applied sciences

MDPI

Article

Development of Integrated Driving Evaluation Index by Proportion of Autonomous Vehicles for Future Intelligent Transportation Systems

Minkyung Kim [1], Hoseon Kim [1] and Cheol Oh [1,2,*]

[1] Department of Smart City Engineering, Hanyang University Erica Campus, 55 Hanyangdaehak-ro, Sangnok-gu, Ansan 15588, Republic of Korea; alsrud587@hanyang.ac.kr (M.K.); hoseon94@hanyang.ac.kr (H.K.)
[2] Department of Transportation and Logistics Engineering, Hanyang University Erica Campus, 55 Hanyangdaehak-ro, Sangnok-gu, Ansan 15588, Republic of Korea
* Correspondence: cheolo@hanyang.ac.kr; Tel.: +82-31-400-5158

Abstract: As the market penetration rate (MPR) of autonomous vehicles increases, it is expected that the safety of mixed traffic situations will change due to interactions between vehicles. A proactive safety analysis of mixed traffic situations is needed for future intelligent transportation systems; thus, it is necessary to determine the driving safety evaluation indicators that have a significant impact on identifying hazardous sections of actual roads by each MPR. The purpose of this study is to simulate autonomous vehicle behavior by analyzing real-world autonomous vehicle data and to derive a promising integrated driving safety evaluation index for mixed traffic. Autonomous vehicle driving data from an autonomous mobility testbed in Seoul were collected and analyzed to assess autonomous vehicle behavior in VISSIM. The simulation environment was established to match the real road environment. Decision tree (DT) analysis was adopted to derive the indicators influencing the classification of hazardous sections of real roads by MPR. The vehicle–vehicle interaction indicators used to evaluate driving safety were applied as the input variables of the DT, and the classification of real-world hazardous road sections was the output variable. An integrated evaluation index was developed using the promising evaluation indicators and information gains derived for each MPR. The most hazardous section and the factors affecting the driving safety of the section based on the integrated evaluation index for each MPR were then presented. The results of this study can be utilized to proactively identify hazardous road sections in the real world through simulations of mixed traffic conditions.

Keywords: autonomous vehicle data; autonomous vehicle MPR; decision tree; driving safety indicator; integrated evaluation index

Citation: Kim, M.; Kim, H.; Oh, C. Development of Integrated Driving Evaluation Index by Proportion of Autonomous Vehicles for Future Intelligent Transportation Systems. *Appl. Sci.* **2024**, *14*, 9322. https://doi.org/10.3390/app14209322

Academic Editor: Nicholas E. Lownes

Received: 13 September 2024
Revised: 4 October 2024
Accepted: 8 October 2024
Published: 13 October 2024

1. Introduction

Mixed traffic is an environment where a mix of autonomous and manual vehicles are driving; mixed traffic situations will continue until all vehicles are autonomous. In addition, the market penetration rate (MPR) of autonomous vehicles is gradually increasing with the development of autonomous driving system technology. Safety can decrease due to the different driving behaviors of autonomous and manual vehicles. Thus, traffic safety can be affected by the MPR [1]. A proactive safety analysis of mixed traffic is needed for future intelligent transportation systems. Currently, safety analysis is mainly based on virtual-environment-based research due to the lack of real-world driving data. In addition, simulations, which are mainly used for virtual-environment-based research, are utilized in many studies because they can freely represent various situations and current road conditions. It is important to simulate the real-world environment, and implementing autonomous vehicle behavior in simulations is essential. Additionally, studies based on

simulation environments lack validation for the indicators selected and explanations for why the indicators used are appropriate.

Therefore, this study simulates the behavior of autonomous vehicles using real-world autonomous vehicle data to develop an integrated evaluation index that can identify hazardous sections of actual roads. The data were collected from autonomous vehicles driving in an autonomous mobility testbed in Seoul to analyze their driving behavior. The driving risk index (DRI), a deceleration-based event occurrence rate, was computed using these collected autonomous vehicle data; this index defines real-world hazardous sections. The simulation was performed with the same traffic volume, signals, road geometry, number of bus stops, etc., as in a real road environment. Then, simulation-based promising evaluation indicators that can identify real-world hazardous sections were derived through a decision tree (DT) analysis. Vehicle-to-vehicle interaction indicators were adopted as candidate evaluation indicators among the driving safety evaluation indicators and used as input variables in the DT analysis; the DRI-based hazardous road sections were the output variables. The DT analysis revealed promising evaluation indicators that are useful for identifying actual hazardous sections, and these indicators were weighted to develop an integrated evaluation index. In addition, the characteristics of the most hazardous sections were presented based on the integrated evaluation index. The results of this study are expected to proactively identify hazardous sections on actual roads through simulation.

This study consists of the following steps. The existing studies that have analyzed the driving safety of autonomous vehicles by MPR are reviewed, and the differences among these studies are presented. Next, a methodology for analyzing the driving safety of autonomous vehicles and developing an integrated evaluation index is presented. Promising simulation-based evaluation indicators and integrated evaluation indices are derived by evaluating the driving behavior of real-world autonomous vehicles. Finally, the conclusions of this study and future research issues are presented.

2. Literature Review

This study analyzes the driving behavior of autonomous vehicles operating in an autonomous mobility testbed to simulate the behavior of autonomous vehicles. Additionally, simulation-based driving-safety-integrated evaluation indices for each MPR of autonomous vehicles are developed. In this regard, simulation-based studies analyzing the driving safety of autonomous vehicles by MPR are reviewed to draw research opportunities.

The use of simulations to analyze driving safety in mixed traffic can be an important factor in proactively enhancing traffic safety [2,3]. Alzoubaidi et al. used VISSIM to evaluate the safety of connected vehicle (CV) technology at signalized intersections [4]. The indicators utilized in the analysis were time-to-collision (TTC), post-encroachment time (PET), rear-end conflict (REC) count, lane change conflict (LCC) count, and the total number of conflicts (TNCs). By comparing the evaluation index values for each MPR at adjusted and unadjusted signalized intersections, it was determined that, the lower the values for TTC and PET and the higher the values for REC, LCC, and TNC, the more dangerous the driving safety. The safety analysis showed that, at the intersections where CV technology was applied, safety increased across all the indicators, regardless of the MPR. Furthermore, the risk decreased with an increasing MPR. Similarly, Hou evaluated the impact of connected and automated vehicles (CAVs) on mobility and safety in mixed traffic under adverse weather conditions, utilizing traffic safety alternative evaluation indicators such as the TTC and time-exposed time-to-collision (TET) [5]. The evaluation indicator values were compared in clear, rain, and snow weather conditions by MPR, using TTC and TET to evaluate driving safety. The analysis revealed that, as the MPR of CAVs increased, the overall traffic performance and safety under adverse weather conditions increased. In addition, the results of existing studies have shown that increasing the MPR of CAVs improves safety [6–8].

