



Article Exploring Factors Influencing Speeding on Rural Roads: A Multivariable Approach

Marija Ferko ^{1,2,3,*}, Ali Pirdavani ^{3,4}, Dario Babić ² and Darko Babić ²

- ¹ Smart View Ltd., Rastočka 8a/2, 10 020 Zagreb, Croatia
- ² Faculty of Transport and Traffic Sciences, University of Zagreb, Vukelićeva 4, 10 000 Zagreb, Croatia; dario.babic@fpz.unizg.hr (D.B.); darko.babic@fpz.unizg.hr (D.B.)
- ³ UHasselt, Faculty of Engineering Technology, Agoralaan, 3590 Diepenbeek, Belgium; ali.pirdavani@uhasselt.be
- ⁴ UHasselt, The Transportation Research Institute (IMOB), Martelarenlaan 42, 3500 Hasselt, Belgium
- * Correspondence: marija.ferko@uhasselt.be

Abstract: Speeding is one of the main contributing factors to road crashes and their severity; therefore, this study aims to investigate the complex dynamics of speeding and uses a multivariable analysis framework to explore the diverse factors contributing to exceeding vehicle speeds on rural roads. The analysis encompasses diverse measured variables from Croatia's secondary road network, including time of day and supplementary data such as average summer daily traffic, roadside characteristics, and settlement location. Measuring locations had varying speed limits ranging from 50 km/h to 90 km/h, with traffic volumes from very low to very high. In this study, modeling of influencing factors on speeding was carried out using conventional and more advanced methods with speeding as a binary dependent variable. Although all models showed accuracy above 74%, their sensitivity (predicting positive cases) was greater than specificity (predicting negative cases). The most significant factors across the models included the speed limit, distance to the nearest intersection, roadway width, and traffic load. The findings highlight the relationship between the variables and speeding cases, providing valuable insights for policymakers and law enforcement in developing measures to improve road safety by determining locations where speeding is expected and planning further measures to reduce the frequency of speeding vehicles.

Keywords: road safety; speeding; prediction modeling; rural roads; spot measurements

1. Introduction

Speeding is known as one of the most common driving violations and one of the leading causes of road crashes, especially ones with severe consequences. According to the European Commission, speeding refers to driving at excessive (exceeding the legal speed limit) or inappropriate speed (driving too fast for the traffic situation, infrastructure, weather conditions, and/or other special circumstances) [1]. In the context of this study, we used the first part of the definition, which refers to the speed limit.

Among the EU countries that monitor levels of speed compliance on urban roads, between 35% and 75% of observed vehicle speeds are above the speed limit, while this share on rural non-motorway roads is between 9% and 63%. When looking at fatalities, in the EU, 37% of fatalities occur on urban roads, and 55% on rural non-motorway roads. The majority of the countries with a significantly lower road crash death rate compared with the EU average (50 deaths per million inhabitants) prescribed a 70 or 80 km/h as a speed limit on rural roads [2]. Additionally, high speed has been recognized as a factor that increases the probability of a crash and increases injury severity [3,4]. A random parameter assessment showed no significant effect of increased speed on the average number of crashes; however, while the model results did not clearly link temporal shifts in parameters to the speed increase, the rise in rollover crash probability in single-vehicle incidents suggests higher



Citation: Ferko, M.; Pirdavani, A.; Babić, D.; Babić, D. Exploring Factors Influencing Speeding on Rural Roads: A Multivariable Approach. *Infrastructures* **2024**, *9*, 222. https://doi.org/10.3390/ infrastructures9120222

Academic Editors: Alessia Ruggeri and Orazio Pellegrino

Received: 30 October 2024 Revised: 25 November 2024 Accepted: 4 December 2024 Published: 6 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). speeds may have contributed to more severe injuries in those crashes [5]. Similarly, higher mean speeds are linked to an increased frequency of severe crashes, while lower speeds are associated with more property damage-only crashes [6].

Speed management is a crucial component of the Safe System approach, with addressing unsafe speeds being the first step to improving a transport system that fails to protect people [7]. Previous research states that operating speed is one of the main factors affecting traffic safety [8]. Some studies primarily focused on drivers' behavior and selected speeds, exploring the relation of road safety with speeds. Different driving styles can be distinguished depending on the category of aggressiveness when driving (from non-aggressive to very aggressive), which means that parameters such as speed, acceleration, and braking will differ from driver to driver [9]. Various factors can influence the speed chosen by a driver on different road sections. Some of the factors mentioned are the psychophysical state of the driver, personal preference, social pressure, vehicle characteristics, and environmental factors such as weather and road characteristics [10]. Further, depending on the part of the road, drivers may misestimate the speed of movement [11]. Hence, it can be concluded that regardless of the statutory speed limit, not all drivers will comply with it.

Obeying the set-up speed limits depends on various factors. Hence, the main objective of this study is to identify these factors. In this research, the emphasis is on linking the main characteristics of the location and type of vehicle with the extent of non-compliance with legal speed limits. Time of day and average summer daily traffic (ASDT) are also considered. To optimally manage speeds, it is necessary to gain insight into the characteristics of traffic flow and operating speeds, both day and nighttime. Furthermore, the aim is to determine which modeling approach fits the objective the best and to point out what factors contribute to or influence the drivers' speeding.

2. Literature Background

Many of the previous studies were based on a behavioral approach. A multilevel logistic regression was utilized on GPS-based data to explore driving behaviors, including speeding [12]. These data were supplemented with drivers' demographics and self-reported speeding behavior, emphasizing the impact of speed zones on speeding behavior. In some studies, a driver behavior questionnaire (DBQ) and the theory of planned behavior model (TPB) were utilized [13]. Regression techniques were applied, and the results show that the components of TPB and DBQ variables can predict drivers' intentions for speeding and overtaking violations; however, it was found that speeding was a more frequent violation than overtaking. A self-assessment questionnaire was used as a data collection tool to investigate speeding behavior in low-visibility conditions [14]. The authors employed structural equation modeling to explore the predictors influencing speed choice under reduced visibility, highlighting driving ability as one of the main factors. In residential areas, critical predictors of speeding intention included affective attitude, descriptive and personal norms, perceived behavioral control, habits, and residential street characteristics [15]. Intention emerged as the sole direct predictor of speeding behavior, with street specifications and facilities significantly influencing it.

