International Journal of Safety and Security Engineering
Vol. 16, No. 9, September, 2026, pp. 2017-2034
Journal homepage: http://iieta.org/journals/ijsse

A Composite Hybrid Severity Index for Two- and Three-Wheeler Crash Assessment Using Multi-Source Artificial Intelligence

Ghaidaa Abdulwahab Majeed, Abeer Khudhur Jameel*

Highway and Transportation Engineering Department, College of Engineering, Mustansiriyah University, Baghdad 10001, Iraq

Corresponding Author Email: abeer_khudhur@uomustansiriyah.edu.iq

Copyright: ©2026 The authors. This article is published by IIETA and is licensed under the CC BY 4.0 license (http://creativecommons.org/licenses/by/4.0/).

https://doi.org/10.18280/ijsse.160908

Received: 11 July 2026
Revised: 21 August 2026
Accepted: 28 August 2026
Available online: 30 September 2026

Keywords:
AHP, Artificial Intelligence, crash severity assessment, composite Hybrid Severity Index, intelligent transportation systems, motorcycle safety, road safety, three-wheeler safety

ABSTRACT

Crash severity assessment for two- and three-wheelers commonly relies on individual sources of evidence, such as police reported crash records, behavioral surveys, or traffic surveillance videos. This study develops a composite Hybrid Severity Index (HSI) that integrates these complementary sources into a single study-area severity indicator. Three independently sampled datasets from Baghdad were analyzed: 800 police reported crashes, questionnaire responses from 400 drivers, and 102 real-world crash videos. Objective and behavioral severity models were developed using supervised machine learning and evaluated on independent held-out test sets, achieving accuracies of 73.75% and 60.00%, respectively. The motion-based Artificial Intelligence (AI) framework combined object detection, tracking, motion analysis, anomaly detection, and rule-based severity estimation, achieving 70.60% three-level classification agreement across the entire 102-video dataset. Component severity scores were aggregated and normalized to a common 0–1 scale, then integrated using Analytic Hierarchy Process (AHP) weights of 0.3848, 0.2422, and 0.3730 for the objective, behavioral, and motion-based AI components, respectively, with a Consistency Ratio (CR) of 0.0210. The resulting normalized component scores were 0.278, 0.255, and 0.526, yielding an integrated HSI of 0.3650, classified as moderate using predefined equal-width interpretive bands. Sensitivity analysis showed that the classification remained Moderate under ±20% perturbations to the component weights, while ablation analysis demonstrated that the motion component had the greatest influence in the present case study. The proposed framework provides a transparent approach for synthesizing historical, behavioral, and dynamic traffic evidence into an interpretable area-level severity indicator for road safety decision support.

1. INTRODUCTION

Road traffic accidents are a serious public health and transport problem and one of the top causes of death and disability globally. The World Health Organization [1] estimated that, on average, 1.19 million people die annually in road traffic collisions and millions more are injured. A significant number of these deaths are attributable to vulnerable road users; among them are motorcyclists and three-wheel vehicle users, as the structure of a passenger vehicle does not protect them.

In many low- and middle- income countries, two- and three-wheelers are a central part of life. They are extremely important for transport due to their affordability, flexibility of use, and ability to operate in congested urban settings [2-4]. Consequently, there has been an increase in crash rates and more severe injuries. Recent studies have found that passengers of powered two- and three-wheeled motor vehicles are at 28–34 times higher risk of death than occupants of passenger cars [3, 5, 6], underscoring the need for accurate measures of crash severity for effective proactive road safety.

The traditional approach to crash severity analysis has mainly relied on official crash databases and statistical/machine-learning models using police-reported crash data. In the last few years, ensemble learning algorithms have demonstrated predictive power by capturing complex nonlinear interactions among roadway, vehicle, environmental, and crash features [7-10]. Although there have been several such advances, official crash databases typically provide minimal detail on driver behavior, risk perception, and pre-crash interactions that significantly impact crash outcomes.

Recognizing the need to fill this gap, transportation safety research has increasingly included behavioral data from survey questionnaires. The Safety Index (SI) and Risk Perception Index (RPI) are composite behavioral indicators used extensively to quantify drivers' safety awareness, compliance with traffic laws, and perception of traffic hazards [11, 12]. These behavioral measures offer important clues about unobserved human factors not captured in crash reports. Behavioral studies, however, are seldom conducted in tandem with objective crash studies and are not as definitive as when added to objective crash analysis.

Meanwhile, the rapid development of Artificial Intelligence (AI), computer vision, and intelligent transportation systems

(ITS) has paved the way for automated crash detection and severity assessment from traffic surveillance videos. Vehicle trajectory, traffic interactions, and abnormal driving behavior can be captured in real time before and during collision events using machine vision and deep learning techniques, and multi-object tracking and motion analysis can provide dynamic data not available in conventional crash databases. However, those previous AI studies have only been applied to video analysis and have provided no link to historical crash data or driver personality.

Therefore, existing crash severity assessment methods are largely single-source, relying on either official crash data, behavior surveys, or surveillance video. While each data source is useful, none alone captures the multitude of factors that combine to determine the severity of a crash—roadway, human, and real-time factors. In addition, studies available so far have not provided a mathematically meaningful framework to combine diverse safety data into a single area-level severity score suitable for engineering decisions.

This study aims to fill this research gap by proposing a composite Hybrid Severity Index (HSI) that combines three complementary sources of motorcycle and three-wheeler crash severity data: the questionnaire-based motorcycle and three-wheeler SI and RPI, and real-world traffic surveillance videos processed by the motion-based AI model. All three predictive models are scaled to numerical values and then combined using weights determined by the experts' pairwise comparisons with the Analytic Hierarchy Process (AHP). This HSI yields an interpretable, continuous measure of crash seriousness, which incorporates historical crash attributes, driver actions, and real-time traffic conditions.

The objectives of this study are to: (1) develop three complementary crash severity prediction models at the study area level based on objective crash records, driver behavioral characteristics, and motion-based AI using traffic surveillance videos; (2) integrate the normalized outputs of these models into an HSI using the AHP; (3) validate the proposed framework through classification performance metrics, confusion matrix analysis, and sensitivity analysis to demonstrate its applicability for ITS and road safety management.

Unlike conventional crash severity prediction or ensemble learning approaches that combine multiple algorithms trained on the same dataset, the proposed framework integrates independent data sources within a transparent mathematical model. This enables the contribution of each information source to be explicitly quantified, improving the interpretability and reliability of severity assessment.

The main contributions of this study are:

•Development of a novel HSI integrating objective crash data, driver behavioral characteristics, and AI-based motion analysis.

•Application of the AHP to derive consistent and interpretable weights for heterogeneous safety information.

•Formulation of a transparent mathematical framework that converts multiple independent severity assessments into a single continuous severity measure.

•Demonstration of the framework's effectiveness for motorcycles and three-wheelers, with potential extension to other vulnerable road users and ITS.

2. METHOD OF THE RESEARCH

2.1 Overall research framework

The proposed framework is shown in Figure 1, which integrates three independent and complementary sources of information to characterize different dimensions of two- and three-wheeler safety: official crash records, driver behavioral questionnaires, and traffic-surveillance crash videos. The three datasets were obtained independently from different sources and were not matched at the individual driver or crash-event level. They should therefore be regarded as complementary rather than observation-matched sources of safety evidence. They also covered different data-collection periods.

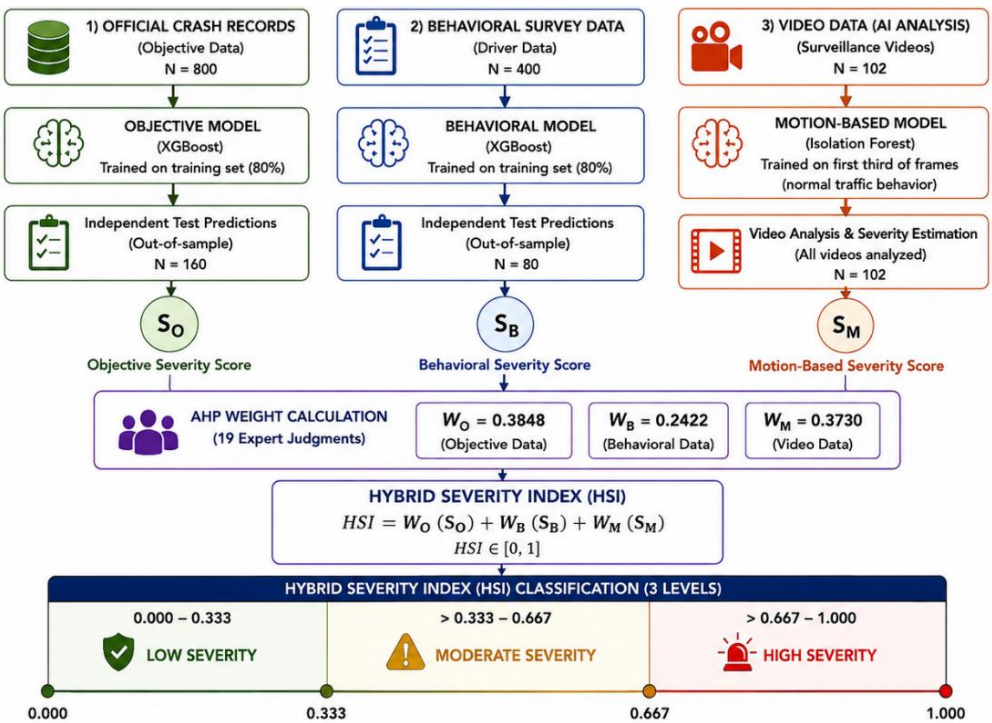


Figure 1. The proposed framework of the research

However, all three datasets shared the same geographical study context, namely the Rusafa area of Baghdad. Therefore, the proposed integration does not assume correspondence between individual observations across the three datasets. Instead, each dataset was analyzed independently through its corresponding sub-model, and integration was subsequently performed at the aggregate study-area level after normalization of the three severity scores. Consequently, their integration was performed at the aggregate study-area level rather than at the individual-record level.

Each data source is processed independently to characterize a distinct safety dimension. The objective model represents the historical crash-severity conditions obtained from official police records; the behavioral model represents driver-related safety behavior and risk characteristics obtained from the questionnaire survey; and the motion-based AI model represents dynamic traffic and vehicle-interaction characteristics extracted from crash videos. Each sub-model therefore produces a separate aggregate component score: the Objective Severity Score (SO), Behavioral Score (SB), and Motion-Based Severity Score (SM).

Integration is performed only after each data source has been independently processed through its corresponding sub-model. For the supervised objective and behavioral sub-models, the component severity scores used for hybrid integration are derived exclusively from predictions generated for the independent test sets, thereby preventing training observations from contributing to the final hybrid score. The motion-based component is derived from the independently analyzed crash-video dataset according to its unsupervised video-analysis protocol. Because the three component scores originate from datasets with different sample sizes, measurement units, and data-generating processes, they are transformed to a common dimensionless scale before integration. The AHP is subsequently used to quantify the relative contribution of the three information sources through expert pairwise comparisons. The normalized component scores are then combined using the AHP-derived weights to construct the HSI.