Shinha et al. utilized VISSIM to identify the impact of CAVs on crash frequency and severity [9]. The VISSIM programming interface was used to develop and implement CAV control algorithms to simulate CAV behavior. A signalized intersection was established in

the simulation, and the MPR of CAVs was increased by 10% to evaluate the impact of CAVs on crash frequency and severity. Sekar et al. also evaluated the impact of autonomous vehicles on safety according to the MPR of CAVs, utilizing two driving logics (cautious and all-knowing) for CoExist and Akins' driving logic to evaluate safety impacts [10]. The travel time and number of conflicts were utilized for performance measures to evaluate mobility and safety, respectively. CoExist's driving logic showed a decrease in the number of conflicts as the MPR increased, while Akins' driving logic showed a decrease in the number of conflicts above an MPR of 50. On the other hand, Olia et al. used the TTC and crash probability as evaluation indicators to evaluate the safety impact of the MPR of CAVs, showing that safety increased up to an MPR of 50, but above this value, there was an increased crash probability [11]. Similarly, existing studies have shown that safety decreases or even increases after a certain MPR [12,13]. Ye and Yamamoto evaluated the impact of CAVs on traffic safety at signalized intersections using the TTC [14]. Conflicts were reduced by 90% for an MPR of 50 and by 50% for an MPR of 30. There was no safety impact for MPRs lower than 20%. Essa and Sayed also analyzed the safety impacts of the MPR of CAVs at signalized intersections and found that the crash rate was reduced by approximately 50% at an MPR of 100%, whereas there was no safety impact at an MPR of less than 20% [15].

Morando et al. analyzed the impact on autonomous vehicle safety by utilizing surrogate safety measures [16]. The driving safety of autonomous vehicles at signalized intersections and roundabouts by MPR was evaluated via simulation. The TTC was selected among the surrogate safety measures and analyzed with different thresholds. An autonomous vehicle MPR of 100% reduced the number of conflicts by 20% compared to an MPR of 0% when a conflict was defined as a vehicle-following event with a TTC of 1 s or less at a signalized intersection. Additionally, the number of conflicts consistently decreased as the MPR of autonomous vehicles increased when a TTC of 0.7 s or less was defined as a conflict. In addition, Elawady et al. analyzed the impact of MPR on the traffic mobility and safety of CAVs at signalized intersections [17]. The evaluation indicators were average delay and TTC, where the two thresholds of TTC-based conflicts were 0.75 s and 1 s. The analysis showed that, as the MPR increased, the road capacity of CAVs increased and the travel time decreased. While the number of conflicts did not change significantly depending on the threshold value of the TTC, the results showed that the number of conflicts was 2% lower when the threshold value was 1 s than when it was 0.75 s for an MPR of 50.

Park and Lee analyzed the mobility and safety impacts of MPR for autonomous vehicles [1]. The parameters of the Wiedemann 99 model were adjusted to implement the behavior of autonomous vehicles based on existing studies. Then, mobility and safety analyses were performed at unsignalized intersections and roundabouts. The average delay decreased as the MPR increased at unsignalized intersections, and the number of crashes decreased when there was less variation in the traffic direction. The average delay per vehicle decreased as the MPR increased at roundabouts, while rear-end and lane change conflicts decreased and intersection conflicts increased. El-Hansali et al. implemented the behavior of autonomous vehicles in VISSIM using the Wiedemann 74 and 99 models and conducted a safety analysis using TTC, PET, and conflict points [18]. The results showed that the number of collisions was greater for manual vehicles than for autonomous vehicles, the number of rear-end collisions was greater for conventional vehicles, and the number of lane change collisions was greater for autonomous vehicles.

Although simulation-based driving safety analyses of autonomous vehicle MPRs have been continuously conducted, the basis for selecting the evaluation indicators selected for these driving safety analyses is not clear. Thus, a rational basis that indicators should be able to represent the real-world safety level of mixed traffic streams consisting of AVs and MVs is required. Therefore, this study collected and analyzed real-world autonomous vehicle data and utilized them in the simulation of autonomous vehicle behavior. Existing studies have mainly used TTC and TTC-based conflict counts to perform safety analyses. This study derived inter-vehicle interaction indicators such as the average spacings and SDI-based

collision counts to evaluate driving safety. Furthermore, simulation-based driving safety evaluation indicators that effectively identify hazardous road sections in the real world via autonomous vehicle MPR were derived. In addition, an integrated evaluation index was developed in this study.

3. Methodology

3.1. Overall Framework

This study utilizes autonomous vehicle data to simulate mixed traffic conditions. Additionally, a methodology is developed to derive an integrated driving safety evaluation index based on the MPR of autonomous vehicles, as shown in Figure 1. First, autonomous vehicle data (AVD) collected from real-world autonomous vehicles are used to perform a driving behavior analysis and define hazardous sections on real roads. Then, automated driving behavior based on the AVD analysis is implemented in a traffic simulation with a network and set of scenarios. Vehicle-to-vehicle interaction indicators are selected as candidate simulation-based driving safety evaluation indicators and utilized as DT input variables. Then, a DT model is developed, with the output variables indicating whether an actual road segment defined as normal or hazardous is unsafe. Through the established model, promising simulation evaluation indicators that can effectively identify actual hazardous road sections by autonomous vehicle MPR are derived.

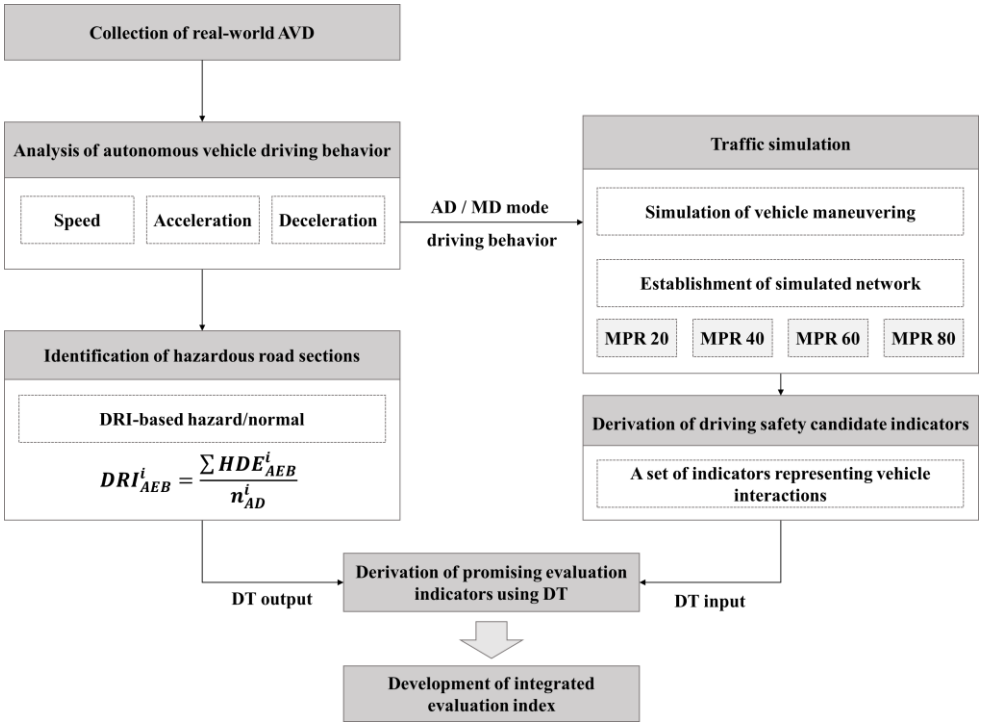


Figure 1. Overall research procedure.

