time prediction of crash probability significantly enhances traffic safety by allowing proactive measures to be implemented before accidents occur. Studies have shown that machine learning models such as Support Vector Machine (SVM) and Multilayer Perceptron (MLP) can effectively predict crash occurrences by analyzing real-time traffic data (Elamrani Abou Elassad et al. 2020). To improve pedestrian safety, Conflict Warning Systems, including sophisticated infrastructure-based systems and Advanced Driver Assistance Systems (ADAS), are crucial. These systems predict the severity of pedestrian-vehicle interactions, enabling them to alert road users about potential high-risk scenarios that could escalate into road crashes (Rasouli and Tsotsos 2020). ADAS, for instance, uses Surrogate Safety Measures (SSMs) such as Time to Collision (TTC) to warn drivers of impending conflicts when, e.g., TTC falls below a certain threshold (Han et al. 2014). TTC is calculated based on the assumption that both pedestrian and vehicle maintain constant speeds on their current paths (Hayward 1972). However, this method often fails to account for the dynamic nature of road interactions, where users frequently adjust their speeds or directions in response to their immediate environment. This limitation can lead to inaccurate risk assessments, either overestimating or underestimating the actual threat of a conflict (Li et al. 2022). Post Encroachment Time (PET), another traditional indicator, evaluates the time difference between when a pedestrian and a vehicle occupy the same space (Cooper 1984). While PET offers insights into past interactions, it cannot predict or assess the potential severity of future conflicts in real-time.

Recent advancements in traffic safety research have explored sophisticated trajectory prediction techniques to enhance pedestrian safety. For instance, Selmoune et al. (2023) developed a Collision Warning Service that integrates closed-circuit television systems (CCTVs), radar, and computer vision to detect objects and predict their paths in real-time using a perspective transformation method and the Kalman filter. Similarly, Ezzati Amini et al. (2022) employed variables such as Minimum Future Relative Distance (MD), Time to Minimum Distance (TMD), and Conflicting Speed (CS) to develop a logit model aimed at enhancing ADAS by predicting the severity of pedestrian-vehicle interactions. Li et al. (2022) proposed a probabilistic framework that employs Gaussian process regression models for predicting the trajectories of both pedestrians and vehicles, integrating maneuver probabilities for comprehensive risk analysis. This method proved to be more effective than traditional TTC in forecasting pedestrian-vehicle conflicts, offering a promising solution for real-time safety management at intersections.

Predictive modeling in pedestrian safety involves multiple steps: predicting the future trajectories of pedestrians and vehicles, identifying potential conflict points, and then calculating traffic conflict indicators like PET or TTC to evaluate the predicted severity of interactions. However, as Golchoubian et al. (2023) noted that trajectory prediction models in mixed traffic conditions can accumulate significant errors over time, leading to unreliable outcomes. To address the complexities and limitations of traditional predictive models, this study introduces a novel approach utilizing a Transformer model architecture known for its exceptional ability to handle time-series data. Unlike traditional methods that often rely on historical data to make linear predictions, our Transformer-based model dynamically incorporates multiple input features, such as speed, direction, and environmental context, in real-time. This allows the model to adapt to changing conditions and more accurately predict the interaction severities. The architecture leverages the Transformer's multi-headed attention mechanism, which focuses on different aspects of the traffic scenario at various time intervals, providing a nuanced understanding of the dynamic interactions between pedestrians and vehicles. This approach enhances the accuracy of predicting immediate risks and significantly improves the interpretability of the model's decisions. By directly analyzing critical features without the need to predict extended future trajectories, our model reduces computational overhead and minimizes the accumulation of prediction errors over multiple time steps. The integration of this advanced modeling technique represents a significant leap forward in real-time traffic safety analysis, offering a robust tool for developing proactive pedestrian safety systems that can seamlessly integrate with modern traffic management technologies.

The objective of this study is to develop and validate an interpretable machine-learning model designed to predict the severity of pedestrian-vehicle interactions at unsignalized intersections, with the ultimate aim of deploying this model to provide timely warnings to road users so that pedestrian crashes can be avoided. This study addresses the complexities and limitations of traditional predictive models by introducing a novel deep learning approach. The proposed model utilizes a feature-based transformer architecture that dynamically incorporates real-time data inputs, allowing for instantaneous adjustments to changing conditions. Unlike conventional methods, this model directly analyzes critical features without predicting extended future trajectories, effectively minimizing error accumulation over multiple time steps and reducing computational overhead. Additionally, this study leverages attention scores generated by the model to dissect and highlight the influence of various features on its predictions. This analysis offers vital insights into the primary factors that determine the severity of pedestrian-vehicle interactions. By processing and prioritizing information in real time, the model not only enhances the interpretability of its decisions but also establishes itself as a more reliable tool for proactive traffic safety management.

Methods

Data sources

To accurately predict the severity of interactions between pedestrians and vehicles, detailed data on the dynamics of

(a) Bahadurgarh Nayagaon-Balaur Intersection: Data collected over a 240-minute duration, with a pedestrian volume of 265 pedestrians per hour and traffic volume of 2400 vehicles per hour.

(b) Khatauli Intersection, Meerut: Data collected over a 240-minute duration with a pedestrian volume of 445 pedestrians per hour and a traffic volume of 3400 vehicles per hour.

Figure 1. Information for selected study sites.

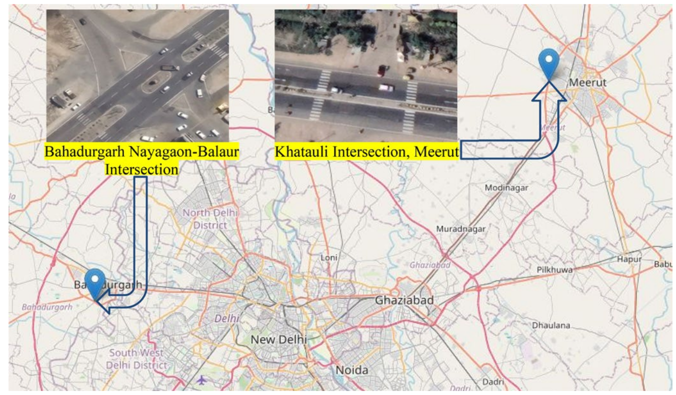

Figure 2. Satellite imagery and mapped locations of the study sites.