Accordingly, the HSI should be interpreted as an aggregate multi-source decision-support indicator that synthesizes complementary historical, behavioral, and dynamic evidence. It is not intended to represent the predicted severity of an individual crash or to imply individual-level correspondence among the three independently sampled datasets.

The proposed framework can be represented mathematically by considering the outputs generated from the three predictive sub-models: SO, SB, and SM. Their relative importance is quantified using the AHP-derived weighting coefficients: Weight of objective score (WO), Weight of behavioral score (WB), and Weight of Motion score (WM), subject to the normalization constraint:

$$WO + WB + WM = 1 \quad (1)$$

The HSI is then computed as:

$$HSI = WO(SO) + WB(SB) + WM(SM) \quad (2)$$

where, *HSI* is a continuous severity measure bounded between 0 and 1; larger values indicate greater crash severity and a higher priority for safety intervention. The computed *HSI* is subsequently transformed into discrete severity categories using predefined decision thresholds, enabling practical application within ITS and road safety decision-support platforms.

2.2 Development of the objective crash severity model

2.2.1 Data processing and feature engineering

The first dataset consisted of 800 officially recorded motorcycle and three-wheeler crashes obtained from the General Traffic Directorate in Baghdad, Iraq. The crash records covered six years and included only motorcycle and tuk-tuk crashes. The outcome variable was police-recorded crash severity, classified according to the official crash records as fatal, injury, or property-damage-only (PDO). This dataset was used to characterize the historical and engineering dimensions of crash severity.

To perform model development, the 800 officially recorded crash data were first preprocessed through steps such as duplicate data removal, consistency checks, missing-value checks, encoding categorical variables, and data normalization. Since several severity classes were imbalanced, the Synthetic Minority Over-sampling Technique (SMOTE) was applied exclusively to the training set after the train–test split. The independent test set was kept unchanged and was not involved in either oversampling or model training, thereby providing an out-of-sample evaluation dataset. The dataset was stratified into 80% training (n = 640) and 20% independent testing (n = 160) subsets to preserve the original class distribution.

To enhance the predictive performance of the machine-learning models and reduce redundancy, feature engineering was performed [13-15]. The first feature set comprised variables corresponding to road environment, roadways, vehicle, collision mechanism, temporal, and crash-related attributes, obtained from official accident reports. In this research, new features were produced by spatial analysis. The DBSCAN algorithm was then applied to generate hotspot indicators, including Zone_Label, Zone_Count, and Hotspot_Flag.

Relationships and distributions between the variables were explored by using exploratory data analysis (EDA). The analysis gave insights into the importance of the variables, data imbalance, and potential noise. These results were then used to select variables for exclusion from the model because they were poorly represented, poorly influential, or redundant. The features were subsequently ranked in a feature-importance analysis during model development to select the variables that most strongly influenced the model's crash-severity estimation. These analyses helped explain the model's performance and identify the engineering factors most similar to those in severe motorcycle and three-wheeler crashes. The wrong parking and illness features, which were within the crash causes variables, were excluded due to very low frequency. Weather state was excluded because it showed no significant patterns in crash frequency, and road condition was excluded due to limited variability in the data. The final predictor variables included roadway characteristics, environmental conditions, vehicle information, collision characteristics, temporal variables, and DBSCAN-derived hotspot indicators.

2.2.2 Model development, validation, and area-level severity score

Three machine-learning algorithms for ensemble learning were tested for objective crash-severity prediction: Random Forest (RF), CatBoost, and Extreme Gradient Boosting (XGBoost) [16-19]. The choice of these algorithms is

motivated by their strong performance with nonlinear relationships, mixed-type predictor variables, and complex interactions, which are frequently observed in traffic safety research. A comparison was made between the algorithms because the same predictor variables were used for training. The hyperparameters were systematically tuned to maximize the classification performance and avoid overfitting.

The processed dataset was first divided using a stratified 80/20 split into a training set (n = 640) and an independent test set (n = 160). Fivefold cross-validation was then performed exclusively on the training data to develop the model and assess robustness. In each fold, four subsets were used for training, and the remaining subset was used for validation, with each subset serving once as the validation fold. The independent test set remained completely unseen during model development and was used only for final out-of-sample evaluation.

The selected algorithms were assessed for predictive performance using classification metrics such as overall accuracy, precision, recall, and F1-score, computed from the confusion matrix [17, 20-23]. These complementary measures provide a thorough evaluation of classification performance, especially when the data is imbalanced with respect to crash severity, and considering only overall accuracy can be misleading.

The independent test set was used exclusively for the final evaluation of the objective severity model's predictive performance. To maintain strict separation between model development and hybrid integration, the SO used in the HSI was also derived exclusively from the out-of-sample predictions generated for this independent test set (n = 160). Training observations were therefore not included in the severity counts used to construct SO.

The proposed HSI is designed as an aggregate multi-source severity indicator rather than as a prediction for an individual crash. For the objective component, the out-of-sample predictions on the independent test set were aggregated by severity class. The SO was calculated by multiplying the number of test observations predicted in each severity category by its corresponding severity weight and summing the weighted values across the three classes. For the objective model, the score was formulated as:

$$SOscore = W_F(f_F) + W_I(f_I) + W_{PDO}(f_{PDO}) \quad (3)$$

where, f_F, f_I, and f_{PDO} denote the numbers of observations in the independent test set predicted by the objective sub-model as Fatal, Injury, and PDO, respectively. The *W* terms represent the corresponding severity weights. SO is the resulting objective severity component score used for hybrid integration

2.3 Developing the behavioral severity model

2.3.1 Questionnaire, validation, and index construction

The second dataset consisted of a questionnaire survey of 400 motorcycle and three-wheeler drivers in the same study area in Baghdad City. The survey collected information on demographic and operational characteristics, driving exposure, safety behavior, crash history, and traffic-risk perception. This dataset was used to characterize the behavioral dimension of the proposed framework rather than to represent the same individual crashes contained in the police database.

The developed questionnaire was divided into four parts. The first part collected demographic and operational data: age, level of education, driving experience, type of vehicle, monthly income, frequency of trips, and daily driving hours. The second part assessed safety-related behavior using a series of Likert-type statements on helmet use, speeding, adherence to traffic rules, driver distraction, overtaking, and protective equipment use. The third part evaluated drivers' perceptions of hazardous traffic situations, including attitudes towards speeding, unsafe overtaking, traffic violations, roadway conditions, and interactions with other road users. The last section collected data on past and self-reported crashes. A three-point Likert-type scale (0–2) was used to code the safety- and risk-related questionnaire items. Item coding was oriented consistently before index construction so that the resulting scores reflected the intended direction of the corresponding safety or risk construct.

Prior to model development, the questionnaire responses were screened for completeness, consistency, and validity. Incomplete questionnaires and inconsistent responses were excluded from further analysis. Numerical encoding was subsequently applied to categorical variables to facilitate machine-learning modeling. Internal consistency of the questionnaire was evaluated using Cronbach's alpha, and construct validity was assessed via exploratory factor analysis prior to calculating the composite indices. Consequently, the number of questionnaire items included in the final stage of predictive modeling was reduced due to low response variability, weak explanatory power, redundancy with other items, or insufficient representation in the data. This reduced noise, enhanced data consistency, minimized dimensionality, and improved the ability of the developed machine-learning models to generalize.

Exploratory Factor Analysis (EFA) was conducted to assess the construct validity and underlying structure of the safety and risk-perception items. The suitability of the data for factor analysis was first examined using the Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity. A KMO value greater than 0.60 together with a statistically significant Bartlett's test ($p < 0.05$) was considered evidence of adequate factorability. Factor extraction was followed by Varimax rotation to obtain a clearer and more interpretable factor structure. Items with factor loadings greater than 0.40 were retained. The analysis identified two factors for the Safety construct and three factors for the Risk Perception construct, supporting the multidimensional structure of the questionnaire measures.

The behavioral model was developed by constructing the Self-reported Safety Index (Self-SI) and RPI. The Self-SI is a composite score representing the respondent's self-reported safety-related driving behavior. It was calculated by averaging the coded safety-related questionnaire items after ensuring that the items were consistently oriented, as expressed in Eq. (4).

$$SI = \frac{1}{n} \sum_{j=1}^{n} XSI_{ij} \quad (4)$$

where, SI_i: index value for participant *i*, XSI_{ij}: response of participant *i* to question *j*, and *n*: total number of questions included in the index calculation.

Likewise, the RPI is a composite score representing the respondent's perception of hazardous traffic situations. It was calculated by averaging the coded risk-perception items after ensuring consistent item direction, as expressed in Eq. (5).

$$RP = \frac{1}{n} \sum_{j=1}^{n} XRP_{ij} \quad (5)$$

where, RP_i: index value for participant *i*, XRP_{ij}: response of participant *i* to question *j*, and *n*: total number of questions included in the index calculation.

The resulting SI and RPI scores were subsequently used as quantitative behavioral predictors in the separate behavioral crash-severity model. They were not used as predictors in models developed to classify the indices themselves.

2.3.2 Model development and behavioral severity score

The processed behavioral dataset (n = 400) was split 80:20 into a training set (n = 320) and an independent held-out test set (n = 80). Behavioral data were converted to a feature matrix in a structured format appropriate for supervised machine learning. The set of final predictors included three groups of variables:

•Demographic characteristics (driving experience);

•Operational characteristics (vehicle type, number of trips per day, number of driving hours); and

•Behavioral indicators: the SI, RPI, and past crash involvement.

Redundant predictors were identified, and potential multicollinearity among the candidate variables was assessed using correlation analysis. Complementary information variables were included in the final dataset, enhancing the model's stability and interpretability.

Machine learning techniques were used to develop the behavioral crash severity model. The processed dataset was divided into separate training (n = 320) and independent testing sets (n = 80). Fivefold cross-validation was performed on the training data during model development to assess stability and inform model selection, while the independent test set remained unseen during training and selection. Several classification algorithms were evaluated under the same predictor structure. Final predictive performance was reported exclusively on the independent test set using accuracy, precision, recall, and F1-score derived from the confusion matrix. The best-performing model was subsequently selected as the behavioral severity sub-model.

The behavioral sub-model produces four outcome categories: No Accident, Low, Moderate, and High. The No Accident category was excluded from the area-level behavioral-severity aggregation because it represents the absence of an experienced crash rather than an ordered behavioral-severity level. The remaining Low, Moderate, and High categories were treated as ordered levels of the behavioral component.