3.2. AVD Collection and Preprocessing

AVD were collected from autonomous vehicles driving in an autonomous mobility testbed in Sangam, Seoul. AVD are collected when the autonomous vehicle is driven in autonomous driving mode (AD mode) and when it is driven in manual driving mode (MD mode). The AVD include the driving date, driving speed, latitude, longitude, and driving mode. The data for this study come from five autonomous vehicles that were driven for a total of three months between 1 August and 31 October 2022. The AVD from the period when the autonomous vehicle was actually operating were used, and the acceleration and jerk were computed from the speed. In addition, the intersection influence zone and noninfluence zone were categorized according to the criteria for determining the intersection influence zone in the Rules on the Connection of Roads with Other Roads of

the Road Act. The intersection influence zone was defined as 30 m from the stop line before entering the intersection and 30 m after the intersection square. The coordinates of the influence and noninfluence zones on the map were collected, and the latitude and longitude data in the AVD were utilized to identify the influence zone. The preprocessed data were used to analyze the driving behavior of vehicles in the influence and noninfluence zones by driving mode.

3.3. Definition of Hazardous Road Sections

The AVD were used to define hazardous sections of real-world roads. Hazardous road sections were determined by whether the autonomous emergency braking (AEB) system, which detects unexpected events based on longitudinal deceleration, was activated. The driving risk index (DRI) was defined as the proportion of events exceeding 0.5 g (4.9 m/s^2), the maximum deceleration of the AEB function proposed in the international standard ISO-22179, among all events in which the AEB system was activated [19]. A total of 219 sections were divided into sections based on the intersection influence area of 30 m. However, school zones, within which the AD mode is not allowed under current Korean law, were excluded from the analysis; thus, a total of 186 sections were utilized in the analysis. The DRI was computed for each section, and a section was defined as hazardous if its DRI was greater than the average for the entire section; otherwise, the section was considered as normal. The DRI-based hazardous sections and normal sections were the output variables of the DT analysis. The *DRI* computation is presented in Equation (1).

$$DRI^{i}_{AEB} = \frac{\sum HDE^{i}_{AEB}}{n^{i}_{AD}} \quad (1)$$

where DRI^{i}_{AEB} is the autonomous driving risk index based on hazardous events in segment i, HDE^{i}_{AEB} is the number of hazardous events occurring in segment i, and n^{i}_{AD} is the number of driving events in AD mode for segment i (every second).

3.4. Simulation of Autonomous Vehicle Driving Behavior

VISSIM, a microscopic traffic simulation tool, was utilized to evaluate the safety of mixed traffic conditions on urban roads depending on the MPR of autonomous vehicles. The Wiedemann 74 model, which is mainly used for urban roads, was selected to construct the car-following model. The parameters that can adjust the behavior of autonomous vehicles are presented in Table 1 [20]. Three distribution parameters could be used to adjust the behavior of autonomous vehicles in the simulations using the analysis results for autonomous vehicle driving behavior on actual roads. 'Desired speed distribution' is a parameter that allows the speed distribution to be adjusted according to the intersection influence and noninfluence zones and also allows the speed distribution of vehicles to be adjusted. Additionally, 'desired acceleration functions' and 'desired deceleration functions' are parameters that adjust the acceleration and deceleration distributions of vehicles. The driving behavior of vehicles is adjusted based on their speed, acceleration, and deceleration distributions according to the driving mode. For other parameters, such as car-following and lane change parameters, values derived from the literature were applied [21].

Table 1. Adjustment of VISSIM driver behavior parameters for each driving mode.

Division	Parameter	MD Mode	AD Mode
Distribution	Desired speed distribution Desired acceleration functions Desired deceleration functions	Results of driving behavior using AVD	
Car following [1] (Wiedemann 74)	Average standstill distance [m]	2.0	2.0
	Additive part of safety distance [m]	2.0	1.5
	Multiplicative part of safety distance [m]	3.0	2.5

Table 1. *Cont.*

Division	Parameter	MD Mode	AD Mode
Lane change [1]	Waiting time before diffusion [s]	60.0	60.0
	Minimum headway (front/rear) [m]	0.5	0.2
	Safety distance reduction factor [m]	0.6	0.3
	Maximum deceleration for cooperative braking [m/s^2]	−3.0	−3.0

[1] Parameter values adopted from prior research.

The simulation analysis section was selected as the autonomous mobility testbed in Seoul, Sangam. The simulated sections were divided in the same way as the DRI-based hazardous sections to compare the actual hazardous road sections and the simulated hazardous sections. The traffic volume was obtained from View-T, a national traffic database used when establishing a network. Then, signals were applied using the actual signals at the intersections. In addition, bus stops, crosswalks, bicycle priority roads, and other road facilities were constructed to obtain an environment similar to that of the actual roads. Four scenarios were established with the MPR of autonomous vehicles ranging from 20% to 80% in 20% increments. In total, 90 min of simulation experiments were conducted, of which 30 min was a warm-up period.

3.5. Derivation of Promising Evaluation Indicators and Development of an Integrated Evaluation Index

Eleven vehicle interaction indicators were selected for the analysis of this study. The indicators included the TTC-based conflict count, stopping distance index (SDI)-based crash count, average deceleration rate to avoid a crash (DRAC), DRAC-based conflict count, crash potential index (CPI), average spacing, standard deviation of spacing, time-varying volatility (VF) of spacing, average headway, standard deviation of headway, and time-varying volatility of headway.

The TTC-based conflict count is defined as the number of conflicts with a TTC of less than 1.5 s [22]. The SDI is a safety evaluation indicator that judges safe/unsafe conditions based on the difference between the minimum stopping distance of the leading vehicle and that of the trailing vehicle. When the stopping distance of the trailing vehicle is greater than the stopping distance of the leading vehicle, the scenario is unsafe, as shown in Equation (2). A conflict is defined as an SDI of less than 0 [23]. DRAC is the collision avoidance deceleration when the trailing vehicle recognizes a hazard and starts to slow down. As the DRAC increases, the safety decreases; the formula for this is shown in Equation (3). Additionally, if the DRAC exceeds 3.35 m/s^2, it is defined as a conflict [24]. The CPI is the probability that the DRAC exceeds the maximum available deceleration rate (MADR) in a given time interval [25]. As the CPI increases, safety decreases; this computation is shown in Equation (4). The time-varying volatility is the variability of the indicator value over time [26]. As the time-varying volatility increases, the safety decreases, as shown in Equation (5).

$$SDI_t = d_{LV,t} - d_{FV,t} \tag{2}$$

where $d_{FV,t}$ is the stopping distance of the trailing vehicle at time t (m) and $d_{LV,t}$ is the stopping distance of the leading vehicle at time t (m).

$$DRAC_t = \frac{(V_{FV,t} - V_{LV,t})^2}{2s_t} \tag{3}$$

where $V_{FV,t}$ is the velocity of the trailing vehicle at time t (m/s), $V_{LV,t}$ is the velocity of the leading vehicle at time t (m/s), and s_t is the spacing at time t (m).

$$CPI = \frac{\sum \Pr(MADR \leq DRAC_t)}{T} \tag{4}$$

where $MADR$ is the maximum available deceleration rate (m/s^2), $DRAC_t$ is the maximum deceleration at time t (m/s^2), and T is the analysis unit time.

$$VF = \sqrt{\frac{1}{n-1}\sum_{t=1}^{T}(r_t - \bar{r})^2} \quad (5)$$

where r_t is the relative change in the indicator value at time t (%), n is the number of data points, and t is the sampling interval.