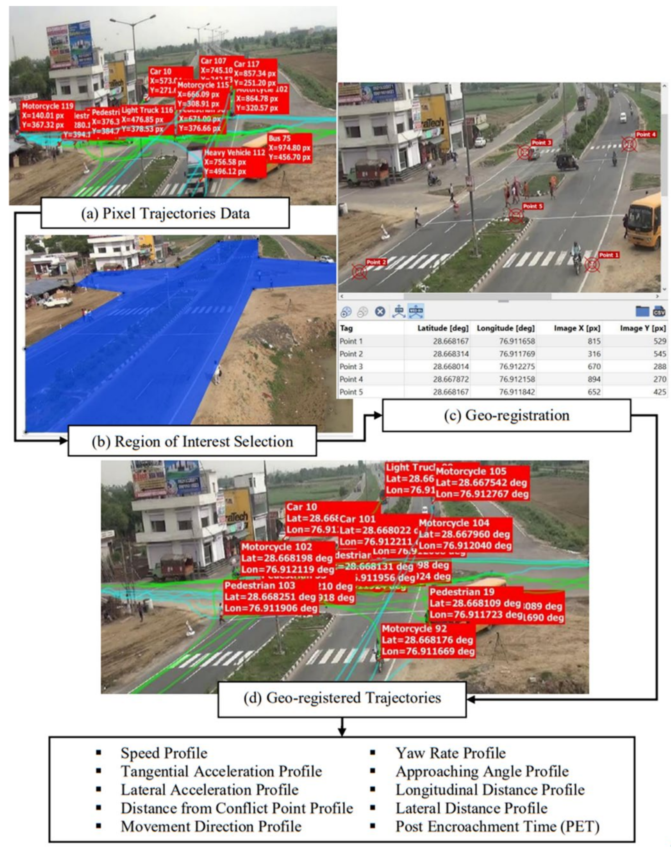

Figure 3. Manual georeferencing in DataFromSky, showing region of interest (ROI) delineation and coordinate mapping.

their encounters is essential, including a substantial number of critical events to encompass the range of severity levels. For this purpose, two unsignalized intersections known for frequent pedestrian incidents were selected for data collection. Information about each location, such as the number of vehicles and pedestrians, is detailed in Figure 1. A total of 8h of video footage, 4h from each intersection, was recorded to document the interaction dynamics. High-quality optical cameras were strategically placed at each site to ensure comprehensive capture of road users' movements. The selection of study sites was grounded in a comprehensive evaluation of historical crash data, high pedestrian activity, and consultations with local traffic management authorities. These intersections were identified due to their significant incidence of pedestrian-vehicle interactions. The study periods were specifically chosen to encompass both morning and evening peak hours, which are well-known for high traffic volume and increased rates of pedestrian-vehicle interactions. This approach ensures that the data collected is representative of typical and critical conditions. The presence of local markets nearby adds to the complexity of traffic dynamics at these locations, contributing to varied pedestrian volumes throughout the day. Additionally, data was collected in clear weather conditions to minimize the impact of environmental factors and the dynamics of pedestrian and vehicle interactions. Both sites feature a dual carriageway with two lanes in each direction, divided by a physical median and equipped with distinct lane markings and a zebra crossing. Figure 2 provides satellite imagery and detailed maps of the study sites.

The data processing was carried out using DataFromSky (DataFromSky 2024), a commercial software, to extract high-resolution trajectories of road users from video footage at each site. Its ability to analyze complex traffic scenarios from video data offers unparalleled detail, making "DataFromSky" particularly suitable for studies requiring precise movement patterns and interaction analyses, and it has been validated in several peer-reviewed studies (Adamec et al. 2020; Dhatbale et al. 2021). Initially, video frames were preprocessed on the DataFromSky web platform, generating trajectory data in pixel units. Subsequently, the DataFromSky viewer application transformed these pixel coordinates into real-world coordinates through manual georeferencing. During this process, the Region of Interest (ROI) was defined, and the geographical coordinates of five selected points within the ROI were inputted. Figure 3 illustrates the detailed steps of this transformation. The resulting geo-registered trajectories were analyzed to extract the necessary input variables and traffic conflict indicators.

Dataset preparation

The trajectory data were analyzed to determine the PET for each pedestrian-vehicle pair with intersecting paths. PET measures the time interval between when one road user leaves a potential conflict zone and the next enters it (Allen et al. 1978). For detailed behavioral observation and interaction severity estimation, only interactions with a PET of less than 10s were considered. This threshold was used because a PET exceeding 10s generally does not indicate a significant interaction. This criterion is supported by the findings of Paul and Ghosh (2020), who observed that the maximum PET values at different unsignalized intersections across India ranged from 8 to 11s, reinforcing the practicality of the 10-second threshold.

Behavior observation and interaction severity estimation

In traffic safety research, methodologies that rely on observer ratings are essential for studying and understanding the dynamics of traffic conflicts and behaviors of road users. Earlier studies, such as those conducted by Van der Horst and Kraay (1986), Ni et al. (2016), and Kathuria and Vedagiri (2020), highlighted the efficacy of video observation in capturing complex interactions between road users. These studies employed video observation techniques to classify pedestrian-vehicle interactions into varying levels of severity. Our study adopts a similar approach, utilizing the AI-enhanced platform DataFromSky Viewer for advanced analysis of trajectories, speeds, and acceleration profiles of all road users (Figure B1, Supplementary material). This tool provides a comprehensive means to assess interaction dynamics.

With DataFromSky Viewer, observers can:

- **Analyze Trajectories:** Enhance understanding of movement patterns and pinpoint potential conflict points.
- **Monitor Speed and Acceleration Profiles:** Evaluate changes in speed and acceleration to understand how road users react to potential conflicts.
- **Estimate Severity:** Use detailed movement data to inform assessments of interaction severity.