To construct the SB, an Equivalent Property Damage Only (EPDO)-derived ordinal weighting scheme was adopted. Importantly, these weights are not interpreted as establishing equivalence between behavioral-severity categories and observed crash outcomes. Instead, the EPDO-derived coefficients are used only as monotonic numerical weights that assign progressively greater influence to progressively higher behavioral-severity levels during aggregation. The behavioral component score is given as:

$$SBscore = W_{HR}(f_{HR}) + W_{MR}(f_{MR}) + W_{LR}(f_{LR}) \quad (6)$$

where, *W* terms refer to the corresponding severity weights. *SB* is the predicted SB. f_{HR}, f_{MR}, and f_{LR} represent the number of observations in the independent test set predicted by the behavioral sub-model as high, moderate, and low severity, respectively. This classification aligns with previous research [24]. The "No Accident" category was excluded from the severity aggregation because it represents the absence of an experienced crash rather than a crash-severity level. The low, moderate, and high categories were treated as ordered behavioral severity levels and subsequently used to construct the test-set-based behavioral component score for normalization and hybrid integration.

The resulting area-level SB was subsequently normalized to a common numerical scale and integrated with the independently derived objective and SMs at the study-area level to construct the HSI.

2.4 Development of a motion-based AI model

The third part of the proposed HSI is the motion-based AI model. The third dataset consisted of 102 real-world motorcycle and three-wheeler crash videos, collected from the same defined study area in Baghdad City, and analyzed using computer vision and motion-based techniques. These observations provided dynamic information on vehicle trajectories, motion characteristics, and traffic interactions that is not available from either the police crash records or the questionnaire survey.

The motion-based framework estimates crash severity from traffic surveillance videos by analyzing the temporal evolution of vehicle movements before and during detected crash events.

The proposed AI solution combines computer vision, multi-object tracking, motion analysis, and anomaly detection to automatically estimate the severity of detected traffic crashes based on real-world traffic videos. The entire modeling pipeline can be decomposed into five successive steps: (i) video pre-processing, (ii) detection of objects (people or vehicles), (iii) multi-object tracking, (iv) motion feature extraction, and (v) estimating the severity. The last prediction generated by the framework is the SM.

You Only Look Once version 8 (YOLOv8) was used to detect vehicles. YOLOv8 is a one-stage deep-learning detector [25-27]. It can detect and classify objects in a single forward pass, achieving high accuracy while processing in real time. The detector identifies motorcycles, passenger cars, heavy vehicles, pedestrians, bicycles, and other road users by creating a bounding box and class label for every video frame. In the second stage of tracking, the detected objects form the input.

While object detection can detect traffic participants separately in each frame, it is important to maintain each vehicle's identity throughout the video sequence for continuous crash analysis. For this reason, the objects detected were tracked using the Simple Online and Real-time Tracking (SORT) algorithm [26, 28]. To estimate the object trajectories across successive frames while minimizing computation, SORT fuses the Kalman filter with the Hungarian assignment algorithm. Continuous tracking allows the determination of the vehicle's image-plane speed, acceleration, path, and other driver interactions. The instantaneous image-plane speed of each tracked vehicle was estimated from the displacement of its bounding-box centroid between two consecutive frames as follows:

$$Speed = \frac{\sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}}{\Delta t} \quad (7)$$

where, x_t, y_t = centroid coordinates of the tracked vehicle in the current frame. x_{t-1}, y_{t-1} = the corresponding centroid coordinates in the previous frame. Δt = 1/FPS is the time interval between consecutive frames. Speed = estimated object

speed in pixels per second (px/s). Image-plane speed was used because the surveillance videos did not provide the consistent camera calibration parameters required to convert pixel displacement into real-world metric speed.

The trajectories of the tracked vehicles were then analyzed to examine dynamic motion properties associated with crash severity. Dense Optical Flow [29-31] was used for motion analysis, calculating pixel motion between video frames to quantify the magnitude and direction of vehicle movement.

$$M = \sqrt{u^2 + v^2} \tag{8}$$

where, u and v are the horizontal and vertical optical-flow components, respectively, and M is the resulting motion magnitude.

A collection of motion descriptors was calculated for each tracked vehicle, such as: instantaneous image-plane speed, acceleration, trajectory deviation, motion magnitude, directional variation, and vehicle interaction characteristics. These features describe unusual traffic activity that occurs just prior to collision events and provide a dynamic view of the development of a crash that is not available in traditional crash databases.

An Isolation Forest algorithm was applied to the extracted motion features to determine if there were any irregular vehicle behaviors related to collisions. Isolation Forest (IF) is an unsupervised anomaly detection algorithm that isolates observations that need to be detected as anomalies with fewer recursive partitions than normal traffic movements. Therefore, an abnormal vehicle interaction involved in a crash will receive a higher anomaly score.

$$f_i = [f_1, f_2, \ldots, f_{14}] \tag{9}$$

where, f_i is the 14-dimensional feature vector extracted from frame i,

$$A_i = IF\ (f_i) \tag{10}$$

where, A_i = anomaly score of frame I, and IF denotes the fitted Isolation Forest anomaly-detection model.

The 14-dimensional motion-feature vector was used by Isolation Forest for anomaly detection and was not directly combined through a weighted linear formulation to calculate SS.

Following anomaly detection, the severity of each detected crash event was estimated from the motion characteristics of the tracked vehicles. Vehicle image-plane speed was calculated from the Euclidean displacement of the vehicle centroid between successive frames multiplied by the video frame rate, and was therefore expressed in pixels per second (px/s). For each detected event, the maximum vehicle image-plane speed (V_{max}) and mean vehicle image-plane speed ($\overline{V}$) were calculated. The event severity score (SS) was then determined separately using the rule-based severity-estimation procedure described in Eqs. (11).

$$SS = \begin{cases} \min\left(10, \left[\frac{V_{max}}{10}\right]\right), V_{max} > 1.5\,\overline{V} \\ \min\left(10, \left[\frac{\overline{V}}{10}\right]\right) \quad Otherwise \end{cases} \tag{11}$$

where, V_{max} is the maximum tracked-vehicle image-plane speed (px/s) and $\overline{V}$ is the mean tracked-vehicle image-plane speed (px/s) during the detected event. The factors 1.5 and 10 are fixed rule-based parameters used in the implemented severity-estimation algorithm.

Because the floor operation in Eqs. (11) produces integer values, SS is a discrete integer-valued severity score rather than a continuous score; therefore, no intermediate values occur between 3 and 4 or between 6 and 7.

Low: $0 \leq SS \leq 3$

Moderate: $4 \leq SS \leq 6$

High: $7 \leq SS \leq 10$

This three-level representation was adopted to provide a consistent and interpretable basis for comparison with the official crash-severity records and is consistent with the use of multi-level severity categorization in previous crash-severity studies [32, 33].

The ground-truth severity labels for the crash videos were obtained from the official records accompanying the surveillance videos provided by the Ministry of Interior. The documented crash outcomes were represented on the same three-level ordinal scale for validation: PDO as Low, Injury as Moderate, and Fatal as High. Thus, the Low, Moderate, and High classes were not direct outputs of the Isolation Forest algorithm. Isolation Forest was used for anomaly detection within the motion-analysis pipeline, whereas the subsequent severity-estimation stage generated a 0–10 SS and mapped it to the three severity levels. The predicted severity level of each video was then compared with its corresponding official ground-truth level using the severity-matching procedure and confusion-matrix-based evaluation.

No independent held-out test subset was used for the three-level severity evaluation of the motion-based AI component. Isolation Forest was applied as an unsupervised anomaly-detection method to the extracted motion features, while the subsequent rule-based severity-estimation procedure generated the Low, Moderate, and High severity classes. The performance of this severity-classification procedure was evaluated across the complete available dataset of 102 crash videos by comparing each video's predicted severity level with its corresponding official crash-severity label. Accordingly, the reported performance metrics for this component represent classification agreement across the complete 102-video dataset and should not be interpreted as independent held-out test performance.

The severity estimation model is written as

$$SMscore = W_{HR}(f_{HR}) + W_{MR}(f_{MR}) + W_{LR}\ (f_{LR}) \tag{12}$$

where, f_{HR}, f_{MR}, and f_{LR} denote the predicted frequencies of high-, moderate-, and low-severity cases, respectively. W_{HR}, W_{MR}, and W_{LR} are their corresponding severity weights. SM represents the resulting area-level motion severity score.

2.5 Output normalization and weight determination

The severity component scores obtained from the three sub-models were transformed to a common dimensionless scale before hybrid integration. For the objective and behavioral sub-models, the component scores were derived from the independent test-set predictions, whereas the motion-based component was derived from the full set of 102 evaluated crash videos. Because the three sub-models differ in sample size and output magnitude, normalization was applied before weighting and integration. The normalized severity score was computed as

$$Si^* = \frac{Si - Smin}{Smax - Smin} \quad (13)$$

where, Si^* is the normalized area-level severity score; *Si* is the original predicted severity score; *Smin* and *Smax* denote the theoretical minimum and maximum severity scores defined from the corresponding component's weighting structure and number of observations included in aggregation. The normalization procedure transforms all model outputs into a common interval between 0 and 1 while preserving their relative ranking.

For each component, the normalization bounds were defined theoretically from the corresponding severity-weighting structure. The minimum score was set to zero, representing the absence of weighted severity burden, whereas the maximum score represented the hypothetical case in which all observations included in that component were assigned to the highest severity category. For the objective component, SOmax = 542 × 160 = 86,720. For the behavioral component, after excluding the No-Accident cases, 56 observations remained, yielding SBmax = 542 × 56 = 30,352. For the Motion component, all 102 videos were included, yielding SMmax = 542 × 102 = 55,284.

The relative contribution of the three predictive models was determined using the AHP, a structured multi-criteria decision-making technique proposed by Saaty (1980) [34]. AHP was selected because it enables the systematic incorporation of expert judgment while providing a quantitative mechanism for evaluating the consistency of pairwise comparisons. In the proposed framework, AHP was employed to determine the weighting coefficients assigned to the three sub-models before their integration into the HSI.

Experts conducted pairwise comparisons using the Saaty 9-point scale, with 1 signifying equal importance and 9 signifying extreme importance, as shown in Table 1.

Table 1. Saaty pairwise comparison scale

Scale	Interpretation
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extreme importance
2,4,6,8	Intermediate values

The first step involved constructing a pairwise comparison matrix using Saaty's nine-point scale, in which experts compared the relative importance of each predictive model for crash severity assessment.

Because multiple transportation safety experts participated in the evaluation, their individual judgments were aggregated using the geometric mean (GM) in Eq. (14). This approach minimizes the influence of extreme judgments while preserving the reciprocal structure of the pairwise comparison matrix.

$$GM = (\prod_{i=1}^{n} Xi)^{1/n} \quad (14)$$

where, *GM* is the geometric mean value, *Xi* is the comparison value provided by expert *i*, and n is the total number of experts. *GM* values obtained from the experts were used to construct the aggregated pairwise comparison matrix (A). The reciprocal values were then computed according to Eq. (15).

$$a_{ji} = \frac{1}{a_{ij}} \quad (15)$$

where, a_{ij} is the relative importance of criterion *i* compared with criterion *j*, while a_{ij} is the reciprocal comparison of criterion j compared with *i*. Thereby satisfying the reciprocal property required by the AHP methodology.