Both autonomous vehicle and manually driven vehicle driving data were utilized to derive the aforementioned evaluation indicators for the evaluation of mixed traffic conditions. For the simulation-based hazardous road sections, 186 sections were selected, which were the same as the sections defined as actual hazardous road sections. The evaluation indicators were calculated for each road section. The definitions and equations for obtaining the candidate evaluation indicators are presented in Table 2.

Table 2. Candidate indicators for driving safety analysis.

Variable Name	Definition	Equation
TTC-based conflict count	A safety assessment metric that determines safe/unsafe conditions based on TTC	$TTC_t = \frac{s_t}{V_{FV,t} - V_{LV,t}}$
SDI-based conflict count	A safety assessment metric that determines safe/unsafe conditions based on the difference of minimum stopping distances between the leading and following vehicles	$SDI_t = d_{LV,t} - d_{FV,t}$
Avg. DRAC	The average collision evasion deceleration rate when the following vehicle begins to perceive a hazardous situation and initiates deceleration	$DRAC_t = \frac{(V_{FV,t} - V_{LV,t})^2}{2s_t}$
DRAC-based conflict count	When the DRAC is greater than 3.35 m/s^2, it is defined as a conflict	$Number\ of\ DRAC >$ 3.35 m/s^2
CPI	The probability of the DRAC exceeding the MADR within a given time interval	$CPI = \frac{\sum \Pr(MADR \le DRAC_t)}{T}$
Avg. spacing	The average distance from the rear of the leading vehicle to the front of the following vehicle	$\bar{s} = \frac{1}{n}\sum_{t}^{T} s_t$
Std. spacing	The standard deviation of vehicle spacing	$\sigma_s = \sqrt{\frac{\sum(s_t - \bar{s})^2}{n}}$
VF spacing	The spacing variability over time	$VF_s = \sqrt{\frac{1}{n-1}\sum_{t=1}^{T}(r_t - \bar{r})^2}$
Avg. headway	The time difference between the preceding vehicle and following vehicles when passing a specific point on the road	$\bar{h} = \frac{1}{n}\sum_{t}^{T} h_t$
Std. headway	The standard deviation of headway	$\sigma_h = \sqrt{\frac{\sum(h_t - \bar{h})^2}{n}}$
VF headway	The headway variability over time	$VF_h = \sqrt{\frac{1}{n-1}\sum_{t=1}^{T}(r_t - \bar{r})^2}$

The DT analysis technique was used to select promising evaluation indicators that can be used to effectively identify hazardous sections on actual roads. The 11 simulation-based vehicle interaction safety evaluation indicators were used as input variables in the DT analysis. Then, binary indicators representing DRI-based hazardous road sections and normal sections were used as output variables to derive promising evaluation indicators that affect the classification of real hazardous road sections. A DT is an analytical method that categorizes or predicts data by plotting decision rules in a tree structure according to specific criteria. DTs are intuitively easy to understand and are often used to determine which variables have the most significant impact by providing key variables and separation criteria, allowing for the easy interpretation of results [27]. It is necessary to apply an optimal parameter, a hyperparameter, to prevent overfitting and select an appropriate tree

size for DT analysis. The information gain is a quantification of the independent variable and the criterion value that results in the lowest entropy between parent and child nodes when classifying data [28]. The information gain quantifies the impact of each input variable on the classification of the output variable and can be used to determine the importance of variables. Information gain is a value between 0 and 1, where higher values mean that the independent variable has a greater impact on classification. Thus, indicators with large information gains are good indicators for identifying hazardous sections. An evaluation indicator with a nonzero information gain is defined as a promising evaluation indicator that increases the accuracy of real-world hazardous road section classification, which can be computed through the formula in Equation (6).

$$Information\ gain = Entropy(before) - \sum_{j=1}^{k} Entropy(j, after) \tag{6}$$

where $Entropy(before)$ is the prior entropy, $Entropy(j, after)$ is the posterior entropy, and j is a child node.

The definition of an integrated evaluation index according to MPR requires weighting by promising indicators. Weights are assigned using the information gain, which allows for comparisons of the importance of variables. Since the sum of the information gains of all promising indicators is 1, the derived value can be utilized as a weight. Next, normalization should be performed when computing the integrated index due to the different units of each promising indicator and the range of indicator values. Min–max normalization was used, a normalization method that represents different indicators as values between 0 and 1, as shown in Equation (7) [29]. For each scenario, a total of four integrated indices were computed, with a higher value indicating a more hazardous section. The integrated index can be computed by using Equation (8) for each scenario.

$$V_x = \frac{x - Min.(x)}{Max.(x) - Min.(x)} \tag{7}$$

where V_x is the normalized value of the evaluation indicator and x is the evaluation indicator value.

$$Y_{MPR_j} = w_1 * (\overline{V_1}) + w_2 * (\overline{V_2}) + \ldots + w_n * (\overline{V_n}) \tag{8}$$

where Y_{MPR_j} is the integrated evaluation index for the MPR j, w_i is the information gain weight of evaluation indicator i, and $\overline{V_i}$ is the average normalized value of evaluation index i for each section.

The variables that normalize the evaluation indicators used to develop the integrated evaluation index are presented as follows. First, the TTC-based conflict count is V_{TTC}^{CC}, the SDI-based conflict count is V_{SDI}^{CC}, and the DRAC-based conflict count is V_{DRAC}^{CC}. Second, the average spacing and average headway indicate greater safety as their values increase. Therefore, the integrated evaluation index uses the inverse of these two indicators, so that larger values represent more hazardous sections. This is shown as $V_{spac.}^{1/Avg.}$ and $V_{hdwy.}^{1/Avg.}$. Finally, the normalized average DRAC is expressed as $V_{DRAC}^{Avg.}$, and the time-varying volatility and standard deviation are represented as $V_{indicator}^{VF}$ and $V_{indicator}^{Std.}$.