Building on the established behaviors identified in the works of Ni et al. (2016) and Kathuria and Vedagiri (2020), our observers classified each interaction into one of three severity levels: safe passage (0), critical event (1), and conflict (2). The typical interactions observed as pedestrians and vehicles approach potential conflict zones include:

1. **Pedestrian Withdrawal:** The pedestrian slows down, stops, or moves away from the conflict zone, yielding to the approaching vehicle.
2. **Pedestrian Acceleration:** The pedestrian increases their speed to cross before the vehicle arrives, indicating a judgment of safe passage.
3. **Path Change by Pedestrian:** The pedestrian alters their path, such as moving behind the vehicle, to avoid interaction, reflecting adaptive safety maneuvers.
4. **Vehicle Yielding:** The vehicle slows or stops to allow the pedestrian to cross, showing recognition of the pedestrian's presence and avoiding potential conflict.
5. **Vehicle Acceleration:** The vehicle speeds up to clear the conflict zone before the pedestrian enters, aiming to minimize interaction time.
6. **Negotiated Passage:** Both the pedestrian and vehicle adjust their movements, achieving a safe passage through mutual adjustment without direct communication.
7. **No Action:** Neither party changes their speed or path, yet the interaction concludes safely without incident.

These behavioral observations, which include subtle cues and context-specific behaviors, were combined with quantitative values of speed, acceleration, trajectory plots for visualizing path changes, and PET values to gauge temporal closeness. This comprehensive approach empowered the observers to assess interaction severity by visually perceiving the scene and utilizing well-informed quantitative knowledge about movement dynamics and traffic conflict indicators.

Classification schemes

Safe passage

Interactions where pedestrians and vehicles navigate the intersection without imminent risk, maintaining a comfortable and safe distance and timing. No sudden maneuvers are required, and the interaction is managed smoothly. These interactions can be associated with high PET values (greater than 3 s) indicating sufficient temporal separation, low or moderate speeds (up to 10 m/s) with minimal or controlled acceleration/deceleration (below 1 m/s^2), and Smooth and predictable trajectories for both pedestrians and vehicles. Observers perceive these interactions as calm and controlled, with clear intent from both parties. Some behavioral examples relevant to safe passages include vehicle yielding, where the vehicle slows or stops well in advance to allow the pedestrian to cross; negotiated passage, involving slight adjustments by both the pedestrian and vehicle for a smooth passage; and no action, where neither party changes speed nor path, yet the interaction concludes safely.

Critical event

Interactions with moderate risk, where timely and appropriate responses are required to avoid conflict. These interactions involve closer temporal and spatial proximity but are managed effectively. These interactions can be associated with moderate PET values (between 1 and 3 s) that indicate potential risk but allow for evasive action, moderate speeds (between 5 to 10 m/s) with some level of acceleration or deceleration (typically between 1 and 3 m/s^2) indicating an active response, and some deviation from a smooth path, indicating adjustments made to avoid conflict. Observers perceive these interactions as heightened alertness and active maneuvers from either party to avoid collision. Some behavioral examples relevant to critical events include pedestrian withdrawal, where the pedestrian stops or moves back to yield to the vehicle; pedestrian acceleration, where the pedestrian speeds up to cross before the vehicle arrives; and path change by the pedestrian, where the pedestrian alters their path to avoid the vehicle.

Conflict

High-risk interactions involving near-misses, where evasive actions are either insufficient or too late. These represent the most severe interactions. These interactions can be associated with low PET values (typically less than 1 s), indicating very close temporal proximity, high speeds (greater than 10 m/s) and/or abrupt acceleration or deceleration (greater than 3 m/s^2), indicating emergency maneuvers, sharp deviations, and erratic paths indicating emergency responses. Observers perceive the interaction as chaotic and dangerous, often involving sudden stops or swerves. Some behavioral examples relevant to conflicts include vehicle acceleration, where the vehicle speeds up to clear the conflict zone before the pedestrian enters, and emergency maneuvers, involving sudden braking or swerving by either party to avoid a collision.

To ensure consistency and reliability in our data, six trained observers evaluated a total of 4,315 interactions, independently assigning severity levels based on the described criteria. The observers were rigorously trained to utilize the DataFromSky Viewer and understand the categorization scheme. The robustness of our severity estimations was enhanced through the use of Fleiss Kappa statistics (Fleiss 1971) to measure inter-rater agreement among the observers. This statistical method helped confirm the reliability of the observers' assessments, ensuring that our findings were based on consistent and accurate evaluations.

Assessment of observer agreement in severity level estimation

In this study, we employed the Fleiss Kappa statistic (Fleiss 1971) to assess the reliability of agreement among multiple observers evaluating the severity of pedestrian-vehicle interactions. We organized the data into an $n \times k$ matrix, where $n = 4{,}315$ denotes the total observed events and $k = 3$ represents the severity categories, with each matrix element reflecting the count of raters assigning each event to a category. We calculated the proportional agreement among raters for each event and derived the overall mean proportional agreement (P_{mean}). Expected agreement by chance (Pe) was determined from the rating proportions across categories. The Fleiss Kappa coefficient, calculated using Equation (1), yielded a value of 0.75.

$$\kappa = \frac{P_{\text{mean}} - P_e}{1 - P_e} \tag{1}$$

This indicates a substantial level of inter-rater agreement beyond what could occur by chance, validating the reliability of our observational methodology as per the standards set by Landis and Koch (1977).

Input variables

To predict the severity of pedestrian-vehicle interactions and to implement a prototype system in an AI-enhanced pedestrian crossing, it is essential that the driver and the pedestrian receive an advance warning for critical interactions θ time units before they occur. Considering the reaction time of drivers and pedestrians, θ was established as 2 s. The trajectories of interacting road users 2s before reaching their minimum separation distance were analyzed to extract the necessary input variables, which represent Pedestrian Dynamics, Vehicle Dynamics, and Interaction Dynamics, presented in Table 1, Table 2, and Table 3, respectively. For training the proposed machine learning model Pedestrian-Vehicle Interaction Severity (Ordinal: Class 0 - Safe Passage, Class 1 - Critical Event, Class 2 - Conflict), is used as the dependent variable. The prediction horizon of 2s was adopted, drawing on the findings of Nie et al. (2021), which says that pedestrians' decision and execution times for avoidance actions range from 1.69 to 2.08 s.

To standardize the input data for the proposed model, the mode of the lengths of all extracted time series, calculated to be 60, has been determined as the standard length for training. Interactions that yielded time series shorter than the mode were padded with zeros, whereas those exceeding the mode were truncated to ensure uniformity. This methodological rigor ensures that the model is trained on consistently formatted data, enhancing its predictive accuracy and reliability in real-world applications.