The aggregated values form the final pairwise comparison matrix used in subsequent calculations. To obtain comparable values, each element of the pairwise comparison matrix was normalized by dividing it by the sum of its corresponding column. The normalized matrix transforms all comparison values into dimensionless quantities whose column sums equal one. The weighting coefficient for each predictive model was obtained by averaging the normalized values within each row; these coefficients quantify each model's relative contribution to the HSI.

To verify the reliability of the expert judgments, the maximum eigenvalue of the comparison matrix was first estimated. Eq. (16) was applied to find the sum of column *j*.

$$C_j = \sum_{i=1}^{n} a_{ij} \quad (16)$$

where, C_j = sum of column *j*, a_{ij} = comparison value in row *i* and column *j*, and *n* = number of criteria.

The next step in the AHP was to determine the relative importance of each criterion. For the present three-criterion AHP matrix, the priority weights were estimated using the normalized column-row average method. Each element of the pairwise comparison matrix was first divided by the sum of its corresponding column, and the priority weight of each criterion was then obtained by averaging the normalized values across its row, as expressed in Eq. (17).

$$w_i = \frac{1}{n}\sum_{j=1}^{n} n_{ij} \quad (17)$$

where, w_i is the priority weight of criterion *i*, n_{ij} is the normalized matrix element in row *i* and column *j*, and *n* is the number of criteria.

Because pairwise comparisons rely on subjective evaluation, inconsistencies may arise when experts inadvertently express contradictory preferences. To evaluate the reliability of the comparison matrix, the AHP employs the Consistency Ratio (CR). The CR measures the degree to which the pairwise comparisons satisfy the principle of logical transitivity. The consistency evaluation involves four sequential calculations: the weighted sum vector, the principal eigenvalue, the Consistency Index (CI), and the CR. The weighted sum vector is obtained by multiplying the aggregated comparison matrix by the priority vector, as shown in Eq. (18).

$$AW = A \times W \quad (18)$$

where, *A* = aggregated pairwise comparison matrix, *W* = priority (eigenvector) weights, and *AW* = weighted sum vector.

Each element of the weighted sum vector was divided by its corresponding priority weight to obtain an individual eigenvalue estimate using Eq. (19).

$$\lambda_i = \frac{(AW)_i}{w_i} \quad (19)$$

where, λ_i = eigenvalue estimate for criterion i; $(AW)_i$ = ith element of the weighted sum vector; w_i = priority weight of criterion i.

The λ_{max} is equal to the number of criteria and is compared with the mean of the three eigenvalues. The difference between the two values is used to estimate the CI, which measures the deviation of the comparison matrix from perfect consistency, as given by Eq. (20).

$$CI = \frac{\lambda_{max} - n}{n-1} \quad (20)$$

This value is used to estimate the CR, which compares the obtained CI with the Random Index (RI) for the size of the comparison matrix, as given by Eq. (21).

$$CR = \frac{CI}{RI} \quad (21)$$

According to Saaty's criterion, expert judgments are considered sufficiently consistent when CR < 0.10. If the calculated CR exceeds this threshold, the pairwise comparisons should be re-evaluated until an acceptable level of consistency is achieved.

After confirming the consistency of the comparison matrix, the resulting priority weights were adopted as the weighting coefficients of the three predictive models. These coefficients were subsequently combined with the normalized severity scores to compute the proposed HSI.

Following the computation of the HSI, the continuous normalized HSI values, ranging from 0 to 1, were classified into three ordered severity categories to facilitate practical interpretation and decision-making.

The three-level classification structure (Low, Moderate, and High) is consistent with multi-level severity categorization adopted in previous road-safety studies [35, 36]. However, the specific boundaries used in the present study, shown in Table 2, were predefined as equal-width interpretive intervals on the normalized 0–1 HSI scale and were not derived from or validated by those previous studies.

Higher HSI values indicate greater overall crash severity. The continuous HSI value was retained for quantitative analysis, while the three severity categories were used only for interpretation and reporting. This three-level structure also aligns with the Low–Moderate–High severity representation used across the component models.

Table 2. Classification of the composite Hybrid Severity Index (HSI)

HSI Range	Severity Level
0.000–0.333	Low Severity
>0.333–0.667	Moderate Severity
>0.667–1.000	High Severity

3. RESULTS AND DISCUSSION

3.1 Performance of the predictive models

The predictive capability of the three sub-models was assessed individually using standard classification metrics, including accuracy, precision, recall, and F1-score, together with confusion matrix analysis. These metrics provide complementary information regarding the classification performance of the developed models and their suitability for subsequent integration. A comparative summary of the individual model performance is presented in Table 3.

The three sub-models demonstrated different levels of classification performance, with the objective model achieving the highest overall accuracy.

Table 3. The individual models' performance

Model	Best Algorithm	Performance Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Objective Crash Severity	XGBoost	73.75	74	74	74
Behavioral Severity	XGBoost	60	60	60	60
Motion-Based AI	Isolation Forest	70.60	79.7	74.4	69.5

Note: The objective and behavioral models were evaluated using their respective independent held-out test datasets. For the motion-based AI component, the reported metrics were calculated by comparing the predicted three-level severity classifications with the corresponding official severity labels across the entire available dataset of 102 crash videos; therefore, the 70.6% metric does not represent independent held-out test performance.

3.1.1 Performance of objective sub-model

The best overall classification result for the evaluated machine-learning algorithms was obtained by the XGBoost algorithm. This aligns with previous research, as XGBoost can model nonlinear interactions among roadway characteristics, environmental conditions, vehicle attributes, and crash-related variables.

The confusion matrix for the test dataset, shown in Figure 2, provides insight into the model's prediction distribution. Fatal and injury crashes were correctly classified in most cases, and some PDO crashes were misclassified as injury crashes. The confusion matrix shows that 33 crashes were correctly classified as PDO, and 14 PDO crashes were misclassified as injury crashes. In addition, about 47 crashes were correctly classified as injury crashes, while 15 crashes were misclassified as PDO crashes. Regarding fatal crashes, 38 out of 46 were correctly classified as such, indicating that severe crashes exhibit more distinct features.

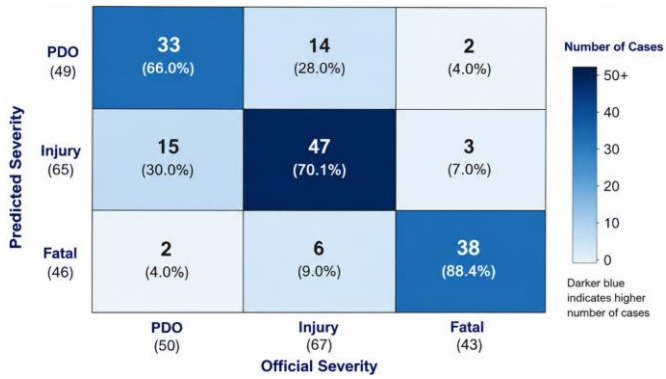


Figure 2. Confusion matrix and predicted severity distribution for the independent test dataset (n = 160)

This analysis showed that most misclassifications occurred between the PDO and Injury categories. This misclassification may have resulted from overlapping aspects of lower-severity

crashes, making it difficult to identify the factors that differentiate them. Fatal collisions, in contrast, were more accurately predicted because stronger patterns were associated with fatal outcomes. The confusion aligns with findings from previous studies in Pakistan [21] and China [35-37].

The predictive performance of the objective severity model was evaluated exclusively using the independent test dataset comprising 160 crashes (20% of the complete dataset). Of these observations, 118 were correctly classified (33 PDO, 47 Injury, and 38 Fatal), resulting in an independent test accuracy of 73.75%.

For out-of-sample area-level aggregation, only predictions obtained from the independent held-out test dataset (n = 160) were used. The test-set predictions comprised 50 PDO, 67 Injury, and 43 Fatal crashes. These predicted class frequencies were used exclusively to calculate the SO, ensuring that the objective component of the HSI was derived entirely from observations not used during model training.

The area-level severity score for the SO was derived from the model-predicted class distribution. The number of observations assigned to each severity category was multiplied by its corresponding severity weight, and the weighted values were summed to quantify the aggregate severity burden of the study area. The confusion matrix was reported separately to evaluate agreement between the observed and predicted classes.

To determine the severity score, the EPDO method, as adopted in the Highway Safety Manual (HSM) [38], was used. The EPDO methodology was adopted because it is a well-established approach in highway safety engineering for representing crashes of different severities on a common numerical scale. Rather than using EPDO to evaluate model performance, this study applies EPDO as a severity-weighting framework to convert the predicted distribution of crash severity classes into an engineering severity score. This transformation enables outputs from heterogeneous predictive models to be represented on a comparable numerical scale prior to normalization and weighted integration within the HSI.

For the independent test set, the predicted distribution comprised 50 PDO, 67 Injury, and 43 Fatal crashes. Substitution of these frequencies into Eq. (3) yielded the following SO:

$$SO\ Score = 542(43) + 11(67) + (50) = 24{,}093$$

The predictive performance of the objective severity model was evaluated exclusively using the independent held-out test dataset (n = 160), as presented in Figure 2. The same out-of-sample predictions were subsequently used to construct the SO. The predicted test-set distribution comprised 50 PDO, 67 Injury, and 43 Fatal crashes. Thus, both model performance evaluation and the severity distribution used for the hybrid integration were based exclusively on observations not used during model training.

To normalize the SO, Eq. (13) was applied using the theoretical bounds of the EPDO-weighted score for the independent test set. The theoretical minimum score was set to zero, while the maximum score corresponded to the hypothetical case in which all 160 test crashes were assigned to the Fatal class. Accordingly, S_{max} = 542 × 160 = 86,720.

$$SO* = \frac{24{,}093 - 0}{86{,}720 - 0} = 0.278$$

3.1.2 Performance of perception/behavioral sub-model

The behavioral severity sub-model was developed based on subjective attributes related to crash severity and used the following factors: demographic characteristics, driving exposure, SI, and RPI. After comparing the models, XGBoost was chosen to create a SB.

As shown in Figure 3, the confusion matrix indicates that the behavior model correctly classified the different levels of safety behaviors and perceived crash risk among drivers. A higher level of overlap between adjacent severity classes was observed compared with the objective model. However, this is not surprising, since behavioral characteristics do not describe the actual characteristics of crash events; rather, they reflect the latent factors that directly affect crash severity.

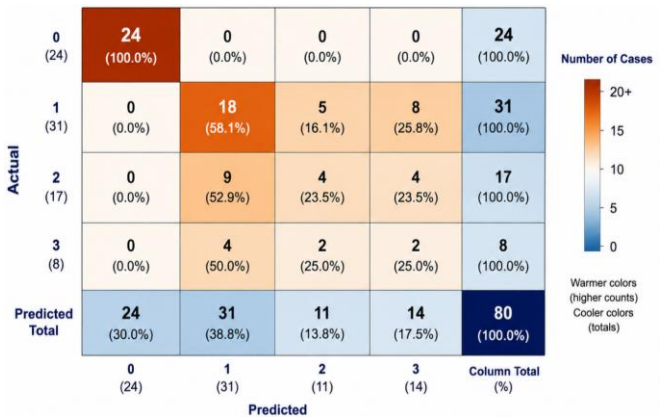


Figure 3. Confusion matrix for the independent test set (n = 80)

Note: 0 = No Accident, 1 = Low Severity, 2 = Moderate Severity, and 3 = High Severity.