4. Results

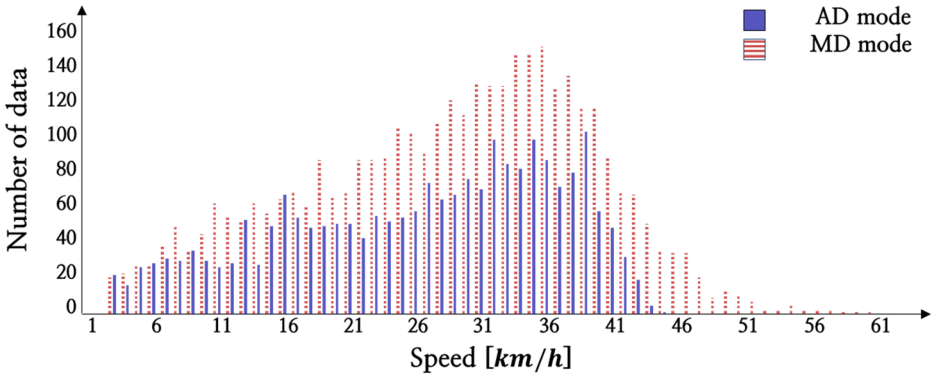
4.1. Adjustment of VISSIM Parameters Based on Driving Behavior Analysis

Descriptive statistics were obtained to identify the characteristics of the different driving behaviors of autonomous vehicles and manual vehicles. The descriptive statistics of speed for each driving mode in the areas influenced and not influenced by intersections were compared using real-world AVD, and the results are presented in Table 3. The average and standard deviation of the AD mode speed were lower than those of the MD mode speed in the intersection-influenced zone, which can be interpreted as more stable driving. Since the AD mode had a slightly greater average speed and lower standard deviation of

speed than the MD mode in the noninfluence zone, the AD mode can be interpreted as leading to faster driving than the MD mode in the noninfluence zone and slower driving as the vehicle entered the intersection-influenced zone.

Table 3. Results of speed driving behavior analysis.

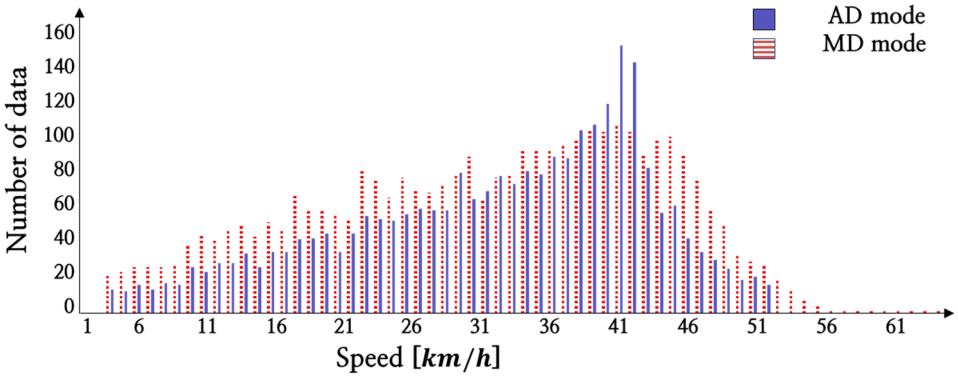
(a) Intersection Influence Zone



Descriptive statistics

Mode	N	Avg.	Std.	Var.	85th%	Range	Min.	Max.
AD	2191	25.56	10.64	113.23	36.00	42	1	43
MD	3585	27.57	11.31	127.98	39.00	67	1	68

(b) Intersection Noninfluence Zone



Descriptive statistics

Mode	N	Avg.	Std.	Var.	85th%	Range	Min.	Max.
AD	2975	29.67	11.24	126.41	40.00	49	1	50
MD	4279	29.22	12.67	160.53	42.00	70	1	71

The VISSIM parameters that could be adjusted based on the results of the driving behavior analysis were the desired speed distribution, the desired acceleration functions, and the desired deceleration functions. The desired speed distribution can be applied to the simulation in two ways. First, the behavior of the vehicles can be adjusted, in which case, the speed distribution is applied without distinguishing between the intersection-influenced zone and the noninfluence zone. Second, the desired distribution can be applied to the speed distribution of the road, in which case, it is different in the influenced zone and the noninfluence zone.

The desired acceleration functions and desired deceleration functions are parameters that can only be adjusted by utilizing all three following values: average, maximum, and minimum acceleration or deceleration. Both parameters can be applied in the form of a distribution that shows the acceleration or deceleration values for a specific speed.

The three distribution parameters of VISSIM were adjusted based on the results of the driving behavior analysis in terms of speed.

4.2. Derivation of Actual Hazardous Road Sections

The first step was to define the actual hazardous road sections to select the indicators that were effective in identifying such sections from among the candidate indicators. A total of 3,238,883 autonomous driving mode data points and 867,065 deceleration data points were collected from the AVD. There were 9658 events with AEB values greater than 4.9 m/s^2. These were used to obtain the DRI for each section. The descriptive statistics of deceleration in AEB-based risk events and the descriptive statistics of DRI by section are presented in Table 4. Also, an example of a DRI-based hazardous section is shown in Figure 2. The sections were chosen as an example of a hazardous section when three or more consecutive road sections were hazardous. The example of a hazardous section shown in the figure is a section where autonomous vehicles perform a right turn.

Table 4. Descriptive statistics of deceleration for risky event and DRI by section.

Descriptive Statistics								
Statistics	**N**	**Avg.**	**Std.**	**Var.**	**85th%**	**Range**	**Min.**	**Max.**
Deceleration for risk event	9658	−0.798	0.726	0.527	−1.111	13.833	−13.889	−0.056
DRI by section	186	0.338	0.947	0.896	0.464	5.708	0	5.708

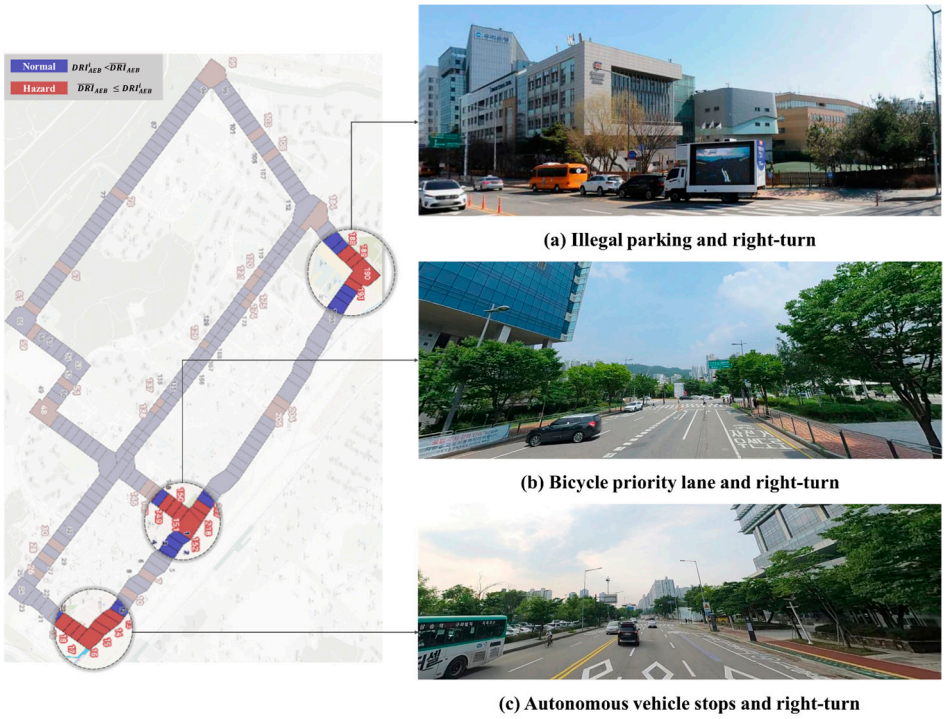


Figure 2. Example of DRI-based hazardous road section.