Position coordinates from the trajectory data, given in latitude and longitude, necessitated the use of the Haversine formula to calculate distances essential for determining input variables such as approach angle, longitudinal and lateral distances, speed, and acceleration. The Haversine distance (d) between two points is calculated using the following formula as presented in Equation (1).

Table 1. Input variables representing pedestrian dynamics.

Variable	Data type	Feature description
Pedestrian Speed	Continuous variable	Measured in meters per second (m/s)
Pedestrian Tangential Acceleration	Continuous variable	Measured in meters per second squared (m/s^2)
Pedestrian Lateral Acceleration	Continuous variable	Measured in meters per second squared (m/s^2)
Pedestrian Distance from Conflict Point	Continuous variable	Measured in meters (m)
Pedestrian Direction	Continuous variable	Measured in degrees
Pedestrian Yaw Rate	Continuous variable	Measured in degrees per second (°/s)

Table 2. Input variables representing vehicle dynamics.

Variable	Data type	Feature description
Vehicle Speed	Continuous variable	Measured in meters per second (m/s)
Vehicle Tangential Acceleration	Continuous variable	Measured in meters per second squared (m/s^2)
Vehicle Lateral Acceleration	Continuous variable	Measured in meters per second squared (m/s^2)
Vehicle Distance from Conflict Point	Continuous variable	Measured in meters (m)
Vehicle Direction	Continuous variable	Measured in degrees
Vehicle Yaw Rate	Continuous variable	Measured in degrees per second (°/s)
Type of Vehicle	Nominal (Categorical)	Motorcycle (0), Auto Rickshaw (1), Car/Taxis/Vans & LMV (2), Bus (3), Truck/Lorry (4)

Table 3. Input variables representing interaction dynamics.

Variable	Data type	Feature description
Approaching Angle	Continuous variable	Measured in degrees
Longitudinal Distance	Continuous variable	Measured in meters (m)
Lateral Distance	Continuous variable	Measured in meters (m)

$$d = 2R \times antitan\left(\sqrt{\frac{\left[\sin\left(\frac{lat2 - lat1}{2}\right)\right]^2 + \cos(lat1) \times \cos(lat2) \times \left[\sin\left(\frac{lon2 - lon1}{2}\right)\right]^2}{1 - \left(\left[\sin\left(\frac{lat2 - lat1}{2}\right)\right]^2 + \cos(lat1) \times \cos(lat2) \times \left[\sin\left(\frac{lon2 - lon1}{2}\right)\right]^2\right)}}\right) \tag{1}$$

Where $R = 6.3781 \times 10^6$ m represents the radius of Earth, and lat1, lat2, lon1, and lon2 are the latitudes and longitudes of the two points, respectively.

Deep learning model overview

This study introduces a novel deep-learning model using a Transformer architecture to predict the severity of pedestrian-vehicle interactions at unsignalized crosswalks. The Transformer is excellent at handling time-series data, which is crucial for analyzing the sequence of events in pedestrian-vehicle interactions. The model uses a mechanism called Multi-Head Attention to focus on the most important parts of the data for predicting interaction severity. Further details of the model's components are discussed in the subsequent sections.

Transformer encoder

The Transformer encoder processes all input data, converting sequences of interactions into a format the model can understand. This includes two main parts: multi-headed attention and a fully connected network. The model ensures stability and efficiency through techniques called residual connections and layer normalization. As shown in Figure 4, the encoder processes sequences of data points, each representing different moments in time. The model uses these data points to understand and predict the dynamics of pedestrian-vehicle interactions. It does this by separating the input into three types of information: Queries, Keys, and Values, which help the model assess the relevance of each data point to the severity of the interaction.

Attention mechanism

The attention mechanism of the Transformer allows the model to focus on the most relevant data points by creating a score for each, indicating its importance. These scores help the model prioritize which events and times are crucial for predicting the severity of interactions.

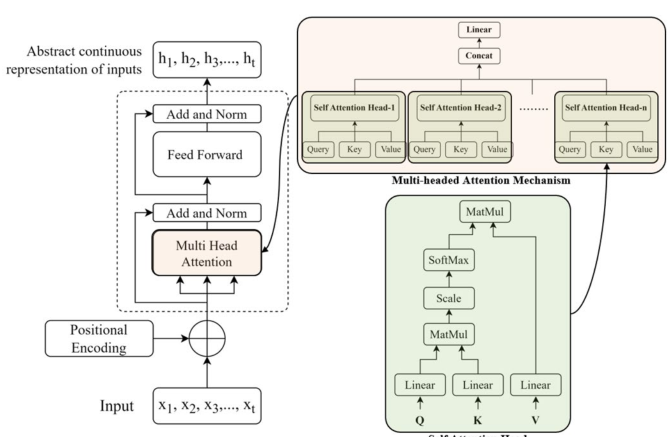

Figure 4. Transform encoder architecture.
Note: MatMul = Matrix Multiplication; SoftMax = Softmax Function; Q = query; K = key; V = value.

Normalization and network Layers

The model uses layer normalization to maintain consistency in processing. It also includes a feed-forward network that refines the information further. A Global Average Pooling layer reduces the computational load by simplifying the output from the Transformer encoder, making the model more efficient.

Dense Layers and output

The network includes several dense layers with activation functions designed to process the refined data effectively. A Dropout layer is incorporated to prevent the model from overfitting. The final output of the model is generated through a Dense layer with a softmax activation function, which classifies the severity of each interaction into categories.

For a detailed technical explanation of the background theory of the Transformer model, please refer to the Appendix (supplementary file).

Calculation of feature importance

Assessing the importance of each variable within the proposed predictive model is crucial for elucidating the factors influencing the severity of pedestrian-vehicle interaction. By calculating attention scores through the Transformer's attention mechanism, it becomes possible to identify which variables the model focuses on and to evaluate their respective impacts on interaction severity, whether positive or negative. This approach offers a detailed insight into how the model processes and prioritizes information, thereby shedding light on the critical factors that drive its predictions. In the Transformer model, the attention mechanism computes the scaled similarity score (S_{ij}) by taking the dot product between the Query (Q_i) of a feature and the Keys (K_j) of all features normalized by dividing the square root of the dimension of the Keys (d_k), as shown in Equation (4).

$$S_{ij} = \frac{Q_i . K_j}{\sqrt{d_k}} \tag{4}$$

The scaled similarity scores (S_{ij}) are normalized using a softmax function to generate the final attention weights (A_{ij}), emphasizing one feature, as shown in Equation (5).

$$A_{ij} = \frac{e^{S_{ij}}}{\sum e^{S_{ik}}} \tag{5}$$

The feature importance is determined by calculating the average attention weight assigned to a feature across all instances in the dataset. The formula used to calculate the importance of a feature i is presented in Equation (6).

$$Importance(i) = \frac{\sum A_{ij}}{N} \tag{6}$$

N represents the number of samples used to calculate the attention weights of the model.