The speeding problem among young drivers was recognized, and a qualitative analysis was performed by conducting a focus group experiment including 60 young drivers [16]. Findings revealed that the following factors influence the prevention of speeding: legal consequences, fear of injury, and speed awareness monitors. Factors perceived to contribute to violating speed restrictions included perceiving it as safe, a perceived norm to speed, emotions, and unintentional speeding.

Factors influencing speeding behavior among Indian long-haul truck drivers were explored using data collected through individual interviews and a questionnaire [17]. Further analysis of predicting speeding behavior included conventional modeling (binary logit approach) and more advanced machine learning algorithms (Decision Tree, Random Forest, Adaptive Boosting, and Extreme Gradient Boosting), with random forest showing the best performance. The obtained results from the variable importance plot showed that

the eight important factors influencing speeding behavior are pressured delivery of goods, sleeping and driving duration per day, age and size of the truck, monthly income, driving experience, and the driver's age.

In addition to using self-reported data from questionnaires and focus groups, some studies utilized naturalistic driving data. Safe, unsafe, and safe but potentially dangerous behaviors were identified based on continuous speed data obtained from smartphoneequipped vehicles on tangent and curve road sections [10]. The findings indicate that with increasing age and driving experience, behavior tends to be safe, or drivers tend to drive at low speeds, which can be dangerous for road traffic; however, if the driver lacks habit, the behavior tends to be unsafe. Thus, young people with low driving experience are more inclined toward unsafe driving behavior in terms of speeding. In another study, speeding behavior was examined using naturalistic driving data gathered from field experiments on typical two-lane mountainous rural highways in five provinces of China [18]. A speeding prediction model was developed using random forest, achieving an accuracy of over 85%. Logistic regression was also used to investigate factors influencing speeding behavior, with an accuracy of around 70%. The speeding prediction model identified current acceleration and driving speed as the most critical variables. Visual environment parameters, such as visual curve length in the "near scene" and visual curve curvature in the "middle scene," are followed in importance. Additionally, drivers' age and driving experience significantly affected speeding behavior, and different roadside landscapes were found to lead to distinct speeding behaviors. Speed modeling utilizing data from smartphone sensors was conducted using linear regression to establish models for various road types and times of day, and a general model was developed [19]. Similarly, naturalistic data from smartphones were used to create an overall model applicable to all road environments, along with separate models for urban and rural roads [20]. This study found that trip distance and mobile phone use while driving were statistically significant factors positively correlated with speeding.

Another approach to examining speeding involves collecting spot speed data and utilizing distinct modeling methods, focusing more on infrastructure characteristics. For instance, the investigation of operating speeds on curved rural road sections was carried out using regression models and artificial neural networks (ANNs) [21]. In the initial analysis, regression models were employed to study the relationship between V₈₅ and horizontal alignment as well as roadway factors, with separate predictive models proposed for cars and trucks. The subsequent ANN analysis revealed better predictive performance. The curve radius was the most influential variable affecting V₈₅ for cars, while for trucks, it was the median width. Curve radius emerged as the most significant factor for the car ANN model, followed by median width. For the truck ANN model, the median width was the most influential variable, with the deflection angle coming next. In another study, a Beta regression model was employed to analyze the proportion of speeding using probe speed data, incorporating a grouped random parameter modeling structure to account for varying effects of speed management strategies and other road attributes across different road types (urban and suburban arterials) [22]. A fixed beta model was also developed for comparison. The results indicated that the grouped random parameter model outperformed the fixed beta model, offering better insights into how road features and other factors influence speeding on various road types. Seven variables were significant in both models: AADT, daily transit frequency, asphalt pavement, an indication of low-speed limits, outer shoulder width, and the number of lanes.

In a recent study, speeding frequency was examined using roadside observational surveys along with spatial and temporal attributes of selected locations [23]. A random parameter negative binomial model was developed to analyze speeding behavior, incorporating unobserved heterogeneity across speeding locations, accounting for temporal, road geometric, and built environment factors. The findings highlight significant variability in speeding behavior at different locations. Based on the results, the authors suggest that implementing temporary speed-calming measures during non-peak hours and weekends

could be effective. Additionally, the use of speed humps, rumble strips, or enhanced law enforcement and developing well-connected roads with frequent intersections and traffic signals could also serve as a strategy to discourage speeding. Another study employed a negative binomial statistical model to analyze data from traffic cameras, considering both temporal and environmental factors [24]. The model revealed the significance and likelihood of speeding tendencies by incorporating variables such as year, month, number of lanes, dwelling unit types, school-related factors, and open green space. The results indicated that aggregating speeding data tends to underestimate the influence of these factors. For instance, the impact of posted speed limits was found to be up to twice as significant in disaggregated models compared with aggregated ones. Additionally, speeding violations in summer months were about 25% higher in aggregated models than 40% in disaggregated models. Camera enforcement was associated with a 25% reduction in speeding over four years. Built environment factors showed varied effects, with one-unit dwellings linked to increased speeding, whereas proximity to schools was associated with a speed decrease.

Further, it is also possible to investigate speed data in artificial environments, such as driving simulators. A mathematical model for an intelligent speeding prediction system was developed, categorizing inputs into three types: model inputs and related in-vehicle technology, a mathematical model along with a data processing module, and warning messages combined with a human-machine interface [25]. The system was tested using a driving simulator, and experimental data were utilized to validate models predicting intentional and unintentional speeding, showing no statistically significant time difference between the modeled and experimental results. A study involving a driving simulator investigated drivers' speed compliance behavior in urban and rural environments, employing a Generalized Linear Model with speed difference as the dependent variable and driving environments and driver attributes as predictors [26]. The results indicated better speed compliance in urban settings compared with rural ones. Additionally, drivers' age was positively correlated with speed compliance. Male drivers exhibited lower speed compliance than female drivers, while those with postgraduate or graduate education demonstrated better compliance than those with only secondary education. Driving experience negatively impacted speed compliance, and drivers with prior crash history showed better compliance. Factors such as vehicle type and preferred driving time did not significantly affect speed compliance. Another study used a driving simulator and numerical analysis to examine road infrastructure design and operating speeds for establishing credible speed limits on Italian roads [27]. The research concluded that increasing speed limits, combined with safety countermeasures, could lead to a 23% reduction in crashes.