The behavioral sub-model's overall predictive performance was lower than that of the objective crash model, but it can provide information unavailable in official databases. Thus, the SB can be used to supplement the engineering data obtained from police-reported crashes and enhance the comprehensive multidimensional characterization of crash severity in the proposed HSI.

For out-of-sample hybrid integration, the behavioral component score was calculated exclusively from predictions obtained from the independent held-out test dataset (n = 80). After excluding the No-Accident category from severity scoring, the predicted distribution comprised 31 Low-, 11 Moderate-, and 14 High-severity observations. The EPDO-derived ordinal weighting scheme was applied to preserve the ordered severity emphasis of the three behavioral severity categories. These weights are used as severity-weighting coefficients and do not imply equivalence between the behavioral categories and actual PDO, Injury, or Fatal crash outcomes. After excluding the 24 No-Accident observations, 56 test observations remained for severity scoring, comprising 31 Low-, 11 Moderate-, and 14 High-severity predictions.

$$SB\ Score = 542(14) + 11(11) + (31) = 7{,}740$$

The resulting weighted score was subsequently normalized using Eq. (13) before integration into the HSI.

$$SB* = \frac{7{,}740 - 0}{30{,}352 - 0} = 0.255$$

3.1.3 Performance of motion-based AI sub-model

The overall classification accuracy of the selected AI framework for the third sub-model was 70.60%, indicating its ability to recognize severe traffic conflicts in real-world surveillance videos. As shown in the confusion matrix in Figure 4, the model identified most crash events, and most classification errors occurred between adjacent severity classes with similar motion.

The motion-based model differs from the objective and behavioral models because it is not based on historical crash records or questionnaire responses. Instead, it captures dynamic traffic conditions and vehicle interactions immediately before and during collision events. This supplemental data helps the proposed HIS capture traffic operating characteristics that are not captured in a traditional crash database. The AI-generated calibrated prediction was used as the SM for the dynamic part of the proposed HIS.

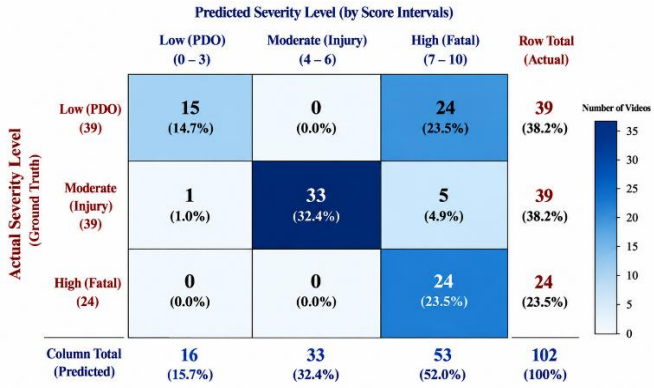


Figure 4. Confusion matrix for three-level crash-severity classification across the complete 102-video dataset

The three-level severity classification performance of the motion-based AI component was evaluated across the entire available dataset of 102 crash videos. For each video, the severity level generated by the rule-based severity-estimation procedure was compared with the corresponding official crash-severity label. The resulting overall accuracy was 70.60%, with precision, recall, and F1-score of 79.7%, 74.4%, and 69.5%, respectively. Because these metrics were calculated across the complete 102-video dataset rather than an independent held-out test subset, they should be interpreted as classification agreement for the available video dataset rather than as independent out-of-sample predictive performance.

Across the complete 102-video dataset, the predicted distribution comprised 16 Low-, 33 Moderate-, and 53 High-severity cases. Accordingly, the motion severity score was calculated as:

$$SM\ Score = 542(53) + 11(33) + 1(16) = 29105$$

The confusion matrix was used separately to assess agreement between the predicted severity categories and the corresponding official severity labels.

The motion severity score was subsequently normalized using Eq. (13) and the theoretical bounds defined in Section 2.5:

$$SM* = \frac{29{,}105 - 0}{55{,}284 - 0} = 0.526$$

3.2 Analytic Hierarchy Process weight determination

A hybrid approach was designed to combine the results of the three sub-models: the objective model, the behavioral model, and the motion-based/AI Model. The following subsections describe the results of applying the AHP methodology for weight determination.

An expert assessment was conducted using AHP to evaluate the relative contribution of each sub-model. 19 experts completed the survey. The participants included universities, government institutions, and engineering consultancy firms working in transportation engineering, road safety, traffic management, and ITS. Organizations involved in the project were Mustansiriyah University, University of Baghdad, University of Kufa, University of Anbar, University of Technology, University of Al-Qadisiyah, East Carolina University, University of Wisconsin–Madison, the Ministry of Higher Education and Scientific Research, Bechtel, Terracon, HVJ Associates, and Alkyal Engineering Consultancy. Table 4 shows the numbers and percentages of experts by years of experience. The results show that about 73.7% of the experts have more than 11 years of professional experience in traffic engineering and road safety, which helps to increase the reliability of the evaluation process.

After applying Eq. (14), GM values shown in Table 5 were obtained.

Table 5 indicates that transportation experts generally considered the objective sub-model more influential than the behavioral sub-model, as reflected by a geometric mean greater than one (1.3643). Likewise, the objective sub-model was judged to be slightly more important than the motion-based AI sub-model (1.2059). Conversely, the comparison between the behavioral and motion-based AI sub-models produced a value below unity (0.5540), indicating that the motion-based AI sub-model was generally preferred over the behavioral sub-model.

Following the calculation of GMs, the A was established, as shown in Table 6. The diagonal elements of the matrix are equal to one because every criterion has equal importance relative to itself. The reciprocal values were computed automatically according to Eq. (15).

Table 4. The numbers and percentages of experts by years of experience

Experience Years	Percentage%	Number of Experts
<5 years	15.8	3
5–10 years	10.5	2
11–20 years	26.3	5
>20 years	47.4	9

Table 5. Geometric Mean values for pairwise comparisons

Comparison	Geometric Mean
SO/SB	1.3643
SO/SM	1.2059
SP/SM	0.5540

Table 6. Aggregated pairwise comparison matrix (A)

	SO	SB	SM
SO	1.0000	1.3643	1.2059
SB	0.7330	1.0000	0.5540
SM	0.8293	1.8050	1.0000
Total (C_j)	2.5623	4.1693	2.7599

Table 7 shows the results of the Normalized pairwise comparison matrix. The sum of each column in the normalized matrix is equal to 1. The normalized matrix provides information on the relative importance of the criteria for each comparison column and is used to approximate the principal eigenvector, as discussed in the next section.

The normalized pairwise comparison matrix clearly represents the relative importance of the three sub-models. The objective sub-model shows the highest normalized values in two of the three comparison columns, while the motion-based/AI sub-model also demonstrates a relatively strong contribution. In contrast, the behavioral sub-model has comparatively lower normalized values. The final priority weights were subsequently obtained by averaging the normalized values across each row, as defined in Eq. (17), and are presented in Table 8.

Table 7. Normalized pairwise comparison matrix

	SO	**SB**	**SM**
SO	0.3903	0.3272	0.4369
SB	0.2861	0.2398	0.2007
SM	0.3236	0.4330	0.3624
Total (C_j)	1	1	1

Table 8. Final priority weights of the developed sub-models

Sub Model	**Weight**	**Rank**
Objective sub-model	0.3848	1
Motion-based/AI sub-model	0.3730	2
Behavioral sub-model	0.2422	3
Total	1.0000	

Based on the results, the objective sub-model received the highest priority weight (0.3848), followed closely by the motion-based/AI sub-model (0.3730). The behavioral sub-model received the lowest priority weight (0.2422). These weights represent the relative importance assigned by the experts to the three sub-models and were subsequently used as weighting coefficients in the HSI.

By substituting the matrix in Table 6 into Eq. (18):

$$A = \begin{bmatrix} 1.000 & 1.3643 & 1.2059 \\ 0.7330 & 1.0000 & 0.5540 \\ 0.8293 & 1.8050 & 1.0000 \end{bmatrix}$$

$$\text{The priority vector is } W = \begin{bmatrix} 0.3848 \\ 0.2422 \\ 0.3730 \end{bmatrix}$$

Therefore,

$$AW = \begin{bmatrix} 1.000 & 1.3643 & 1.2059 \\ 0.7330 & 1.0000 & 0.5540 \\ 0.8293 & 1.8050 & 1.0000 \end{bmatrix} \begin{bmatrix} 0.3848 \\ 0.2422 \\ 0.3730 \end{bmatrix} = \begin{bmatrix} 1.1650 \\ 0.7309 \\ 1.1293 \end{bmatrix}$$

Table 9 shows the estimated eigenvalues for the three indices.

The λ_{max} is equal to the number of criteria, 3; it is then compared with the mean of the three eigenvalues, 3.0243, which is too close. This indicates a high degree of logical consistency of the expert judgments exhibited. The CI is then estimated using Eq. (20); the CI value is 0.01215.

The CR is estimated as CR = 0.01215/0.58 = 0.0210, which is substantially lower than the commonly accepted threshold of 0.10. This result confirms that the pairwise comparisons provided by the participating experts are logically consistent and reliable. Therefore, the derived priority weights are suitable for constructing the final hybrid model. The final results are shown in Table 10. The low CR demonstrates that the participating experts exhibited a high degree of agreement regarding the relative importance of the three predictive models. Consequently, the derived weighting coefficients are considered mathematically reliable and suitable for constructing the proposed HSI.

Table 9. The results of the estimated eigenvalues for the three indices

Sub Model	**Weighted Sum Value (AW_i)**	**Priority Weight (wi)**	λ_i
Objective	1.1650	0.3848	3.0276
Behavioral	0.7309	0.2422	3.0178
Motion-Based AI	1.1293	0.3730	3.0276
Mean (λmax)			3.0243

Table 10. Consistency analysis results

Parameter	**Value**	**Formula Meaning**
λmax	3.0243	Average of λ values
Consistency Index (CI)	0.01215	(λmax - n)/(n-1)
Random Index (RI) (n = 3)	0.5800	RI for 3 criteria
Consistency Ratio (CR)	0.0210	CI/RI

The AHP analysis demonstrated that the three predictive models contribute differently to the proposed HSI because they differ in nature and reliability in providing information. The final weighting coefficients were 38.48% for the objective crash severity model, 37.30% for the motion-based AI model, and 24.22% for the behavioral severity model. All three models have potential value in crash severity estimation, but objective traffic and crash characteristics, as well as dynamic traffic information, are deemed more important than subjective behavioral factors.