4.3. Derivation of Promising Evaluation Indicators and the Integrated Evaluation Index

A DT analysis was conducted to derive the promising evaluation indicators that should be used to identify hazardous sections on actual roads from among the candidate interaction safety evaluation indicators. In establishing the DT model, simulation-based interaction safety evaluation indicators were used as the input variables; the output variables were the DRI-based hazardous road sections and normal sections defined using AVD. All scenarios were analyzed with a 7:3 ratio of training data to test data, with a total of 186 data points. The hyperparameters of the model were determined from this analysis. The definitions of

the parameters and the optimized values of the parameters for each scenario are presented in Table 5 [30].

Table 5. Parameter optimization results.

Parameter	Definition	MPR 20	MPR 40	MPR 60	MPR 80
Class weight	Whether to apply the weight of each class	None	None	None	None
Criterion	The function to measure the quality of a split	Gini	Entropy	Entropy	Entropy
Splitter	The strategy used to choose the split at each node	Best	Best	Best	Best
Max depth	The maximum depth of the tree	3	4	4	4
Min samples split	The minimum number of samples required to split an intermediate node	13	13	7	13
Min sample leaf	The minimum number of samples required to be at a leaf node	7	13	11	13
Max features	The number of features to consider when looking for the best split	Sqrt	Log2	Sqrt	Log2

The confusion matrix and information gain of MPR 20 are shown in Table 6 as an example of the results of the decision tree analysis. The MPR 20 scenario resulted in a classification accuracy of 87.50%, while the MPR 40, 60, and 80 scenarios resulted in 78.57%, 85.71%, and 80.36% accuracies, respectively. The number of promising indicators that affected the classification of hazardous sections varied by scenario. First, four promising indicators of MPR 20 and their respective information gains are as follows. The average spacing was 0.359, the SDI-based conflict count was 0.332, the time-varying volatility of headway was 0.270, and the standard deviation of headway was 0.039. These values were used as weights to develop an integrated index. Second, three promising indicators were derived for MPR 40, namely, the average spacing, TTC-based conflict count, and headway-based time-varying volatility, which were used as weights to develop the integrated evaluation index. For MPRs of 60 and 80, the same method was used to develop the integrated evaluation index. The formulas for the integrated evaluation indices for the scenarios with MPRs of 20, 40, 60, and 80 are presented in Equations (9), (10), (11), and (12), respectively. The average spacing was found to be a promising indicator and was included in the integrated index in all scenarios. Additionally, it was the indicator that best identified hazardous road sections among the candidate evaluation indicators for each scenario. Regardless of the MPR of autonomous vehicles, spacing is an indicator that affects driving safety; the shorter the distance is, the more hazardous driving is. The developed integrated evaluation indices were utilized to derive hazardous road sections for each scenario.

$$Y_{MPR_{20}} = 0.359 * \left(\overline{V_{spac.}^{1/Avg.}}\right) + 0.332 * \left(\overline{V_{SDI}^{CC}}\right) + 0.270 * \left(\overline{V_{hdwy.}^{VF}}\right) + 0.039 * \left(\overline{V_{hdwy.}^{Std.}}\right) \quad (9)$$

$$Y_{MPR_{40}} = 0.671 * \left(\overline{V_{spac.}^{1/Avg.}}\right) + 0.180 * \left(\overline{V_{TTC}^{CC}}\right) + 0.149 * \left(\overline{V_{hdwy.}^{VF}}\right) \quad (10)$$

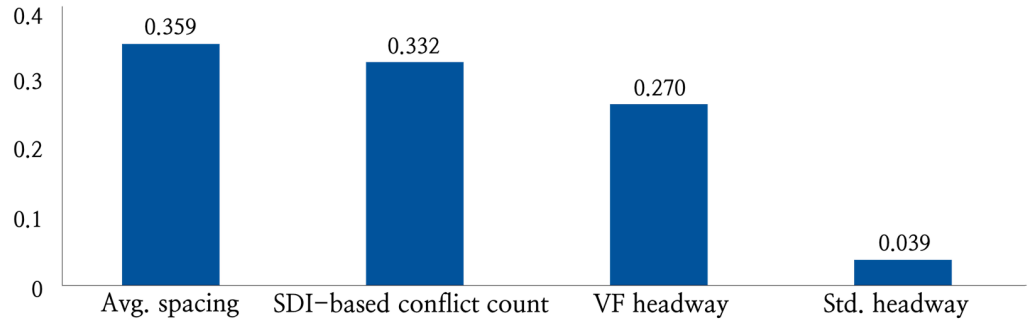
$$Y_{MPR_{60}} = 0.316 * \left(\overline{V_{spac.}^{1/Avg.}}\right) + 0.205 * \left(\overline{V_{spac.}^{VF}}\right) + 0.127 * \left(\overline{V_{SDI}^{CC}}\right) + 0.126 * \left(\overline{V_{hdwy.}^{1/Avg.}}\right) + 0.116 * \left(\overline{V_{hdwy.}^{Std.}}\right) + 0.110 * \left(\overline{V_{DRAC}^{Avg.}}\right) \quad (11)$$

$$Y_{MPR_{80}} = 0.419 * \left(\overline{V_{spac.}^{1/Avg.}}\right) + 0.229 * \left(\overline{V_{SDI}^{CC}}\right) + 0.151 * \left(\overline{V_{hdwy.}^{VF}}\right) + 0.117 * \left(\overline{V_{hdwy.}^{Std.}}\right) + 0.084 * \left(\overline{V_{DRAC}^{CC}}\right) \quad (12)$$

Table 6. An example of DT analysis results.

Confusion Matrix				
Criterion		Prediction		Recall (%)
		Normal	Hazard	
Actual	Normal	41	2	69.57
	Hazard	5	8	
Precision (%)		80.00		87.50

Information Gain



The road characteristics of the most hazardous sections according to the integrated indicators for each scenario are shown in Figure 3. The most hazardous section in the MPR 20 scenario was segment 93, as shown in Figure 3a. Segment 93 is a four-lane approach that includes an exclusive right-turn lane. Slowing down to make a right turn can result in a rear-end collision with a vehicle that does not maintain a safe distance. Additionally, when a vehicle stops due to pedestrians or bicyclists moving to the traffic island, it may interact with the following vehicle, increasing the potential for conflict. The most hazardous section in the case of MPR 40 was segment 23, as shown in Figure 3b, where a lane drop exists. In segment 23, the lanes decrease to one lane, and a right turn is necessary to join the main road. As right-turning vehicles in this section join the main road, there are interactions with vehicles going straight on the main road, which would increase the potential for conflicts. In addition, pedestrians crossing at the unsignalized crosswalk can cause collisions due to the difficulty of maintaining a safe distance from following vehicles when stopping.