Experiment setup

Data preprocessing for model input

In this study, data preprocessing was conducted to optimize the deep learning model's performance. Numerical variables were standardized by zero-centering their mean and scaling their variance to one, while categorical variables like 'Type of Vehicle' were converted to numerical values through label encoding. To ensure uniform sequence lengths, crucial for consistent model performance, data for each interaction was sequenced and padded as needed.

The dataset, consisting of 4,315 pedestrian-vehicle interactions from eight hours of video across two sites, was divided into training (70%), validation (15%), and testing (15%) sets. The majority of interactions (3,567) were 'Safe Passages', with the remainder being 'Critical events' (507) and 'Conflicts' (214), highlighting a class imbalance. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was employed within the training set to artificially enhance the representation of the less frequent categories. This approach is a popular technique used in the field (Elamrani Abou Elassad et al. 2020; Zhang and Abdel-Aty 2022; Muduli et al. 2024; Muduli and Ghosh 2024) to improve the model's ability to learn from a more balanced distribution of classes, thereby enhancing the robustness and accuracy of the training process.

Model training details

The proposed model combines Transformer encoders and a Multilayer Perceptron (MLP) to predict pedestrian-vehicle interaction severity at unsignalized crossings, processing 16 input features. Each transformer encoder block includes multi-head attention, layer normalization, and a feed-forward network, stacked to form the model's core. Key hyperparameters such as head size, number of attention heads, feed-forward dimension, and encoder blocks are optimized using Keras Tuner's RandomSearch algorithm, which systematically explores the model space. The training process iteratively adjusts MLP unit sizes and dropout rates to identify configurations that maximize validation accuracy, achieving an optimal balance of performance and efficiency. Table 4 details the search space and optimal values for each parameter.

After optimizing the hyperparameters, the model architecture is structured as a sequence-based deep learning framework where input batches containing temporal sequences of pedestrian-vehicle interaction features are processed through a Transformer Encoder. This encoder uses multi-head attention to identify key temporal segments for severity prediction and incorporates a Global Average Pooling (GAP) layer to reduce dimensionality and prevent overfitting. The processed data is then fed into a Multi-Layer Perceptron (MLP), which consists of two layers with 256 and 128 units, respectively, and includes a 10% dropout rate to promote model generalization. The final output is generated by a dense layer with a softmax activation function that classifies the interactions into severity categories: 'Safe passage', 'Critical event', and 'Conflict'. The model's architecture is fine-tuned with a head size of 96, 4 attention heads, a feed-forward dimension of 192, and 2 transformer blocks, ensuring a balance of accuracy and computational efficiency. The details of this architecture and training process are depicted in Figures 5 and 6.

The model uses the Adam optimizer, sparse categorical cross-entropy loss function, and accuracy as the evaluation metric, trained over 110 epochs with a batch size of 32. An early stopping technique monitors the validation loss, ending training if no improvement is seen over ten epochs and reverting to the best model weights. This practice prevents overfitting and optimizes computational resources.

Table 4. Optimal hyperparameters for transformer model.

Parameter	Range/Search Space	Optimal Value
Head Size (H)	$32 \le H \le 128$, $\Delta H = 32$	96
Number of Heads (N_h)	$2 \le N_h \le 8$, $\Delta N_h = 2$	4
Feed Forward Dimension (F)	$64 \le F \le 256$, $\Delta F = 64$	192
Number of Transformer Blocks (B)	$1 \le B \le 5$, $\Delta B = 1$	2
MLP Units (U)	{'64_32', '128_64', '256_128'}	'256_128'
Dropout (D)	$0 \le D \le 0.5$, $\Delta D = 0.1$	0.1
MLP Dropout (D_m)	$0 \le D_m \le 0.5$, $\Delta D_m = 0.1$	0.1

Model testing

The testing phase of the model was systematically executed to ensure a comprehensive evaluation of its performance on unseen data (text data). The trained model with optimum hyperparameters, equipped with its refined weights, was employed to generate predictions for the test dataset. To quantitatively assess the model's predictive efficacy, key performance metrics, as summarized in Table 5, were calculated:

Results

The experimental evaluation of the model yielded promising results, as delineated by a detailed confusion matrix (Figure 7) and comprehensive classification metrics (Table 6). It is important to emphasize that the evaluation of the proposed model was conducted on a test dataset derived exclusively from the original dataset, consisting of non-synthetic data, to ensure an accurate assessment of the model's performance in real-world conditions. For Class 0 (Safe Passage), the model showcased a precision of 0.78, a high recall of 0.91, and an F1-score of 0.84, indicating a strong propensity to correctly identify safe scenarios, despite a slight tendency to misclassify few safe interactions as other classes. A precision of 0.78 for safe passage can be acceptable for a warning system where the

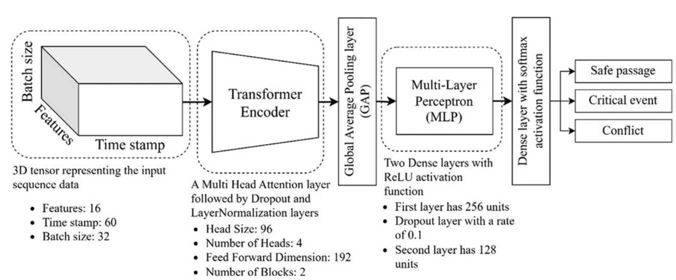

Figure 5. Architecture of the proposed transformer-based model for prediction of pedestrian-vehicle interaction severity.

Table 5. Key performance metrics used for evaluating model's predictive efficacy.

Metric	Formula	Description
Precision	$\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$	Measures the accuracy of the model in identifying only true positives out of all positive predictions.
Recall (Sensitivity)	$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$	Assesses the model's ability to capture all actual positive instances out of the actual condition positives.
F1-Score	$\frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$	Provides a balanced measure of precision and recall, useful in cases of uneven class distributions.
Accuracy	$\frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$	Calculates the proportion of total correct predictions made by the model across all categories.