The official crash records had the highest weighting coefficient (38.48%), and experts considered them the most reliable source of severity data. As anticipated, police-reported crash data is expected to yield this result because it contains verified information from the standardized procedures police investigators use to gather roadway geometry and traffic control information, weather conditions, collision mechanisms, vehicle characteristics, and crash outcomes.

Engineering factors, such as physical and operational roadway characteristics, strongly influence the severity of crashes. The type of collision, car speed, road alignment, road condition, lighting, and traffic conditions all directly affect how much energy is transferred during a crash and, thus, the severity of any injuries. These factors are important in past crash severity studies and are the most complete source of engineering evidence available. Thus, the transportation experts' greatest confidence in the accuracy, consistency, and utility of official crash data is given to the objective model, which is assigned the highest weight.

In addition, the motion-based AI model was assigned the second-highest weighting coefficient (37.30%), just 1.18 percentage points less than the objective model. The small gap suggests that the dynamic traffic information captured by surveillance videos was considered nearly as important as the information provided by conventional crash records.

The AI model, which is based on motion rather than the

official crash databases that describe what happens after a crash, can report how a vehicle's motion changes just before the collision. By detecting objects, tracking multiple objects, and analyzing motion, the framework identifies abnormal interactions among vehicles, speed changes, trajectory deviations, and conflict patterns that police reports rarely capture. These dynamic properties provide valuable insight into crash mechanisms and can be used to evaluate operational conditions during severe collisions. This is also because computer vision and ITS have become increasingly important in today's road safety management, as reflected in the high weight assigned to the motion-based model. With the growth of surveillance systems, AI traffic monitoring can fill this gap by delivering real-time, objective, and always available safety data, complementing existing crash data. Thus, the motion-based model is an important step toward moving from conventional post-crash analysis to proactive estimation of crash severity.

The lowest weighting coefficient was for the behavioral severity model (24.22%). However, this does not mean that driver behavior is not a key factor in road safety. Rather, it is a measure that reflects the increased uncertainty of a questionnaire-based measure compared with objectively observed crash characteristics and traffic dynamics captured on video. The behavioral model is based on the SI, RPI, demographic characteristics, driving exposure, and previous crash history. These are variables representing potential, unobserved human-related factors that affect crash rates and severity through decision-making, risk-taking, and adherence to traffic laws. While all of these behavioral measures are important in understanding crash causation, they are almost entirely self-reported data and are subject primarily to recall bias, response bias, and subjectivity.

However, the behavioral model provides information not found in either crash data or surveillance videos. The model includes psychological factors such as drivers' attitudes, safety awareness, and perceived risks, which are not directly observable through engineering measurements. For this reason, a model with an objective and motion component was complemented by the behavioral severity model to represent the human aspect of crash severity.

The results support the rationale of the proposed framework that no single data source provides all dimensions required for a comprehensive characterization of crash severity. Engineering evidence comes from official crash databases, motion-based AI delivers dynamic operational evidence, and behavioral indicators provide complementary information on human-related and perceptual factors associated with crash severity. Combining these complementary information sources in the context of AHP thus yields a more comprehensive and balanced severity assessment than any single predictive model.

The relatively balanced weighting coefficients also indicate that the proposed HSI is not overly dependent on any one information source. Rather, it fuses objective engineering evidence and observations of live traffic with behavioral attributes, all within a clear mathematical process. This integrated structure provides an interpretable framework that may support transportation agencies in crash-severity assessment, prioritization of safety interventions, and intelligent transportation system applications.

3.3 Construction of the composite Hybrid Severity Index

Following the development and evaluation of the three complementary sub-models, the proposed HSI was constructed by integrating their normalized severity scores with weighting coefficients derived from the AHP. The final hybrid model was formulated as follows:

$$HSI = 0.3848SO + 0.2422SB + 0.3730SM \quad (22)$$

where, HSI: final hybrid area-level severity score, SO: severity predicted by the objective sub-model, SB: severity predicted by the behavioral sub-model, and SM: severity predicted by the motion-based AI sub-model.

The outputs of the three predictive models were normalized to a common numerical range using Min–Max normalization before integration. Min–Max normalization was applied to the output of the three predictive models prior to integration. The severity scores were then normalized and multiplied by the weighting coefficients calculated using the AHP to calculate the weighted contribution of each model. The normalized scores, weighting coefficients, and weighted contributions are shown in Table 11.

Table 11. Construction of the composite Hybrid Severity Index (HSI)

Predictive Model	Normalized Score	AHP Weight	Weighted Contribution
Objective Severity	0.278	0.3848	0.1070
Behavioral Severity	0.255	0.2422	0.0618
Motion-Based AI	0.526	0.3730	0.1962
HSI	—	—	0.3650

Note: Analytic Hierarchy Process (AHP).

The goal of the calibration step is to map the results of the three separately trained machine learning models to a similar quantitative scale before they are combined to implement the proposed Hybrid Model. Due to differences in datasets, feature spaces, and classification schemes used to build the three models, it is inappropriate to compare one model's output with any of the others. Thus, the prediction outcomes of each model were converted to a normalized machine-learning SI for calibration.

The proposed framework generates an integrated crash severity estimate of 0.3650 in this case. The weighted integration shows that all three predictive models contribute differently to the final HSI, due to their varying abilities to capture different dimensions of crash severity. The objective severity model had the largest AHP weight (38.48%); however, the motion-based AI model produced the largest weighted contribution (0.1962), owing to its higher normalized severity score (0.526). Therefore, the motion-based model accounts for approximately 53.8% of the total HSI value, while the objective and behavioral models contribute approximately 29.3% and 16.9%, respectively.

This study's result emphasizes the extent to which dynamic traffic has influenced the integrated severity assessment. The motion-based AI model can record vehicle paths, changes in

vehicle speed, unusual driving behaviors, and pre-crash interactions that are not seen in traditional crash databases. Therefore, the actual traffic conditions at the time of the incident strongly influence the final severity assignment for the investigated crash.

The second-largest contribution to the HSI comes from the objective severity model. The relatively strong influence results from the significance of verified engineering data from official crash investigations regarding roadway geometry, collision characteristics, environmental conditions, and vehicle attributes. These variables represent physical mechanisms underlying crash occurrence and injury severity and thus remain fundamental factors in determining crash severity.

The behavioral severity model accounts for the least of the final HSIs. However, it remains a key latent human factor that engineering observations or traffic videos cannot capture. The SI, RPI, driving experience, and prior crash experience offer complementary information about driver decision-making and risk-taking behavior. Behavioral information is not dominant in the prediction but adds to the completeness of the integrated framework by capturing psychological attributes that affect the likelihood of crashes.

In general, the results from the three predictive models suggest that they contribute well to the proposed HSI. The proposed framework integrates complementary information sources, namely the historical crash record, the self-reported behavioral characteristics, and the traffic dynamics of the video, in a transparent mathematical framework, instead of relying solely on the historical crash record, the self-reported behavioral characteristics, or using only the traffic dynamics from the video. This holistic approach enables a more complete and meaningful measure of an accident's severity, allowing transportation agencies to prioritize their road safety interventions, budgets, and plans based on evidence.

For interpretive purposes, the continuous HSI was classified using the three predefined equal-width intervals described in Section 2.5: low severity (0.000–≤0.333), moderate severity (>0.333–≤0.667), and High Severity (>0.667–1.000). The resulting HSI of 0.3650 therefore falls within the moderate severity category. These categories are used as interpretive bands rather than empirically derived or externally validated risk thresholds. Accordingly, the Moderate classification indicates the relative position of the study-area HSI within the proposed framework and should not be interpreted as a universally validated intervention threshold. Future applications should evaluate and, where necessary, recalibrate these boundaries using independent datasets from other study areas.

The usefulness of this classification is not limited to crash analysis. The HSI can assist transportation agencies in prioritizing limited resources by identifying locations or traffic scenarios according to their integrated severity level and supporting appropriate engineering, operational, and behavioral safety interventions.

For locations or traffic scenarios classified as moderate severity, the following safety interventions may be considered:

•Improved monitoring, including periodic safety audits and ongoing monitoring of traffic to recognize changes in operational circumstances.

•Focused traffic enforcement, which includes speeding, unsafe passing, wearing helmets, and obeying traffic laws by motorcycle riders and three-wheelers.

•Pavement surface conditions, pavement friction, pavement markings, traffic sign programming, street lighting, and intersection geometry (where appropriate) improvements.

•Traffic management measures such as signal timing optimization, speed management strategies, and conflict reduction treatments at high-risk locations.

•Driver education and awareness campaign on defensive driving, hazard perception, compliance with Road Safety Rules for better SI and RPI.

The interventions demonstrate excellent practical applications of the proposed HSI as a decision-support tool. Instead of just determining the severity of a collision, the HSI allows transportation agencies to make safety decisions based on the combined effects of engineering, behavioral, and dynamic traffic factors. This feature enables better use of financial and operational resources and proactive road safety management.

Lastly, the continuous nature of the HSI offers additional benefits over traditional categorical severity indicators. When conditions on the roadway, driver behavior, or traffic system operations change, the index value reflects these changes, which can be quantified and measured; this enables agencies to track the effectiveness of the safety countermeasures implemented over time. Thus, the proposed framework is also applicable to continuous evaluation of road safety performance in ITS and to evidence-based transportation planning.

Table 12 compares the characteristics of the individual predictive models with those of the proposed HSI.

Table 12. Comparison of the individual models and the proposed composite Hybrid Severity Index (HSI)

Framework	Primary Information Source	Main Characteristics	Output
Objective Crash Severity Model	Official crash database	Engineering, roadway, environmental, and crash characteristics	Historical severity assessment
Behavioral Severity Model	Driver questionnaire	Human behavior, Safety Index (SI), Risk Perception Index (RPI)	Behavioral severity assessment
Motion-Based AI Model	Traffic surveillance videos	Vehicle trajectories, motion dynamics, and traffic interactions	Dynamic severity assessment
Proposed Composite HSI	Integrated multi-source data	Engineering + behavioral + dynamic traffic information	Comprehensive crash severity assessment

The proposed HSI can be considered a step towards integrated safety assessment from a transportation engineering perspective. While traditional methods estimate crash severity in real time using official crash data, the proposed method integrates historical engineering data, driver behavior, and dynamic traffic data into a continuous severity indicator.

The multidimensional representation allows transportation agencies to more fully identify conditions that can pose hazards on roadways, prioritize engineering countermeasures, and allocate limited resources based on the combined effects

of infrastructure, human behavior, and traffic operations.

In addition, the HSI is not based on a black-box prediction but rather on a transparent mathematical formulation, enabling engineers and decision-makers to better understand how different information sources contribute to the final severity estimate. This transparency builds trust in the framework and enables its use in roadway safety management, network screening, ITS, and in the creation of evidence-based policies.

The proposed framework is quite different from previous crash severity studies. Most published research has focused on improving prediction accuracy from a single data source, whether police-reported crash data, questionnaire-based behavioral analysis, or computer vision models. This study, on the other hand, proposes a multi-source hybrid framework that mathematically combines the three complementary domains into a HSI.