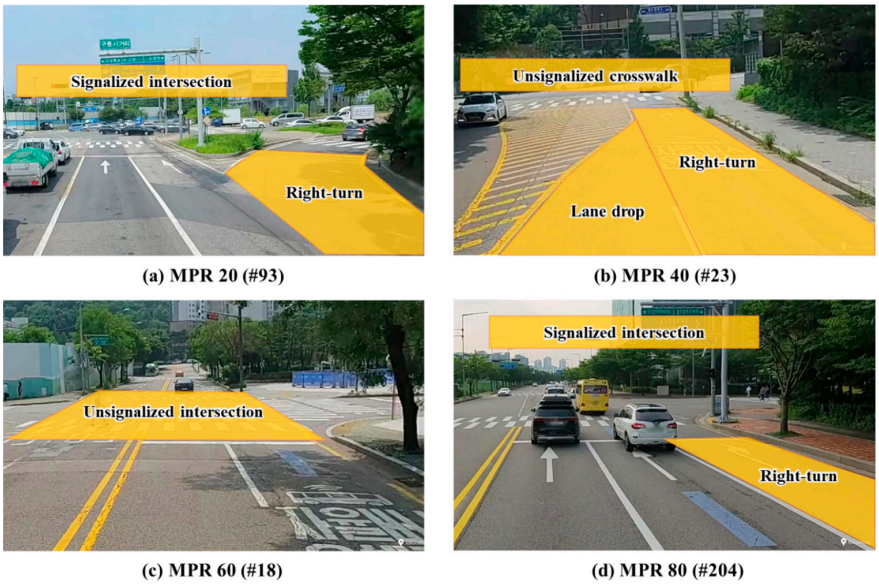


Figure 3. Most hazardous section based on integrated evaluation index by different MPRs.

The most hazardous section in the case of MPR 60 was segment 18, as shown in Figure 3c, with an approach to an unsignalized intersection. A vehicle stopping when a pedestrian crosses the unsignalized intersection can result in a rear-end collision with a following vehicle, while driving cautiously by decelerating before crossing the unsignalized intersection can result in an interaction with a following vehicle, which can cause a crash. In the MPR 80 scenario, the most hazardous section based on the integrated index was segment 204, as shown in Figure 3d. Segment 204 is a signalized intersection with three lanes for each direction, where frequent stop-and-go traffic occurs. The rightmost lane of the section is a right-turn lane, which requires vehicles that aim to continue straight from the right-turn lane or perform a right turn from the straight lane to perform a lane change. Safety can be degraded due to interactions between vehicles changing lanes.

When comparing the integrated evaluation index values of the most hazardous sections by MPR, MPR 20 had a value of 0.462, MPR 40 had a value of 0.791, MPR 60 had a value of 0.466, and MPR 80 had a value of 0.496. The results for the four MPR scenarios in this analysis show that the most hazardous section of MPR 40, section 23, had the highest integrated evaluation index value of 0.791. Segment 23, which requires a lane change due to a lane drop, a right turn, and a stop due to an unsignalized crosswalk, had a high value of the integrated evaluation index because it is a segment where various driving influence factors should be considered.

5. Conclusions

Mixed traffic conditions continue to persist; in addition, the MPR of autonomous vehicles on roads is expected to gradually increase as autonomous driving technology develops. Recently, studies have been performed that consider the MPR of autonomous vehicles for future intelligent transportation systems, and it is important to include a realistic representation of the real-world road environment in a simulation to ensure the reliability of such research. In addition, it is necessary to select realistic evaluation indicators that effectively identify hazardous sections for driving safety on actual roads from among various evaluation indicators in the simulation environment.

Driving data of autonomous vehicles in the autonomous mobility testbed were collected to analyze driving behavior by driving mode. Using this information, promising evaluation indicators for determining hazardous sections on actual roads were derived after establishing a realistic simulation environment, and integrated evaluation indices were developed. The intersection-influenced zone and noninfluence zone were separated by driving mode when analyzing driving behavior. The results showed that the AD mode exhibited more stable driving behavior than the MD mode did. The driving behavior analysis results were used to simulate the behavior of autonomous and manual vehicles according to the driving behavior of vehicles on actual roads. Additionally, the simulation network was established to be the same as the autonomous mobility testbed by including factors such as bus stops and crosswalks, and scenarios were examined in which the MPR of autonomous vehicles varied. A total of 11 interaction evaluation indicators were selected to evaluate driving safety according to the interactions between vehicles under mixed traffic conditions. Then, a DT analysis was conducted to derive evaluation indicators that affect the identification of hazardous road sections, which is defined as a classification problem. A DT model was developed with the interaction safety evaluation indicators as input variables and the actual hazardous road sections as output variables. The results of the DT analysis showed that the promising evaluation indicators that improved the accuracy of hazardous section classification varied by scenario. The promising evaluation indicators for each scenario and the information gains of the evaluation indicators were used as weights to develop an integrated evaluation index. Therefore, this study applied driving behavior in simulations according to analyses of real-world autonomous vehicles. The parameters adjusted the vehicle behavior, and this technique can be used when conducting research using urban road simulation. In addition, the numbers of lanes, signals, crosswalks, illegal parking vehicles, bus stops, etc., were constructed similarly to those of the actual road environment when establishing the simulation network, which ensured the reliability of

the results of this study. Furthermore, the proposed methodology for developing the integrated index is expected to be utilized for the selection of evaluation indices in future simulation-based driving safety analyses of vehicle interactions. Policymakers can develop effective safety measures to address issues associated with mixed traffic situations that would be affected by MPRs. In addition, proper evaluation indicators can be selected in the consideration of the level of autonomous vehicles being developed and promote the development of autonomous vehicle technology.

Further research should be conducted to obtain more generalized results. First, more AVD should be collected. Since the AVD used in this study can only support the analysis of longitudinal individual vehicle driving behavior, this was the type of analysis performed. In addition, because the definition of actual hazardous road sections is based on longitudinal deceleration, it is necessary to apply the methodology of this study again after collecting further data in the future. Second, mapping actual crash data with the integrated index needs to be conducted to evaluate the feasibility of the outcomes of this study. There is a need for a comparative analysis of accident occurrence zones and hazard zones based on the integrated evaluation index derived in this study using the accident history data of autonomous vehicles. Additional incident history data can be collected to perform a comparative analysis in order to improve the reliability of the results of this study. Finally, real-world autonomous vehicle data can be widely used to support policymakers' decision making. As autonomous vehicle technology develops, real-world data need to be collected, and the proposed index in this study needs to be derived based on such real-world data.