Previous studies on speeding behavior have employed various methodologies to explore its influencing factors and to develop predictive models. Behavioral approaches have highlighted how intentions and self-reported behavior can predict speeding, including drivers' demographics, self-assessment questionnaires, driver behavior questionnaires (DBQ), and the theory of planned behavior (TPB). Legal consequences, fear of injury, and speed awareness monitors were found to be influential in preventing speeding, while factors such as perceived safety, norms, emotions, and unintentional speeding contributed to speeding violations. Several studies utilized naturalistic driving data to analyze speeding behavior. For instance, based on these data, prediction models with high accuracy highlight acceleration, driving speed, trip distance, mobile phone, and visual environment parameters as significant predictors. Studies employed several techniques, such as regression models, random forests, and artificial neural networks, revealing different infrastructural factors influencing operating speeds. The results showed that factors differ on urban and rural roads.

A limited number of studies have focused on spot speed measurement data, which is essential for capturing real-time speeding behavior at specific locations. This gap is particularly significant in rural road environments, where unique challenges such as varying road conditions, limited enforcement, and distinct driving behaviors, compared with urban areas, complicate speed management. The diversity of speeding behaviors across different locations, influenced by cultural, environmental, and infrastructural factors, underscores the need for targeted research in rural settings. Although some research has addressed infrastructure characteristics, a more comprehensive analysis that integrates traffic flow, road design, and speed management is necessary to develop effective interventions for reducing speeding.

3. Methodology

3.1. Data Collection

Raw data for this paper are based on two years of data collection (for July) and were collected through traffic counters installed on state roads (secondary roads) in the Republic of Croatia. Road authorities use them to control the amount and the heterogeneity of traffic and gain insight into the operating speed at a particular location. Stationary traffic counters with electromagnetic inductive loops were installed in characteristic places (tangent road parts on regularly maintained pavement and with the appropriate lane width) on two-way, two-lane roads. The type of these counters, QLD-6CX nano (vehicle detection accuracy > 99.9%, vehicle classification precision ~97%, speed measurement error: at 50 km/h < 2.8% and at 160 km/h < 8%), can measure the times of individual vehicles passing through the measuring point and their current speeds [28], and which is why they are placed on the road parts where free traffic flow is expected. The counter has a built-in software solution for recognizing and eliminating double counting of vehicles that pass through the counter by occupying two lanes simultaneously. The vehicle is counted in the lane corresponding to the movement's direction.

Selecting tangent road sections for measuring and assessing speeding is justified since the highest and the most uniform speeds are expected on these sections [29]. Moreover, significant acceleration is expected on the tangents following the curve and decelerating on the tangents preceding the curve [30]. The white central line and edge road markings were painted at all measuring locations, emphasizing road alignment. Since no law enforcement cameras are installed at the measuring locations, nor are there regular police patrols, drivers know they cannot be fined based on the counters installed on the roadway. This is another factor that speaks in favor of freely choosing the speed of movement.

Most measuring points on Croatian state roads meet the requirements for speed measurement, allowing drivers to choose their speed freely based on subjective assessments of road conditions and speed limits. Free traffic flow is defined as vehicle movement in one direction, on straight roads, outside intersections, with dry pavement and no speed-restricting factors, where vehicles are far enough apart to drive independently. Ideally, vehicle speeds in free flow should be measured in dry conditions, but the current system cannot guarantee this; however, the impact of wet roadways is minimized by analyzing speed data from July and August [28], which is in accordance with the previous study that showed speeding is most likely to occur in summer months [24].

Data for this research were collected from 20 traffic counters on 15 distinct Croatian state roads, considering five different speed limits. The traffic counters are located on the roads referred to as rural since they are outside the urban area, although some pass through small, non-urban settlements. After filtering out empty rows and rows with unclassified vehicles, the sample consisted of 4,623,852 unique records. After rejecting irregular records and considering that an error could have occurred with the counter, speed was not observed as a continuous variable. Still, the dependent variable was set as "Speeding," distinguishing cases in which the amount of overspeed was >0 km/h as "Yes" and all other cases as "No".

In addition to speeding, additional data were collected through the field inspection and from road authorities: roadside state, whether the location is inside the settlement, posted speed limit, whether overtaking was allowed, the width of the roadway, the distance to the nearest intersection, and the average summer daily traffic (ASDT). The last two were considered especially important, assuming they indicate the characteristic of the traffic flow (i.e., where there is higher traffic and the intersection is close, the speed might decrease). In addition, the roadside was considered, given the presumption that unsafe roadside elements could impact the severity of crash injuries [31]. Based on the above, the measurement locations were carefully chosen to encompass a variety of conditions, such as different speed limits, varying traffic loads, distances to the nearest intersection, and other traffic and infrastructure characteristics.

3.2. Data Analysis

After the previous research examination, several methods were selected to find the bestfitting model to explain each independent variable's importance. All statistical analyses were performed using the statistical software SPSS (version 29.0) and R (version 4.3.3).

Since the primary focus was understanding the factors influencing the decision to speed rather than the degree of speeding, we treated speeding as a binary variable. This approach aligns with legal standards, where any speed above the limit constitutes a violation. While different levels of speeding may have distinct implications, binary classification provides clear interpretability, making it valuable for policymakers and practitioners. It also helps mitigate the impact of measurement inaccuracies in spot speed data.

The aim was to use conventional binary logistic regression since "Speeding" was set as the dependent/outcome variable with a binary result (1—speeding occurred; 0—speeding did not occur); however, given the extensive data sample available in this study, the potential for utilizing some of the machine learning algorithms was noticed.

Regression analysis is a technique used to predict the relationship between a dependent (outcome) variable and one or more independent (predictor) variables using a mathematical equation called a model [32–34]. Since the proposed dependent variable is categorical, logistic regression was preferable over linear regression.

Artificial neural networks mimic the functioning of the human brain through a vast network of interconnected processing nodes. They excel at recognizing patterns. A usual neural network comprises numerous simple, interconnected processing elements known as neurons [35,36]. Each neuron generates a series of real-valued activations for the target outcome. The mathematical model of an artificial neuron includes inputs (X_i), weights (w), bias (b), a summation function (Σ), an activation function (f), and the corresponding output signal (y), as shown in Figure 1.



Figure 1. Scheme an artificial neuron mathematical model [35].