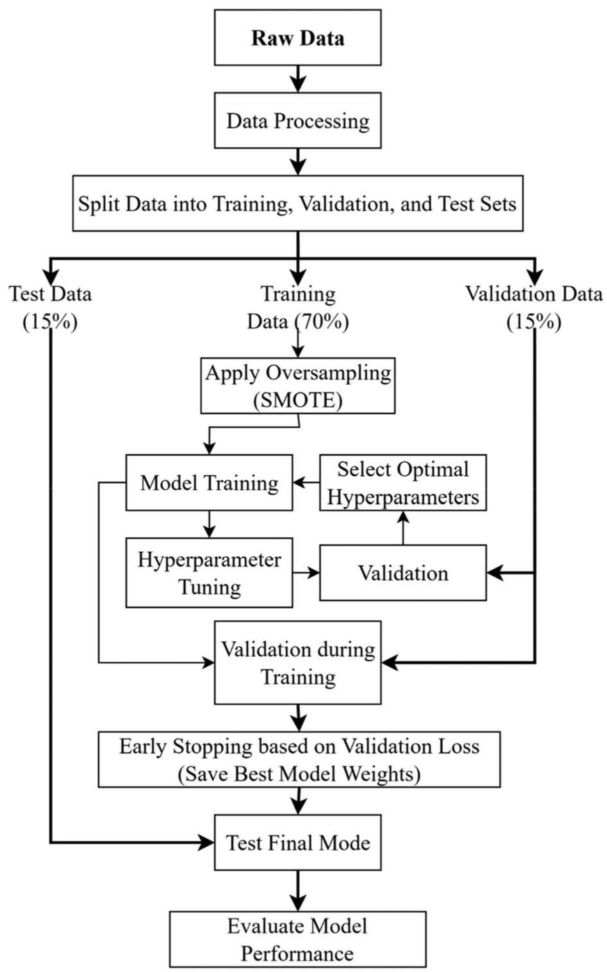

Figure 6. Overview of the model training procedure.

cost of missing a potentially dangerous situation is much higher than the inconvenience of a false alert. This level of precision can increase caution among drivers and pedestrians, potentially reducing accidents while causing minor disruptions. In the case of Class 1 (Critical Event), the model achieved an impressive precision of 0.93, with a recall of 0.85 and an F1-score of 0.89, demonstrating its accuracy in identifying critical events, though with a small margin of missed true events. Class 2 (Conflict) saw the model at its best performance, with a precision of 0.95, recall of 0.86, and an F1-score of 0.90, suggesting exceptional precision in identifying conflicts. The overall accuracy across all classes stood at 0.87, denoting a high degree of correct predictions. The macro and weighted averages for precision, recall, and F1-score were consistently above 0.87, reflecting a uniform performance across classes and indicating the model's reliable predictive quality, even with class imbalance considerations. These results affirm the model's robustness and its applicability for classifying diverse scenarios effectively.

Figure 7. Confusion matrix of model predictions for three classified outcomes.

Table 6. Comprehensive classification metrics for three categories, with accuracy and average values.

	Precision	Recall	F1-Score
Class 0 (Safe Passage)	0.78	0.91	0.84
Class 1 (Critical event)	0.93	0.85	0.89
Class 2 (Conflict)	0.95	0.86	0.9
Accuracy			**0.87**
Macro avg	0.89	0.87	0.88
Weighted avg	0.88	0.87	0.88

Discussion

In assessing factors influencing pedestrian-vehicle interaction severity, the proposed Transformer model's attention mechanism identified 'Vehicle Speed' as having a significant positive impact (0.3765). This correlation is supported by research from Aarts and Van Schagen (2006) and Gargoum and El-Basyouny (2016), which demonstrated that increased vehicle speeds escalate the probability of crashes. 'Approaching Angle' received a negative importance score (0.2952), indicating that narrower angles, which reduce visibility and reaction opportunities, may lead to more severe incidents, as discussed by Belkhouche and Bendjilali (2013) and supported by Harris et al. (2023). 'Vehicle Distance from Conflict Point' also showed a negative score (-0.2683), suggesting that larger distances provide more reaction time, thus reducing severity, a concept reinforced by Matsui and Oikawa (2019). Additionally, 'Type of Vehicle' emerged as a positive influencer (0.2182), consistent with Dey et al. (2021), who found that larger vehicles tend to cause more severe pedestrian road crashes. 'Pedestrian Speed' (0.1583) and 'Pedestrian Yaw Rate' (0.2898) were also significant, indicating that faster movements and sudden directional changes increase severity, echoing Cheng et al. (2022). Figure 8 visually illustrates these influences, showing higher scores for

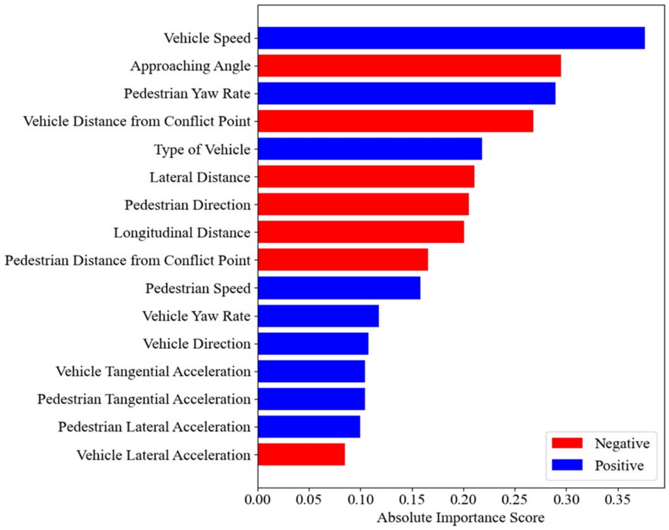

Figure 8. Attention-based variable importance ranking.

factors like vehicle speed and pedestrian yaw rate, which significantly impact severity while highlighting the preventative potential of greater distances from conflict points.

The fact that these findings are generally in line with previous studies suggests that the proposed transformer-based deep learning approach for predicting interaction severity is working correctly, and the model learns to predict the interaction severity in an accurate manner. This alignment reinforces the reliability and validity of the proposed methodology and supports its potential for practical applications in enhancing pedestrian safety.