Author Contributions: Conceptualization, C.O. and M.K.; methodology, C.O. and M.K.; formal analysis, M.K.; data curation, M.K. and H.K.; writing—original draft preparation, M.K.; writing—review and editing, C.O. and H.K.; supervision, C.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant RS-2021-KA160881, Future Road Design and Testing for Connected and Autonomous Vehicles).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Park, S.Y.; Lee, S.Y. Traffic Safety Analysis according to Autonomous driving MPR (Market Penetration Rate). *J. Korea Contents Assoc.* **2023**, *23*, 57–66. [CrossRef]
2. Hussain, M.S.; Bahrha, G.; Goswami, A.K. An integrated VISSIM-SSAM approach to predicting and mitigating pedestrian crashes and severity along urban crossings. *Case Stud. Transp. Policy* **2024**, *15*, 101153. [CrossRef]
3. Killi, D.V.; Vedagiri, P. Proactive evaluation of traffic safety at an unsignalized intersection using micro-simulation. *J. Traffic Logis. Eng.* **2014**, *2*, 140–145. [CrossRef]
4. Alzoubaidi, M.; Zlatkovic, M.; Jadaan, K.; Farid, A. Safety assessment of coordinated signalized intersections in a connected vehicle environment: A microsimulation approach. *Int. J. Inj. Control Saf. Promot.* **2023**, *30*, 26–33. [CrossRef]
5. Hou, G. Evaluating Efficiency and Safety of Mixed Traffic with Connected and Autonomous Vehicles in Adverse Weather. *Sustainability* **2023**, *15*, 3138. [CrossRef]
6. Abdel-Aty, M.; Wu, Y.; Saad, M.; Rahman, M.S. Safety and operational impact of connected vehicles' lane configuration on freeway facilities with managed lanes. *Accid. Anal. Prev.* **2020**, *144*, 105616. [CrossRef]
7. Guglielmi, J.; Yanagisawa, M.; Swanson, E.; Stevens, S.; Najm, W. *Estimation of Safety Benefits for Heavy-Vehicle Crash Warning Applications Based on Vehicle-to-Vehicle Communications*; No. DOT HS 812 429; Department of Transportation, National Highway Traffic Safety Administration: Washington, DC, USA, 2017.
8. Virdi, N.; Grzybowska, H.; Waller, S.T.; Dixit, V. A safety assessment of mixed fleets with connected and autonomous vehicles using the surrogate safety assessment module. *Accid. Anal. Prev.* **2019**, *131*, 95–111. [CrossRef]

9. Sinha, A.; Chand, S.; Wijayaratna, K.P.; Virdi, N.; Dixit, V. Comprehensive safety assessment in mixed fleets with connected and automated vehicles: A crash severity and rate evaluation of conventional vehicles. *Accid. Anal. Prev.* **2020**, *142*, 105567. [CrossRef] [PubMed]
10. Sekar, N.K.; Malaghan, V.; Pawar, D.S. Micro-simulation insights into the safety and operational benefits of autonomous vehicles. *J. Intell. Connect. Veh.* **2023**, *6*, 202–210. [CrossRef]
11. Olia, A.; Abdelgawad, H.; Abdulhai, B.; Razavi, S.N. Assessing the potential impacts of connected vehicles: Mobility, environmental, and safety perspectives. *J. Intell. Transp. Syst.* **2016**, *20*, 229–243. [CrossRef]
12. Xiao, G.; Lee, J.; Jiang, Q.; Huang, H.; Abdel-Aty, M.; Wang, L. Safety improvements by intelligent connected vehicle technologies: A meta-analysis considering market penetration rates. *Accid. Anal. Prev.* **2021**, *159*, 106234. [CrossRef] [PubMed]
13. Yue, L.; Abdel-Aty, A.M.; Wu, Y.; Farid, A. The practical effectiveness of advanced driver assistance systems at different roadway facilities: System limitation, adoption, and usage. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 3859–3870. [CrossRef]
14. Ye, L.; Yamamoto, T. Evaluating the impact of connected and autonomous vehicles on traffic safety. *Phys. A Stat. Mech. Its Appl.* **2019**, *526*, 121009. [CrossRef]
15. Essa, M.; Sayed, T. Self-learning adaptive traffic signal control for real-time safety optimization. *Accid. Anal. Prev.* **2020**, *146*, 105713. [CrossRef]
16. Morando, M.M.; Tian, Q.; Truong, L.T.; Vu, H.L. Studying the safety impact of autonomous vehicles using simulation-based surrogate safety measures. *J. Adv. Transp.* **2018**, *2018*, 6135183. [CrossRef]
17. Elawady, A.; Abuzwidah, M.; Zeiada, W. The benefits of using connected vehicles system on traffic delay and safety at urban signalized intersections. In Proceedings of the 2022 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 21–24 February 2022; pp. 1–6.
18. El-Hansali, Y.; Farrag, S.; Yasar, A.; Shakshuki, E.; Al-Abri, K. Using surrogate measures to evaluate the safety of autonomous vehicles. *Procedia Comput. Sci.* **2021**, *191*, 151–159. [CrossRef]
19. *ISO 22179*; Intelligent Transport Systems-Full Speed Range Adaptive Cruise Control (FSRA) Systems-Performance Requirements and Test Procedures. ISO: Geneva, Switzerland, 2009.
20. Durrani, U.; Lee, C.; Maoh, H. Calibrating the Wiedemann's vehicle-following model using mixed vehicle-pair interactions. *Transp. Res. Part. C Emerg. Technol.* **2016**, *67*, 227–242. [CrossRef]
21. Bhin, M.; Son, S.; Lee, C. A Study of Effectiveness Evaluation of Internet of Things Traffic Control Devices in Preparation of Autonomous Vehicle Operation Using Simulation. *J. Korean Soc. Transp.* **2021**, *39*, 737–752. [CrossRef]
22. El-Basyouny, K.; Sayed, T. Safety performance functions using traffic conflicts. *Saf. Sci.* **2013**, *51*, 160–164. [CrossRef]
23. Oh, C.; Jo, J.I.; Kim, J.H.; Oh, J.T. Methodology for Evaluating Real-time Rear-end Collision Risks based on Vehicle Trajectory Data Extracted from Video Image Tracking. *J. Korean Soc. Transp.* **2007**, *25*, 173–182.
24. Archer, J. Indicators for Traffic Safety Assessment and Prediction and Their Application in Micro-Simulation Modelling: A Study of Urban and Suburban Intersections. Ph.D. Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2005.
25. Cunto, F.J.C.; Saccomanno, F.F. Microlevel traffic simulation method for assessing crash potential at intersections. In Proceedings of the Transportation Research Board 86th annual Meeting, Washington DC, USA, 21–25 January 2007.
26. Jo, Y.; Jung, A.; Park, H.; Park, J.; Oh, C. Prioritizing driving safety indicators using real-world C-ITS data to identify hazardous freeway sections. *J. Korean Soc. Transp.* **2022**, *40*, 863–878. [CrossRef]
27. Park, J.T.; Lee, S.B. Effects Analysis of Traffic Safety Improvement Program Using Data Mining: Focusing on Urban Area. *J. Transp. Res.* **2011**, *18*, 77–91.
28. Yaghoubzadeh-Bavandpour, A.; Bozorg-Haddad, O.; Zolghadr-Asli, B.; Singh, V.P. Computational intelligence: An introduction. In *Computational Intelligence for Water and Environmental Sciences*, 2nd ed.; Bozorg-Haddad, O., Zolghadr-Asli, B., Eds.; Springer: Singapore, 2022; Volume 1043, pp. 411–427.
29. Patro, S.; Sahu, K.K. Normalization: A preprocessing stage. *arXiv* **2015**, arXiv:1503.06462. [CrossRef]
30. Breiman, L.; Friedman, J.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Chapman and Hall/CRC: New York, NY, USA, 1984; p. 368.