For this paper, a Multi-layer Perceptron (MLP) was used. The Multi-layer Perceptron (MLP) is a supervised learning approach and a feedforward artificial neural network [35,37,38]. A typical MLP is a fully connected network comprising an input layer that receives input data, an output layer that makes a decision or prediction about the input signal, and one or more hidden layers between these two that serve as the network's computational engine. The output of an MLP network is determined using various activation functions, also known as transfer functions, such as ReLU (Rectified Linear Unit), Tanh, Sigmoid, and Softmax. An MLP maintains the structure of a single layer but includes one or more hidden layers, with all nodes connecting between layers. The network trains itself using an algorithm called backpropagation.

Random forest extends a decision tree approach by using an ensemble method that constructs multiple decision trees [39,40]. A decision tree is a supervised learning technique primarily employed for classification tasks, though it can also be used for regression. It starts with a root node representing the initial decision point for splitting the dataset based on a single feature that best separates data into distinct classes. Each split leads to a new decision node, which applies another feature to refine these data into more homogeneous groups or a terminal node that provides the final class prediction. This method of dividing data into binary partitions is known as recursive partitioning. Instead of utilizing all features to build each tree, a random subset of features is used for each decision tree in the forest. Each tree then predicts a class outcome, and the final prediction is determined by a majority vote among all the trees [41]. Thus, a random forest model predicts values or categories by aggregating the results from numerous decision trees [42].

The receiver operating characteristics (ROC) curves were generated for the neural network and random forest models, with the calculation of the area under the curve (AUC). The AUC is a scalar value that assesses the overall performance of a binary classifier. The AUC ranges from 0.5 to 1.0, where 0.5 indicates the performance of a random classifier, and 1.0 corresponds to a perfect classifier with zero classification error. AUC is a robust measure for evaluating score classifiers as it accounts for the entire ROC curve, incorporating all possible classification thresholds. The AUC is calculated by summing the areas of successive trapezoids under the ROC curve [43].

Among other tree-growing algorithms, Chi-square Automatic Interaction Detector (CHAID) was employed as one of the modeling approaches. CHAID accommodates nominal, ordinal, and continuous data, with continuous predictors being categorized into groups with approximately equal numbers of observations [44–47]. After identifying the target (dependent variable), which is the decision or classification tree's root, CHAID divides the root into two or more categories, referred to as parent nodes, and further splits these nodes into child nodes. CHAID analysis constructs a predictive model or tree to identify how variables interact to explain the outcome of a given dependent variable.

4. Results

4.1. Variables' Description

Given that the research aims to determine the influencing factors on speeding, the dependent variable is binary (the vehicle exceeded or did not exceed the speed limit). Of the recorded vehicles, 57.7% drove faster than permitted at a particular measuring location (Table 1).

Dependent Variable		Description	Frequency	Cumulative Percent	Mean	St. Dev.
Speeding	(0) (1)	No Yes	1,956,378 2,667,474	42.3 100	0.58	0.494
		Total	4,623,852			

Table 1. Description of the dependent variable.

Before conducting further analysis, a Variance Inflation Factor (VIF) was investigated to check the multicollinearity between independent variables. Since all considered variables' VIFs were <3, all of them were included, as shown below.

Seven categorical variables were included in the further analysis. Table 2 shows the characteristics and frequencies of each categorical variable, with their encoding values.

Variable		Description	Frequency	Cumulative Percent	Mean	St. Dev.
	(0)	Motorcycles	85,351	1.8		
	(1)	Passenger cars with or without trailer	3,915,205	86.5		
Vehicle group	(2)	Vans with or without trailer	284,847	92.7	1.2	0.613
	(3)	Cargo vehicles	300,090	99.2		
	(4)	Buses	38,359	100		
	(1)	Shoulder/Maintained	3,034,433	65.6		
Roadside state	(2)	No shoulder/Not maintained	1,094,420	89.3	1.66	1.228
	(5)	Open drain canal	494,999	100		
T (1) ((0)	No	2,374,668	51.4	0.40	0 5
In settlement	(1)	Yes	2,249,184	100	0.49	0.5
Overtaking	(0)	No	2,991,094	64.7	0.25	0.470
allowed	(1)	Yes	1,632,758	100	0.35	0.478
	(1)	Monday	771,045	16.7		
	(2)	Tuesday	696,168	31.7		2.065
	(3)	Wednesday	553,363	43.7		
Day of the week	(4)	Thursday	565,909	55.9	3.96	
	(5)	Friday	643,012	69.8		
	(6)	Saturday	750,335	86.1		
	(7)	Sunday	644,020	100		
	(1)	Daytime	3,792,118	82.0		
Part of the day	(2)	Twilight	314,927	88.8	1.05	0.01
r art or the day	(3)	Nighttime	268,968	94.6	1.35	0.815
	(4)	Dawn	247,839	100		
		50	1,324,261	28.6		
		60	1,499,306	61.1		
Speed limit		70	321,978	68.0	65.81	14.454
-		80	745,813	84.2		
		90	732,494	100		

Table 2. Descri	ption of the c	ategorical ex	planatory v	variables in	cluded ir	the analysis.

Table 3 describes the continuous variables, where "Width across the roadway" implies the overall width of traffic lanes and nearby roadside, and "Average Summer Daily Traffic" (ASDT) implies average traffic in summer months (July and August). Finally, "Distance to the closest intersection" considers the distance to the intersection nearest to the measuring point.

Table 3. Description of the continuous variables included in the analysis.

	Mean	Std. Dev.	Min.	Max.
Width across the roadway (m)	6.868	0.534	5.00	8.00
Average Summer Daily Traffic—ASDT (vehicles)	10,577.51	10,577.51	2867.00	21,562.00
Distance to the closest intersection (m)	215.720	149.19	85.00	850.00

The variables were chosen based on their potential relevance and availability. The variables depicting the state of the infrastructure are "In settlement," "Speed limit," "Road-side state," and "Width across the roadway." The variable "Roadside state" was included to capture the physical condition and characteristics of the roadside environment, potentially influencing driver behavior and safety outcomes. A well-maintained shoulder with drainage channels, curbs, and sand covering (Shoulder/Maintained) typically enhances safety. In contrast, roads lacking a shoulder or having poorly maintained edges, such as grassy areas without barriers or curbs (No shoulder/Not maintained), may increase risk and uncertainty for drivers. Similarly, an open water canal may represent certain risks and influence driving behavior.