Conclusions

This study introduces a novel approach to assessing pedestrian-vehicle interaction severity at crosswalks, leveraging a novel deep learning model that uniquely combines Transformer-based architectures with Multilayer Perceptrons (MLP). This model integrates the sophisticated time-series data handling capabilities of Transformer architectures renowned for their effectiveness in capturing sequential dependencies through attention mechanisms, which allow the model to focus on the most relevant features of the data over time. In conjunction with the classification prowess of MLPs, which analyze the transformed features to predict interaction severity, the model provides a robust and interpretable framework. This dual approach enables a more nuanced understanding of complex pedestrian and vehicle dynamics at unsignalized intersections, significantly enhancing prediction accuracy and interpretability compared to existing methods, such as those used by Zhang and Abdel-Aty (2022), who achieved the highest accuracy of 0.723 in predicting pedestrian-vehicle interaction severity using an Extreme Gradient Boosting (XGBoost) model after employing various techniques like random forests, support vector machines, logistic regression, gradient boosting machines, and linear discriminant analysis.

Data was initially collected from eight hours of video surveillance at two unsignalized intersections. High-resolution optical cameras captured detailed movements of pedestrians and vehicles, which were processed using the DataFromSky software to extract trajectories and convert them into real-world coordinates through precise georeferencing. The analysis focused on interactions with PET less than 10 s to identify potentially significant interactions. Trained observers then used the DataFromSky Viewer platform to classify these interactions into three severity categories: safe passage, critical event, and conflict. The reliability of these classifications was confirmed using the Fleiss Kappa statistic, indicating substantial inter-rater agreement with a coefficient (κ=0.75).

The model was trained on dynamic input variables representing detailed pedestrian and vehicle dynamics derived from the trajectory data and balanced using SMOTE. It was then evaluated on a separate test dataset, demonstrating high effectiveness across different severity categories. For safe interactions (Class 0), the model achieved a precision of 0.78, a recall of 0.91, and an F1-score of 0.84. Critical events (Class 1) and conflicts (Class 2) showed even higher precision and recall, illustrating the model's ability to accurately identify and classify the most hazardous situations. Overall accuracy was recorded at 0.87, with consistent macro and weighted averages for precision, recall, and F1-score above 0.87. The variable importance analysis using the model's attention mechanism highlighted key factors affecting interaction severity. 'Vehicle Speed' emerged as a significant factor, with higher speeds increasing severity due to greater impact forces and longer stopping distances. In contrast, 'Approaching Angle' and 'Vehicle Distance from Conflict Point' had negative importance scores, suggesting that narrower angles and greater distances might reduce severity by improving visibility and reaction times. Additionally, 'Type of Vehicle', along with pedestrian behaviors such as 'Pedestrian Speed' and 'Pedestrian Yaw Rate', were significant, linking the type of vehicle to the impact severity and pedestrian behaviors to increased risk due to sudden movements. These findings illustrate the complex interaction of behavioral and environmental factors in pedestrian-vehicle dynamics.

This study has certain limitations, including the omission of environmental factors like weather conditions and architectural features, which could have strengthened the assessments. Additionally, intrinsic factors such as psychological traits and individual differences among traffic participants merit further exploration in subsequent studies. Future research should consider integrating contextual variables like time of day and road surface conditions to enhance the precision of severity predictions. Adapting the model for real-time analysis could enable immediate feedback and warnings at high-risk crosswalks, enhancing pedestrian and driver safety. Moreover, expanding the methodology to account for interactions involving vehicular traffic and cyclists would broaden the scope of safety assessments in urban environments. By deploying this model across varied settings, its effectiveness can be validated across different cultural and traffic contexts, contributing significantly to global pedestrian safety initiatives.

Disclosure statement

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

Funding

This work was supported by the Science and Engineering Research Board (SERB), Government of India, under the Core Research Grant Scheme, awarded for the project "Machine Learning Framework for the Development of an Interactive Pedestrian Crossing to Enhance Pedestrian Safety," Sanction Order No. CRG/2023/003239.

ORCID

Kaliprasana Muduli http://orcid.org/0000-0003-3504-2086
Indrajit Ghosh http://orcid.org/0000-0002-1341-0258

Data availability statement

The datasets from this study are available from the authors upon reasonable request.

References

Aarts L, Van Schagen I. 2006. Driving speed and the risk of road crashes: a review. Accid Anal Prev. 38(2):215–224. doi:10.1016/j.aap.2005.07.004.

Adamec V, Herman D, Schullerova B, Urbanek M. 2020. Modelling of traffic load by the datafromsky system in the smart city concept. In: Lopes N, editor. Smart governance for cities: Perspectives and experiences. EAI/Springer innovations in communication and computing. Cham: Springer. p. 135–152. doi:10.1007/978-3-030-22070-9_7.

Allen BL, Shin BT, Cooper PJ. 1978. Analysis of traffic conflicts and collisions. Transp Res Rec. 667:67–74.

Ankit Bansal, Tripta Goyal, Umesh Sharma. 2018. Pedestrian safety on crosswalks in India – Need of the hour. J. Today's Ideas - Tomorrow's Technol. 6(1):35–46. doi:10.15415/jotitt.2018.61004.

Belkhouche F, Bendjilali B. 2013. Dynamic collision risk modeling under uncertainty. Robotica [Internet]. 31(4):525–537. doi:10.1017/S0263574712000550.

Cheng R, Pan Y, Xie L. 2022. Analysis of vehicle-pedestrian accident risk based on simulation experiments. Math Probl Eng. 2022:1–14. doi:10.1155/2022/7891232.

Cooper PJ. 1984. Experience with traffic conflicts in canada with emphasis on "post encroachment time" techniques. International Calibration Study of Traffic Conflict Techniques [Internet]. 5:75–96. doi:10.1007/978-3-642-82109-7_8.

Deep traffic video analysis – DataFromSky. [accessed 2024 Mar 12]. https://datafromsky.com/.

Dey D, Matviienko A, Berger M, Martens M, Pfleging B, Terken J. 2021. Communicating the intention of an automated vehicle to pedestrians: the contributions of eHMI and vehicle behavior. IT – Information Technology [Internet]. 63(2):123–141. doi:10.1515/ITIT-2020-0025/MACHINEREADABLECITATION/RIS.