In this regard, the proposed HSI should be considered a decision-support system that integrates engineering evidence, human factors, and real-time traffic data into a single, interpretable severity metric, rather than an additional machine-learning-based classifier. This holistic approach is the main scientific result of the present research and can serve as a basis for further research into ITS, EAI, and holistic road safety assessment.

3.4 Sensitivity and component-contribution analysis

3.4.1 Sensitivity analysis

A sensitivity analysis was performed by varying the weights of the proposed HSI components, one at a time, to assess the index's robustness. The HSI is computed as a weighted sum of the three model outputs, which are normalized to the same range, thereby helping determine the sensitivity of the final severity estimate to uncertainty in the weights derived by the experts. This was done through a sensitivity analysis, in which four types of perturbations were applied independently to each predictive model (±5%, ±10%, ±15%, and ±20%). In each perturbation scenario, the other weights were adjusted to satisfy the normalization constraint. The HSI was then recalculated using the modified weighting coefficients while maintaining the normalized severity scores unchanged.

The sensitivity analysis shown in Figure 5 demonstrates that the HSI responds differently to variations in the AHP weights assigned to its three components. Among the evaluated components, the motion-based AI component exhibited the greatest influence on the integrated index. A ±20% perturbation in its AHP weight resulted in an approximately ±5.25% change in the HSI. In comparison, the corresponding maximum changes associated with the objective and behavioral components were approximately ±2.98% and ±1.93%, respectively. The greater sensitivity to the motion-based AI weight can be explained by its relatively high normalized component score (SM = 0.526) compared with the objective (SO = 0.278) and behavioral (SB = 0.255) scores. Consequently, increasing the relative contribution of the motion-based AI component increases the integrated HSI, whereas reducing its weight produces the largest decrease in the index. Conversely, increasing the weights assigned to the objective or behavioral components slightly decreases the HSI because their normalized scores are lower than that of the motion-based AI component. This inverse response results from the proportional renormalization of the remaining AHP weights and should not be interpreted as indicating that greater objective or behavioral severity improves safety.

Importantly, the baseline HSI of approximately 0.365 remained within the moderate severity category under all ±5%, ±10%, ±15%, and ±20% weight-perturbation scenarios. Even under the most influential motion-based AI perturbation, the HSI varied only from approximately 0.346 to 0.384, without changing the overall severity classification. This indicates that the categorical interpretation of the HSI is reasonably robust to moderate uncertainty in the AHP-derived weights. Nevertheless, the comparatively greater sensitivity to the motion-based AI component indicates that the reliability of the integrated index depends more strongly on this component, reinforcing the importance of transparent validation of the video-based severity-estimation procedure.

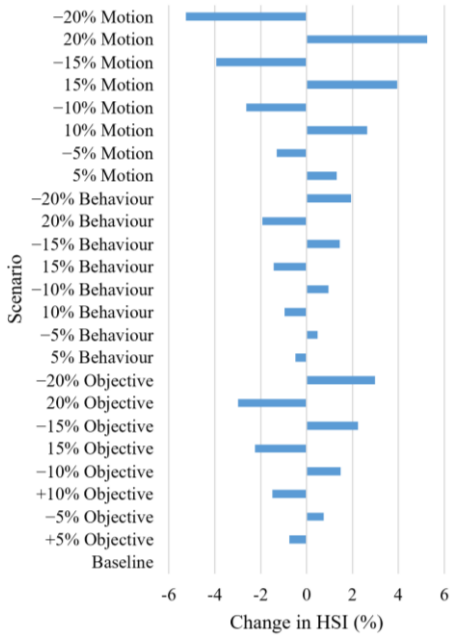


Figure 5. Relative change in the composite Hybrid Severity Index (HSI) under ±5%, ±10%, ±15%, and ±20% perturbations of the Analytic Hierarchy Process (AHP)-derived component weights

In transportation engineering terms, the results show that adopting technologies that capture dynamic traffic interactions, such as intelligent video surveillance and computer vision, could significantly improve crash severity assessment. Meanwhile, the importance of high-quality official crash databases cannot be overstated, as engineering variables continue to impact the HSI significantly. While behavioral data is the lowest in number, it also offers unique insight into human factors and complements objective and motion-based models. One of the key strengths of the proposed multi-source framework over traditional, single-source severity assessment approaches is that it ensures an optimal mix of complementary information without over-relying on any one source.

3.4.2 Ablation analysis

To examine the contribution of each sub-model source, an ablation study was conducted. For each scenario, one weight was dropped, and the remaining weights were re-normalized to keep the total weight equal to 1. When the score of the SO is removed, the HSI is then calculated by:

$$HSI = \frac{0.2422SP + 0.3730SM}{(0.2422 + 0.3730)} \quad (23)$$

Table 13. The relative change in the hybrid index due to removing a sub-model

Scenario	Retained Sub-Models	Renormalized HSI	Relative Change	Classification
Complete hybrid model	Objective + Behavioral + Motion	0.3650	0.00%	Moderate
Without objective model	Behavioral + Motion	0.4193	+14.90%	Moderate
Without behavioral model	Objective + Motion	0.4001	+9.63%	Moderate
Without motion-based model	Objective + Behavioral	0.2691	−26.26%	Low
Objective model only	Objective	0.2780	−23.82%	Low
Behavioral model only	Behavioral	0.2550	−30.12%	Low
Motion-based model only	Motion	0.5260	+44.14%	Moderate

Note: Hybrid Severity Index (HSI).

The Change relative to the complete model is the difference between the original *HSI* and the new one divided by the original *HSI*. The results are shown in Table 13.

The ablation analysis demonstrates that the three components contribute differently to the HSI. When the objective component was removed, the HSI increased from 0.365 to 0.4193 (+14.90%). Similarly, removing the behavioral component increased the HSI to 0.4001 (+9.63%). These increases occurred because the objective (SO = 0.278) and behavioral (SB = 0.255) components had lower normalized scores than the motion-based AI component (SM = 0.526); consequently, their removal and the proportional renormalization of the remaining AHP weights increased the motion-based AI component's relative contribution. In both scenarios, the overall classification remained moderate severity.

In contrast, removing the motion-based AI component reduced the HSI to 0.2691, representing a 26.26% decrease from the baseline and changing the classification from moderate to low severity. This result confirms the comparatively strong influence of the dynamic motion information on the integrated assessment. However, this should not be interpreted as indicating that the motion-based AI component is intrinsically more accurate or more important than the other components. Rather, its greater influence in this case study results from the combination of its relatively high normalized component score and its AHP weight.

The single-component analysis further illustrates the benefit of hybrid integration. The objective-only and behavioral-only scores were 0.278 and 0.255, respectively, both corresponding to low severity, whereas the motion-only score was 0.526 and corresponded to moderate severity. The integrated HSI of 0.3650 remained within the moderate severity category and provided a more balanced representation of the study-area condition than reliance on any individual information source. The purpose of the hybrid framework is therefore not to maximize the severity score, but to combine complementary evidence from historical crash records, self-reported behavioral information, and video-derived dynamic motion characteristics into a multidimensional, less source-dependent study-area-level severity assessment.

4. CONCLUSIONS

This study developed an HSI that integrates three complementary dimensions of two- and three-wheeler safety: historical crash severity derived from official police records, self-reported behavioral information from drivers, and dynamic motion characteristics extracted from traffic surveillance videos. Because the three datasets were independently sampled and were not matched at the individual crash or driver level, integration was performed only at the aggregate study-area level. The resulting HSI should therefore be interpreted as a multi-source area-level decision-support indicator rather than as a prediction of the severity of an individual crash.

The objective and behavioral models were evaluated using independent held-out test sets and achieved classification accuracies of 73.75% and 60.00%, respectively. The motion-based AI component achieved 70.60% three-level classification agreement across the entire dataset of 102 videos. This result reflects agreement within the available dataset rather than out-of-sample predictive accuracy. Following aggregation and normalization, the objective, behavioral, and motion-based AI component scores were 0.278, 0.255, and 0.526, respectively.

AHP assigned weights of 0.3848 to the objective component, 0.2422 to the behavioral component, and 0.3730 to the motion-based AI component. The CR of 0.0210 confirmed acceptable logical consistency of the expert judgments. Combining the normalized component scores with these weights produced an HSI of 0.3650, corresponding to Moderate Severity under the predefined three-level equal-width classification. These thresholds are interpretive categories rather than empirically optimized or externally validated risk boundaries and should therefore not be regarded as universal intervention thresholds.

The sensitivity analyses further clarified the behavior of the proposed index. Under ±5%, ±10%, ±15%, and ±20% perturbations of the AHP weights, the HSI remained within the moderate severity category. The motion-based AI component produced the greatest sensitivity, with approximately ±5.25% change in the HSI at a ±20% weight perturbation, compared with approximately ±2.98% for the objective component and ±1.93% for the behavioral component. Ablation analysis similarly showed that removing the motion-based AI component reduced the HSI from 0.3650 to 0.2691 and changed the classification from Moderate to Low. This greater influence should not be interpreted as evidence that the motion-based AI model is intrinsically more accurate or important; rather, it results from the combination of its relatively high normalized score and its AHP weight in the present case study.

The principal contribution of the proposed framework is therefore the transparent integration of heterogeneous but complementary safety evidence rather than maximization of predictive accuracy. Official crash records characterize historical and engineering conditions, behavioral data provide information on self-reported human and perceptual factors, and video analysis captures dynamic vehicle interactions that are not represented in conventional crash databases. Their integration provides a multidimensional representation of the study-area severity condition while retaining the contribution of each information source.

Several limitations should be considered. The AHP weights

are expert-derived and may vary across geographical and operational contexts. The EPDO-derived coefficients used for the behavioral component are heuristic ordinal weights and do not establish empirical equivalence between behavioral categories and PDO, Injury, or Fatal crash outcomes. The three datasets were independently sampled and were not spatially, temporally, or individually matched. In addition, the Motion-Based AI evaluation was conducted on the complete available video dataset rather than an independent held-out test set, and the HSI classification boundaries were predefined rather than externally validated. Future research should therefore validate the framework using larger and independent datasets, spatially and temporally synchronized multi-source observations, independently tested video data, empirically calibrated behavioral weights and HSI classification thresholds, and applications across different cities and roadway environments.

ACKNOWLEDGMENT

The authors would like to thank Mustansiriyah University (https://uomustansiriyah.edu.iq), Baghdad, Iraq, and the College of Engineering, Highway and Transportation Department, for their support in the present work.

REFERENCES

[1] World Health Organization. (2023). Global status report on road safety 2023. https://www.who.int/publications/i/item/9789240086517.