Furthermore, "Distance to the closest intersection," ASDT, and "Overtaking allowed" are assumed to describe the traffic flow characteristics. More precisely, "Overtaking allowed" (Yes or No) refers to road sections where overtaking is legally permitted and the center line is dashed. "Day of the week" and "Part of the day" describe the time component that could influence the speed selection. "Vehicle group" is a variable that clarifies the observed vehicles' technical characteristics, assuming that the vehicles with the most favorable power/mass ratio (motorcycles) will also be the fastest, i.e., the most likely to overspeed.

4.2. Descriptive Statistics on Speeding

The vehicles were automatically classified into ten groups based on length; however, due to the minor vehicle frequency in some groups, in further analysis, the vehicles were finally sorted into five groups (Table 4). As stated in Table 1, 57.7% (N = 2,667,852) of vehicle speed records were above the legal speed limit; therefore, speeding was only compared between vehicles that were above the speed limit. The test of Homogeneity of Variances confirmed that the variances among the groups are significantly different. Hence, the significance of differences in the means between presented vehicle groups was tested using the Welch test, which confirmed statistically significant differences in the amount of speeding among the groups at the 0.05 level.

Table 4. Descriptives on the amount of speeding by vehicle groups (km/h).

Vehicle Group	Ν	Mean	Std. Dev.	Std. Error
Motorcycles	58,237	24.11	19.41	0.08
Passenger cars with or without trailer	2,289,814	13.22	10.985	0.007
Vans with or without trailer	158,203	13.15	11.16	0.028
Cargo vehicles	142,664	11.39	8.836	0.023
Buses	18,556	10.03	7.616	0.056
Total	2,667,474	13.34	11.252	0.007

Based on the Games–Howell post hoc test results, passenger cars, and vans are the only groups with insignificant mean differences in speeding. Motorcycles proved to be the fastest form of travel. The average amount of speeding is more than 10 km/h higher for motorcycles than other vehicle groups. The results are presented in Table 5.

Table 5. Multiple comparisons on speeding mean difference.

(I) Vehicle Group	(I) Vehicle Group (J) Vehicle Group		Std. Error	Sig.
Motorcycles	Motorcycles Passenger cars with or without trailer		0.081	< 0.001
-	Vans with or without trailer	10.959 *	0.085	< 0.001
	Cargo vehicles	12.717 *	0.084	< 0.001
	Buses	14.083 *	0.098	< 0.001
Passenger cars with or without	Vans with or without trailer	0.075	0.029	0.072
trailer	Cargo vehicles	1.833 *	0.024	< 0.001
	Buses	3.199 *	0.056	< 0.001
Vans with or without trailer	Cargo vehicles	1.757 *	0.037	< 0.001
	Buses	3.124 *	0.063	< 0.001
Cargo vehicles	Buses	1.367 *	0.061	< 0.001

* The mean difference is significant at the 0.05 level.

4.3. Binary Logit Model

Binary logistic regression with the enter method was employed to examine the relevant variables in predicting the occurrence of speeding. The Nagelkerke R Square of 0.380 suggests that the model explains approximately 38% of the variance in the dependent variable (e.g., speeding). At the same time, the Cox and Snell R Square is slightly lower (0.283), indicating a reasonable but not perfect fit. Table 6 illustrates the classification of correctly and incorrectly predicted values (with a cut-off value of 0.5). The model performs very well in predicting "Yes" cases with a high sensitivity of 89.3% (the model's ability to identify speeding cases correctly), while that is not the case for negative outcomes with a specificity of 55.0% (model's accuracy in identifying non-speeding cases).

			Predicte	d
		Spee	eding	Percent Correct
Obser	ved	No	Yes	
Croading	No	1,075,436	880,942	55.0%
Speeding	Yes	284,357	2,383,117	89.3%
Overall Per	centage			74.8%

Table 6. Classification table (binary logistic regression).

The model indicates that all factors significantly influence speeding, with speed limits being the most impactful factor, while ASDT has no meaningful impact on the likelihood of speeding (Table 7). The negative coefficients indicate a strong inverse relationship between the speed limit and the likelihood of speeding, with higher speed limits strongly predicting the outcome "No." Further, as expected, greater distances to intersections increase the odds of speeding (B = 0.003, Exp(B) = 1.003), while locations within settlements have significantly lower odds of speeding compared with those outside settlements (B = -1.148, Exp(B) = 0.317). The width of the roadway has a negative association with speeding, indicating that wider roadways are associated with lower speeding odds (B = -0.295, Exp(B) = 0.744), while allowed overtaking increases the odds of speeding (B = 0.296, Exp(B) = 1.345). The odds of speeding vary by day of the week and the time of the day, with the highest odds observed on Sundays (B = 0.330, Exp(B) = 1.392) and during dawn (B = 0.937, Exp(B) = 2.552). Motorcyclists are significantly more likely to speed than other vehicle types, with passenger cars, vans, buses, and cargo vehicles all exhibiting lower odds.

 Table 7. Regression coefficients, odds ratios, and confidence intervals for variables predicting speeding likelihood.

Variables in the Equation	п	ст	TA7 1 1	16	Sia	Exer(P)	95% C.I. for EXP(B)	
variables in the Equation	D 5.E.		Wald	đf	51g.	схр(б)	Lower	Upper
ASDT	0.000	0.000	372.296	1	< 0.000	1.000	1.000	1.000
Distance to the closest intersection	0.003	0.000	50,940.542	1	< 0.000	1.003	1.003	1.003
Width across the roadway	-0.295	0.003	9260.942	1	< 0.000	0.744	0.740	0.749
Shoulder/Maintained			42,580.405	2	< 0.000			
No shoulder/Not maintained	0.853	0.007	15,876.513	1	< 0.000	2.347	2.316	2.379
Open drain canal	1.753	0.009	42,173.050	1	< 0.000	5.772	5.677	5.870
In settlement (Yes)	-1.148	0.004	73,416.343	1	< 0.000	0.317	0.315	0.320
Speed limit (50 km/h)			712,758.350	4	< 0.000			
Speed limit (60 km/h)	-0.247	0.006	1636.172	1	< 0.000	0.781	0.772	0.791
Speed limit (70 km/h)	-2.832	0.009	107,559.285	1	< 0.000	0.059	0.058	0.060
Speed limit (80 km/h)	-2.556	0.005	219,753.026	1	< 0.000	0.078	0.077	0.078
Speed limit (90 km/h)	-4.578	0.006	582,896.093	1	< 0.000	0.010	0.010	0.010
Overtaking allowed (Yes)	0.296	0.004	5063.941	1	< 0.000	1.345	1.334	1.356