Dhatbale R, Bhargava ·, Chilukuri R. 2021. Deep Learning Techniques for Vehicle Trajectory Extraction in Mixed Traffic. J. Big Data Anal. Transp. 3(2):141–157. doi:10.1007/s42421-021-00042-3.

Elamrani Abou Elassad Z, Mousannif H, Al Moatassime H. 2020. Class-imbalanced crash prediction based on real-time traffic and weather data: a driving simulator study. Traffic Inj Prev. 21(3):201–208. doi:10.1080/15389588.2020.1723794.

Ezzati Amini R, Yang K, Antoniou C. 2022. Development of a conflict risk evaluation model to assess pedestrian safety in interaction with vehicles. Accid Anal Prev. 175:106773. doi:10.1016/J.AAP.2022.106773.

Fleiss JL. 1971. Measuring nominal scale agreement among many raters. Psychol Bull. 76(5):378–382. doi:10.1037/h0031619.

Formosa N, Quddus M, Ison S, Abdel-Aty M, Yuan J. 2020. Predicting real-time traffic conflicts using deep learning. Accid Anal Prev. 136:105429. doi:10.1016/j.aap.2019.105429.

Gargoum SA, El-Basyouny K. 2016. Exploring the association between speed and safety: a path analysis approach. Accid Anal Prev. 93:32–40. doi:10.1016/j.aap.2016.04.029.

Golchoubian M, Ghafurian M, Dautenhahn K, Azad NL. 2023. Pedestrian trajectory prediction in pedestrian-vehicle mixed environments: a systematic review. IEEE Trans Intell Transport Syst. 24(11):11544–11567. doi:10.1109/TITS.2023.3291196.

Han IC, Luan BC, Hsieh FC. 2014. Development of autonomous emergency braking control system based on road friction. IEEE International Conference on Automation Science and Engineering. 2014-January: p. 933–937. doi:10.1109/COASE.2014.6899438.

Harris L, Ahmad N, Khattak A, Chakraborty S. 2023. Exploring the effect of visibility factors on vehicle–pedestrian crash injury severity. Transp Res Rec [Internet]. 2677(11):24–35. doi:10.1177/03611981231164070/ASSET/IMAGES/LARGE/10.1177_03611981231164070-FIG3.JPEG.

Hayward JC. 1972. Near-miss determination through use of a scale of danger. Highway Research Record. 384:24–34.

Hydén C, Linderholm L. 1984. The Swedish traffic-conflicts technique. International Calibration Study of Traffic Conflict Techniques. 5:133–139. doi:10.1007/978-3-642-82109-7_12.

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

Landis JR, Koch GG. 1977. The measurement of observer agreement for categorical data. Biometrics. 33(1):159–174. doi:10.2307/2529310.

Li P, Guo H, Bao S, Kusari A. 2022. A Probabilistic Framework for Estimating the Risk of Pedestrian-Vehicle Conflicts at Intersections [Internet]. [accessed 2023 Apr 15]. https://arxiv.org/abs/2207.14145v1.

Matsui Y, Oikawa S. 2019. Situational characteristics of fatal pedestrian accidents involving vehicles traveling at low speeds in Japan. Traffic Inj Prev. 20(sup1):S1–S6. doi:10.1080/15389588.2019.1587166.

Mohammadian S, Haque MM, Zheng Z, Bhaskar A. 2021. Integrating safety into the fundamental relations of freeway traffic flows: a conflict-based safety assessment framework. Anal Methods Accid Res. 32:100187. doi:10.1016/j.amar.2021.100187.

MORTH. 2022. Road accidents in india 2020. Ministry of Road Transport and Highways [Internet]. [accessed 2022 Jul 22]. www.morth.nic.in.

Muduli K, Ghosh I. 2024. Prediction of Vehicular Yielding Intention While Approaching a Pedestrian Crosswalk. [Internet]. [accessed 2024 Jun 23]. doi:10.1177/03611981241252835.

Muduli K, Sahu V, Ghosh I. 2024. Predicting pedestrian movement in unsignalized crossings: a contextual cue-based approach. 2024 16th International Conference on COMmunication Systems & NETworkS (COMSNETS) [Internet]. [accessed 2024 Mar 5]: 204–209. doi:10.1109/COMSNETS59351.2024.10427345.

Mukherjee D, Mitra S. 2019. A comparative study of safe and unsafe signalized intersections from the view point of pedestrian behavior and perception. Accid Anal Prev [Internet]. 132(April):105218. doi:10.1016/j.aap.2019.06.010.

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

Nie B, Li Q, Gan S, Xing B, Huang Y, Li SE. 2021. Safety envelope of pedestrians upon motor vehicle conflicts identified via active avoidance behaviour. Sci Rep. 11(1):1–10. doi:10.1038/s41598-021-82331-z.

Paul M, Ghosh I. 2020. Post encroachment time threshold identification for right-turn related crashes at unsignalized intersections on intercity highways under mixed traffic. Int J Inj Contr Saf Promot. 27(2):121–135. doi:10.1080/17457300.2019.1669666.

Rasouli A, Tsotsos JK. 2020. Autonomous vehicles that interact with pedestrians: A survey of theory and practice. IEEE Trans Intell Transport Syst. 21(3):900–918. doi:10.1109/TITS.2019.2901817.

Selmoune A, Yun J, Seo M, Kwon H, Lee C, Lee J. 2023. Development of a residential road collision warning service based on risk assessment. J Adv Transp. 2023:1–16. doi:10.1155/2023/7496377.

Shahdah U, Saccomanno F, Persaud B. 2014. Integrated traffic conflict model for estimating crash modification factors. Accid Anal Prev. 71:228–235. doi:10.1016/j.aap.2014.05.019.

Van der Horst R, Kraay RJ. 1986. The Dutch Conflict Technique—DOCTOR. *ICTCT Workshop* Budapest.

WHO. 2018. Global status report on road safety 2018. World Health Organization [Internet]. [accessed 2022 Jul 22]. http://apps.who.int/bookorders.

Zhang S, Abdel-Aty M. 2022. Real-Time Pedestrian Conflict Prediction Model at the Signal Cycle Level Using Machine Learning Models. IEEE Open J Intell Transp Syst. 3:176–186. doi:10.1109/OJITS.2022.3155126.