[2] Konlan, K.D., Hayford, L. (2022). Factors associated with motorcycle-related road traffic crashes in Africa, a scoping review from 2016 to 2022. BMC Public Health, 22(1): 649. https://doi.org/10.1186/s12889-022-13075-2

[3] Das, S., Vierkant, V., Gonzalez, J.C., Kutela, B., Sheykhfard, A. (2023). Bayesian network for motorcycle crash severity analysis. Transportation Research Record, 2677(11): 51-63. https://doi.org/10.1177/03611981231164386

[4] Santos, K., Firme, B., Dias, J.P., Amado, C. (2024). Analysis of motorcycle accident injury severity and performance comparison of machine learning algorithms. Transportation Research Record, 2678(1): 736-748. https://doi.org/10.1177/03611981231172507

[5] Asuelinmen, G.A., Ojolo, S.J., Ajayi, O.O. (2020). Investigating the lateral stability of three-wheeled scooter taxi due to tyre-road forces. Nigerian Journal of Technology, 39(1): 189-195. https://doi.org/10.4314/njt.v39i1.21

[6] Rehman, Z.U., Chaozhe, J., Adanu, E.K., et al. (2025). Assessing heterogeneity in factors influencing three-wheeled motorized rickshaws crash outcomes between weekdays and weekends. Scientific Reports, 15(1): 14164. https://doi.org/10.1038/s41598-025-97847-x

[7] Iranitalab, A., Khattak, A. (2017). Comparison of four statistical and machine learning methods for crash severity prediction. Accident Analysis & Prevention, 108: 27-36. https://doi.org/10.1016/j.aap.2017.08.008

[8] Wahab, L., Jiang, H. (2019). A comparative study on machine learning based algorithms for prediction of motorcycle crash severity. PLoS ONE, 14(4): e0214966. https://doi.org/10.1371/journal.pone.0214966

[9] Aldhari, I., Almoshaogeh, M., Jamal, A., Alharbi, F., Alinizzi, M., Haider, H. (2023). Severity prediction of highway crashes in Saudi Arabia using machine learning techniques. Applied Sciences, 13(1): 233. https://doi.org/10.3390/app13010233

[10] Li, Z., Wang, C., Liao, H., Li, G., Xu, C. (2024). Efficient and robust estimation of single-vehicle crash severity: A mixed logit model with heterogeneity in means and variances. Accident Analysis & Prevention, 196: 107446. https://doi.org/10.1016/j.aap.2023.107446

[11] Sayed, I., Abdelgawad, H., Said, D. (2022). Studying driving behavior and risk perception: A road safety perspective in Egypt. Journal of Engineering and Applied Science, 69(1): 22. https://doi.org/10.1186/s44147-021-00059-z

[12] Junaid, M., Pervez, A., Lee, J.J., Jiang, C., Khan, W.A. (2025). Investigating contributing factors to severe injuries in crashes involving motorized rickshaws in Pakistan. Journal of Safety Research, 94: 157-166. https://doi.org/10.1016/j.jsr.2025.06.004

[13] Sarigiannis, D., Atzemi, M., Oke, J., Christofa, E., Gerasimidis, S. (2024). Feature engineering and decision trees for predicting high crash-risk locations using roadway indicators. Transportation Research Record, 2678(8): 535-548. https://doi.org/10.1177/03611981231217497

[14] Ghasemkhani, B., Balbal, K.F., Birant, K.U., Birant, D. (2025). Ordinal random tree with rank-oriented feature selection (ORT-ROFS): A novel approach for the prediction of road traffic accident severity. Mathematics, 13(2): 310. https://doi.org/10.3390/math13020310

[15] Mostafa, A.M., Aldughayfiq, B., Tarek, M., et al. (2025). AI-based prediction of traffic crash severity for improving road safety and transportation efficiency. Scientific Reports, 15(1): 27468. https://doi.org/10.1038/s41598-025-10970-7

[16] Santos, K., Dias, J.P., Amado, C. (2022). A literature review of machine learning algorithms for crash injury severity prediction. Journal of Safety Research, 80: 254-269. https://doi.org/10.1016/j.jsr.2021.12.007

[17] Ahmed, S., Hossain, M.A., Ray, S.K., Bhuiyan, M.M.I., Sabuj, S.R. (2023). A study on road accident prediction and contributing factors using explainable machine learning models: analysis and performance. Transportation Research Interdisciplinary Perspectives, 19: 100814. https://doi.org/10.1016/j.trip.2023.100814

[18] Hasan, A.S., Jalayer, M., Das, S., Kabir, M.A.B. (2024). Application of machine learning models and SHAP to examine crashes involving young drivers in New Jersey. International Journal of Transportation Science and Technology, 14: 156-170. https://doi.org/10.1016/j.ijtst.2023.04.005

[19] Roudnitski, A. (2024). Evaluating road crash severity prediction with balanced ensemble models. Transport Findings. https://doi.org/10.32866/001c.116820

[20] Dong, S., Khattak, A., Ullah, I., Zhou, J., Hussain, A. (2022). Predicting and analyzing road traffic injury severity using boosting-based ensemble learning models with SHAPley Additive exPlanations. International Journal of Environmental Research and Public Health, 19(5): 2925. https://doi.org/10.3390/ijerph19052925

[21] Zhang, S., Khattak, A., Matara, C.M., Hussain, A., Farooq, A. (2022). Hybrid feature selection-based machine learning classification system for the prediction

of injury severity in single and multiple-vehicle accidents. PLoS ONE, 17(2): e0262941. https://doi.org/10.1371/journal.pone.0262941

[22] Sudheshna, P., AlJawahry, H.M., Phadke, R., Sivasubramanian, S., Kanth, T.V.R. (2024). An effective prediction modelling of traffic accidents based on categorical boosting with grid search. In 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS), Kalaburagi, India, pp. 1-5. https://doi.org/10.1109/ICIICS63763.2024.10859490

[23] Atsir, E.M., Nurmalitasari, N., Sari, A.A. (2025). Traffic accident severity classification system using random forest algorithm. J-INTECH: Journal of Information and Technology, 13(2): 272-281. https://doi.org/10.32664/j-intech.v13i02.2089

[24] Ortelli, N., De Lapparent, M., Varotto, S.F., Bierlaire, M. (2025). An integrated ordered logit and latent variable model for accident injury severity and risk-taking behavior. Transportation Research Part A: Policy and Practice, 192: 104330. https://doi.org/10.1016/j.tra.2024.104330

[25] Kalpana, P., Sowmiya, G., Sri, C.S., Sivapriya, S. (2025). Road traffic accident detection based on YOLOv8 and ByteTrack. AIP Conference Proceedings, 3204(1): 040013. https://doi.org/10.1063/5.0248651

[26] Karim, A., Raza, M.A., Alharthi, Y.Z., et al. (2024). Visual detection of traffic incident through automatic monitoring of vehicle activities. World Electric Vehicle Journal, 15(9): 382. https://doi.org/10.3390/wevj15090382

[27] Minh, Q.N.H., Dinh, N.N., Ho, L.V., Huu, C.P. (2025). Real-time traffic accident detection using YOLOv8. Transportation Research Procedia, 85: 68-75. https://doi.org/10.1016/j.trpro.2025.03.135

[28] Yang, F., Zhang, X., Liu, B. (2022). Video object tracking based on YOLOv7 and DeepSORT. arXiv preprint arXiv:2207.12202. https://doi.org/10.48550/arXiv.2207.12202

[29] Zhou, J., Kwan, C. (2018). Anomaly detection in low-quality traffic monitoring videos using optical flow. In Pattern Recognition and Tracking XXIX, Orlando, FL, USA, pp. 122-132. https://doi.org/10.1117/12.2303651

[30] Kale, K., Pawar, S., Dhulekar, P. (2015). Moving object tracking using optical flow and motion vector estimation. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, India, pp. 1-6. https://doi.org/10.1109/ICRITO.2015.7359323

[31] Nguyen, H.H., Tran-Huu, P.T., Tran, H.M., Ha, S.V.U. (2013). Improved optical flow estimation in traffic monitoring system. In 2013 Third World Congress on Information and Communication Technologies (WICT 2013), Hanoi, Vietnam, pp. 165-169. https://doi.org/10.1109/WICT.2013.7113129

[32] Liao, H., Li, Y., Li, Z., et al. (2024). Real-time accident anticipation for autonomous driving through monocular depth-enhanced 3D modeling. Accident Analysis & Prevention, 207: 107760. https://doi.org/10.1016/j.aap.2024.107760

[33] Bao, W., Yu, Q., Kong, Y. (2021). DRIVE: Deep Reinforced Accident Anticipation with Visual Explanation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, pp. 7619-7628. https://doi.org/10.1109/ICCV48922.2021.00752

[34] Saaty, T.L. (1980). The Analytic Hierarchy Process. New York: McGraw-Hill.

[35] Chakraborty, M., Gates, T.J., Sinha, S. (2023). Causal analysis and classification of traffic crash injury severity using machine learning algorithms. Data Science for Transportation, 5(2): 12. https://doi.org/10.1007/s42421-023-00076-9

[36] Wang, X., Su, Y., Zheng, Z., Xu, L. (2024). Prediction and interpretive of motor vehicle traffic crashes severity based on random forest optimized by meta-heuristic algorithm. Heliyon, 10(16): e35595. https://doi.org/10.1016/j.heliyon.2024.e35595

[37] Chen, F., Liu, X.Q., Yang, J.J., et al. (2025). Traffic accident severity prediction based on an enhanced MSCPO-XGBoost hybrid model. Scientific Reports, 15: 25729. https://doi.org/10.1038/s41598-025-00797-7

[38] Federal Highway Administration. (2026). Highway safety manual. U.S. Department of Transportation. https://highways.dot.gov/safety/data-analysis-tools/highway-safety-manual.

NOMENCLATURE

SI	Safety Index
RPI	Risk Perception Index
ITS	Intelligent Transportation Systems
HSI	Hybrid Severity Index
AHP	Analytic Hierarchy Process
SO	Objective Severity Score
SB	Behavioral Severity Score
SM	Motion-Based Severity Score
CR	Consistency Ratio
WO	Weight of objective score
WB	Weight of behavioral score
WM	Weight of Motion score
PDO	Property Damage Only
SMOTE	Synthetic Minority Over-sampling Technique
EDA	Exploratory data analysis
RF	Random Forest
EFA	Exploratory Factor Analysis
KMO	Kaiser–Meyer–Olkin
AI	Artificial Intelligence
YOLOv8	You Only Look Once version 8
SORT	Simple Online and Real-time Tracking
CI	Consistency Index
XGBoost	Extreme Gradient Boosting
HSM	Highway Safety Manual
EPDO	Equivalent Property Damage Only
GM	The geometric mean
A	Aggregated pairwise comparison matrix
W	Priority (eigenvector) weights
AW	Weighted sum vector
RI	Random Index

Greek symbols

a_ij	The relative importance of criterion i compared with criterion j
w_i	Priority weight of criterion i
n_{ij}	Normalized matrix element

n	Number of criteria
C_j	Sum of column j
λ_i	Eigenvalue estimate for criterion
λ_{max}	the number of the criteria

Subscripts

f_F, f_I, f_{PDO}	The numbers of fatal, injury, and PDO crashes, respectively
according	according