Variables in the Equation	р	сг	Wald	46	Sig	$E_{YD}(\mathbf{R})$	95% C.I. 1	for EXP(B)
variables in the Equation	D	5. E.	vvalu	ai	Jig.	Lxp(D)	Lower	Upper
Monday			7726.487	6	< 0.000			
Tuesday	0.070	0.004	315.058	1	< 0.000	1.073	1.065	1.081
Wednesday	0.039	0.004	86.571	1	< 0.000	1.040	1.032	1.049
Thursday	0.067	0.004	257.313	1	< 0.000	1.070	1.061	1.079
Friday	0.042	0.004	106.092	1	< 0.000	1.043	1.034	1.051
Saturday	0.067	0.004	290.725	1	< 0.000	1.069	1.061	1.077
Sunday	0.330	0.004	6397.624	1	< 0.000	1.392	1.380	1.403
Daytime			36,411.946	3	< 0.000			
Twilight	0.102	0.004	526.301	1	< 0.000	1.107	1.098	1.117
Night-time	0.498	0.005	10,173.651	1	< 0.000	1.645	1.629	1.661
Dawn	0.937	0.006	28,154.951	1	< 0.000	2.552	2.524	2.580
Motorcycles			11,493.604	4	< 0.000			
Passenger cars with or without trailer	-0.644	0.009	5134.531	1	< 0.000	0.525	0.516	0.535
Vans with or without trailer	-0.683	0.010	4710.769	1	< 0.000	0.505	0.495	0.515
Cargo vehicles	-1.004	0.010	10,115.399	1	< 0.000	0.366	0.359	0.374
Buses	-0.670	0.015	1971.906	1	< 0.000	0.511	0.497	0.527
Constant	3.909	0.027	21,414.320	1	< 0.000	49.848		

Table 7. Cont.

4.4. Neural Network Model

The initial dataset is divided into a training set (70%) and a test set (30%) to perform neural network modeling. The input layer comprises 31 units (excluding the bias unit), while the hidden layer contains 10. The hyperbolic tangent activation function was used for the hidden layer, while the Sigmoid function was used for the output layer, with the sum of squares used as the error function. Batch was used as a type of training, with maximum training epochs computed automatically. The model's performance is evaluated using classification metrics on both the training and testing datasets (Table 8).

Table 8. Classification table (neural network).

			Predicted	
Sample	Observed	No	Yes	Percent Correct
Training	No	899,402	470,605	65.6%
-	Yes	279,665	1,587,025	85.0%
	Overall Percent			76.8%
Testing	No	393,561	192,810	67.1%
	Yes	131,495	669,289	83.6%
	Overall Percent			76.6%

The model demonstrated an overall accuracy of 76.8% on these training data, indicating that it effectively learned the patterns in these data. The overall accuracy of these testing data was 76.6%, slightly lower than the accuracy of these training data. This suggests the model generalizes well to unseen data and does not suffer from overfitting. Sensitivity, with 85.0% for the training set and 83.6% for the testing set, reflects the model's strong performance in detecting instances of speeding. Specificity was comparatively lower, at 65.6% for the training set and 67.1% for the testing set. A model with two hidden layers was created for control, but the performance was not better than the initial one-hidden layer model. The factor that proved to be the most influential is the speed limit, followed by the distance to the closest intersection, roadway width, ASDT, and vehicle group.

The AUC for the neural network model was 0.840 for both training and testing datasets, indicating that the model has good discriminative ability (Figure 2). An AUC of 0.840 means an 84% chance that the model will correctly distinguish between a randomly chosen positive



instance (speeding) and a randomly chosen negative instance (no speeding). These results further support the effectiveness of the neural network model in predicting speeding.

Figure 2. ROC curves for training and testing one-hidden layer neural network model.

4.5. Chi-Squared Automatic Interaction Detector (CHAID)

The Chi-squared Automatic Interaction Detector (CHAID) model was employed to analyze factors influencing speeding behavior. The model, validated through a split sample approach (70% training set, 30% test set), identified six significant predictors of speeding: distance to the closest intersection, ASDT, vehicle group, time of the day, day of the week, and width across the roadway. With tree depths of 3 and 160 nodes, including 115 terminal nodes, the model provides detailed insights into how these factors influence speeding behaviors across different contexts.

The risk estimates for speeding obtained from the CHAID model are 0.231 and 0.232, with a standard error of 0.000 for both datasets. These estimates reflect the stability and consistency of the model's predictions across training and test datasets. In classification results for speeding, the training dataset shows an overall correct prediction percentage of 76.9% (Table 9). The sensitivity was high, with values of 85.2% for both the training and testing sets. This suggests that the model is robust in detecting speeding cases. The specificity was lower, with 65.5% on the training set and 65.3% on the testing set, indicating that the model has a moderate error rate in classifying non-speeding instances. These metrics indicate the model's effectiveness in classifying instances of speeding based on the specified variables and the CHAID growing method.

			Predicted	
Sample	Observed	No	Yes	Percent Correct
Training	No	896,729	473,294	65.5%
0	Yes	275,399	1,590,846	85.2%
	Overall Percent			76.9%
Testing	No	383,166	203,189	65.3%
0	Yes	118,407	682,822	85.2%
	Overall Percent			76.8%

Table 9. Classification table for the CHAID model.

The CHAID model demonstrates predictive solid performance, as evidenced by AUC values of 0.838 for the training set and 0.839 for the testing set (Figure 3). These AUC values indicate that the model can discriminate between speeding and non-speeding cases, performing consistently well on both the training and testing datasets. The close similarity in AUC values suggests that the model generalizes effectively to new data, maintaining its accuracy and robustness outside the initial training environment.



Figure 3. ROC curves for training and testing of the CHAID model.

4.6. Random Forest Model

The random forest model trained in this study is capable of predicting speeding behavior appropriately. Using a dataset of 3,236,697 observations (70% training set, 30% test set), the model comprising 500 trees achieved an out-of-bag (OOB) prediction error (Brier score) of 0.1597. On the training set, the model achieved an accuracy of 76.8% (95% CI: 76.74–76.83%), with a sensitivity of 84.4% and a specificity of 66.35%. On the test set, the model maintained a similar accuracy of 76.8% (95% CI: 76.73–76.87%), with a sensitivity of 66.3%. The results are presented in Table 10. These metrics demonstrate the model's consistency and robustness across different datasets. Regarding predictors' importance, the ones with the highest importance are speed limit, distance to the closest intersection, ASDT, and roadway width.

			Predicted	
Sample	Observed	No	Yes	Percent Correct
Training	No	908,674	290,506	66.4%
Ū.	Yes	460,791	1,576,726	84.4%
	Overall Percent			76.8%
Testing	No	389,311	124,255	66.3%
-	Yes	197,602	675,987	84.5%
	Overall Percent			76.8%

Table 10. Classification table for random forest model.

The random forest model demonstrated strong discriminative power, achieving an AUC of approximately 0.840 and 0.841 for the training and the testing dataset, respectively (Figure 4). This high AUC indicates excellent performance in distinguishing between speeding and non-speeding instances.



Figure 4. ROC curves for training and testing random forest model.

5. Discussion

5.1. Average Values of Speeding Concerning Vehicle Groups

The classification of vehicles into five groups, based on length, was a pragmatic approach to ensure sufficient sample sizes of each group for meaningful analysis. This consolidation likely enhanced the robustness of subsequent statistical tests by mitigating the issue of minor vehicle frequency, which could lead to unreliable estimates and conclusions. Given that the specified groups of vehicles differ significantly in their driving-dynamic characteristics, the amounts of speeding for them were observed before analyzing the factors influencing speeding. Furthermore, the groups were compared with the basic assumption that the highest speeding was recorded among motorcyclists.

With a mean speeding amount of 24.11 km/h, motorcycles are the fastest vehicles, significantly surpassing other vehicle groups by more than 10 km/h. This result indicates a higher propensity for speeding among motorcyclists, highlighting them as a critical category. Further, as expected, passenger cars and vans showed similar speeding behavior, with mean speeds of 13.22 km/h and 13.15 km/h, respectively. Finally, cargo vehicles and buses exhibit the lowest speeding amounts, with means of 11.39 km/h and 10.03 km/h, respectively. This may reflect the professional nature of drivers in these categories or inherent vehicle characteristics that limit speed.

Earlier studies confirmed that motorcyclists are more likely to exceed the posted speed limit compared with passenger cars and other motor vehicles [48–51]. Furthermore, the results of this study are consistent with previous research, stating that motorcyclists are more likely to speed on rural roads due to motorcycle maneuverability and riders' enjoying fast riding [52]. Some studies point out that there is also a significant difference in speed between different types of motorcycles (e.g., sports and enduro), which can be an implication for future research [50].

5.2. Speeding Prediction Models

This study applied several modeling techniques to examine the factors influencing speeding behavior and evaluate their predictive performance. The methods compared include binary logistic regression, neural network, CHAID classification tree, and random forest. Each model has distinct characteristics, strengths, and limitations, as discussed below.

All models showed a more accurate prediction of "Yes" cases, with a sensitivity greater than 84%. All approaches showed decent results when observing the models' accuracy (>74%).

Table 11 compares the performance of four classification models across six key metrics: sensitivity, specificity, accuracy, precision, F1-Score, and Cohen's Kappa. The binary logistic model exhibits the highest sensitivity at 89.30%, demonstrating strong performance in identifying true positive cases (e.g., identifying "Yes" instances); however, this is paired with the lowest specificity (55.00%), indicating weaker performance in correctly identifying true negative cases (e.g., identifying "No" instances). On the other hand, the machine learning models show a more balanced performance. Sensitivities for the neural network, CHAID tree, and random forest models hover around 85%, while specificities range from 65.41% to 66.14%. In terms of accuracy, all three machine learning models perform similarly, with the CHAID tree model achieving the highest accuracy at 76.85%, followed closely by the neural network (76.82%) and random forest (76.78%).

Precision and F1-Score further confirm the balanced performance of the machine learning models, with the random forest model showing the highest precision (77.31%) and the CHAID tree model achieving the best F1-Score (80.94%). Cohen's Kappa values, which measure overall agreement between predicted and observed classifications, are also higher for the machine learning models, with the neural network model slightly outperforming the others (0.517), suggesting a better overall predictive quality compared with the binary logistic model (0.459).

Model	Sensitivity	Specificity	Accuracy	Precision	F1-Score	Cohen's Kappa	The Most Significant Factors
Binary logistic	89.30%	55.00%	74.80%	73.04%	80.36%	0.459	Speed limit, distance to intersection, roadway width, roadside state, etc.
Neural network	85.02%	65.65%	76.82%	77.13%	80.88%	0.517	Speed limit, distance to intersection, roadway width, ASDT, vehicle group
CHAID tree	85.25%	65.41%	76.85%	77.05%	80.94%	0.516	Distance to intersection, ASDT, vehicle group, time of the day, day of the week, roadway width
Random forest	84.59%	66.14%	76.78%	77.31%	80.78%	0.516	Speed limit, distance to intersection, ASDT, roadway width, roadside state

Table 11. Summary of models' performance.

Several aspects can explain the similar performance across the models. First, the data quality and structure likely offer transparent relationships between variables, making it easier for simpler models, such as logistic regression, to perform well. Moreover, the complexity of the problem may not demand highly sophisticated models, so machine learning methods such as neural networks or decision trees do not provide significantly better results. Furthermore, without extensive hyperparameter tuning, the more complex models may not fully exploit their potential, resulting in only marginally better performances than more straightforward approaches. Last, the large sample size and the absence of all potential causal parameters in the dataset can influence the models' similar performance. A large dataset often provides enough information for different models to achieve stable and consistent predictions, reducing variability in performance. This can lead to even simpler models capturing essential patterns effectively. Additionally, while the models' performance is satisfying, there is a potential for including some more causal factors. Above all, the similarity of results between training and test sets generally indicates a good model performance and generalizability. The abovementioned explains why the models display similar sensitivity, specificity, and overall accuracy levels.

The table reveals that while speed limit and distance to the closest intersection are consistently significant predictors across multiple models, the importance of other factors such as ASDT, roadway width, and time-related variables varies depending on the model used. These variations suggest that the choice of model should align with the specific needs of the analysis, whether prioritizing simplicity or a more comprehensive, multifactorial approach; therefore, it will be essential for the authorities to evaluate these models against their specific datasets to determine the most suitable one for implementation.

The association between speeding and several explanatory variables was constructed with several modeling approaches, which aligns with the previous research. Given that the performance of all models is similar, binary logistic regression is the basis for further discussion, showing strong sensitivity (89.30%). By utilizing binary logistic regression, the findings are interpretable (outputs in the form of odds ratios (Exp(B), previously shown in Table 7), statistically sound, and actionable, contributing meaningfully to the understanding of speeding behavior and its underlying determinants.

Higher speed limits are significantly associated with lower odds of speeding. This implies that as speed limits increase, drivers are less likely to exceed them. This trend may be attributed to factors such as enforcement practices and adjustments in driver behavior. In addition, it might imply that drivers feel bored or too slow at lower speeds or even think that posted speed limits do not comply with the road design. The difference between the road design and the posted speed limit can also lead to speeding [53]. Roadway characteristics substantially impact speed selection behavior, leading drivers who usually

tend to drive fast to increase their speeds more than slower drivers when opportunities to drive faster are present [54,55]. Another study shows that drivers justify speeding by saying the speed limits are too low, the road conditions allow higher speed, or it is a habit [56]. Kutela et al. concluded that the analysis of speed limits reveals an increase in the likelihood of speeding as the speed limits increase; however, the design difference should be taken into account [24]. Similarly, Cai et al. (2021) confirmed that drivers are more likely to exceed the speed limit when the speed limit is low [22]. Other studies also confirmed that speeding is more likely to occur on low-speed limit roads [57], concluding that drivers choose their operating speed based on the other drivers' speed [58]; therefore, the speed distribution should serve as the foundation for determining suggested speed limits, with the final recommended value considering roadway type, context, safety performance, and other relevant characteristics [59].

Research shows that a wider shoulder is associated with higher speeds, while narrower lanes encourage a speed reduction [22]. Contrary to expectations, this study revealed that a wider roadway is associated with a lower likelihood of speeding; however, this must be considered cautiously since the difference between road width was relatively small (St. Dev. = 0.534 m). Interestingly, the roadside without a well-maintained shoulder and safety barrier is a more significant indicator of speeding than where the shoulder has a curb and/or a safety barrier. That may indicate that, in addition to the safety function, protective barriers and curbs impact the drivers' perception, i.e., drivers are more careful not to hit the roadside object.

Similar to previous research, increased traffic can be connected to reduced speed [60,61]. The fact that the importance of variables such as distance to the closest intersection and ASDT has been shown indicates that traffic flow characteristics influence speed selection. In other words, the influence of other vehicles and increased driver's caution when approaching or passing through an intersection is possible. This is consistent with previous studies showing that the least "smooth driving" is expected in urban areas (i.e., cities) [62]. Since the increased share of speeding vehicles is expected outside urban areas, the selection of measurement locations in this research is justified when discussing factors potentially affecting speeding. Further, it can be expected that speeding will occur outside the settlements, even in rural areas, as well as on road sections where overtaking is allowed, which this study confirmed.

Weekends, notably Sundays, show increased odds of speeding, particularly in combination with nighttime and dawn driving, which is associated with a higher likelihood of speeding. The finding is particularly worrying since crashes, especially single-crashes, often occur during nighttime, at weekends, and under low traffic volume [63]. On the other hand, some research shows that it is more likely that speeding will occur during evening and midday weekend hours [64]. This difference may indicate that it is necessary to consider the geographical and cultural components when modeling speeding behavior or discussing transportation in general since the dynamics of people's lives and habits can vary.

The analysis within our research indicated significant differences in speeding behavior across various vehicle groups. The results suggest motorcycle riders are more likely to be involved in speeding incidents than all other drivers. Specifically, as vehicle type changes from motorcycles to other groups, the likelihood of speeding decreases significantly. This expected result underscores the distinct driving behaviors and speed compliance levels associated with different vehicle types, highlighting motorcycles as a more prominent risk group for speeding-related incidents. Previous studies confirm motorcycle riders are more prone to speeding than other drivers [52]. The above is particularly worrying considering that research shows that excessive speed significantly affects the occurrence of severe injuries and fatalities among motorcyclists [65,66].

While most of this study's findings align with previous research, which confirms the role of factors such as speed limits in influencing speeding behavior, there were also some notable variations. These diverse indicators suggest that geographical and sociological

aspects may play a significant role in shaping speeding tendencies. For example, the analysis revealed that temporal factors, such as the day of the week and time of day, significantly affect speeding likelihood. This indicates that drivers' behavior might be influenced by social or cultural norms tied to specific times, such as weekend driving habits or night-time driving patterns. This variation could also indicate regional differences in driving culture, law enforcement practices, or public awareness of road safety, which more standardized driver behavior models might need to capture fully.

5.3. Limitations and Future Implications

This paper provides valuable insights into the factors influencing speeding behavior and choosing the appropriate modeling approach; however, several limitations should be acknowledged. This study did not account for individual driver characteristics such as age, gender, driving experience, or driving history. This means that only the vehicle and road characteristics were taken into account; however, this information can be unavailable to road authorities since they do not know the drivers' characteristics when posting speed limits. Still, according to some researchers, driver attitude and other driver features strongly correlate with obeying speed limits [67,68]. The problem is solved to some extent by using groups of vehicles since a specific group of drivers is often associated with some typical behavior in the literature.

Further, the speed measurements were point-based, capturing speed at specific locations rather than over a continuous stretch of road. This method may not fully represent the overall driving behavior and could miss variations in speed between measurement points.

Finally, the dataset in this study represents summer measurements only. Although this ensures uniform measurement conditions, future research could include data from other seasons when conditions differ (e.g., snow, fog, etc.). On the other hand, favorable weather conditions are one of the prerequisites for free traffic flow, which is a vital assumption when inspecting speeding.

According to the presented results, the focus of future research could be more specific, for example, more detailed monitoring and data collection on motorcyclists' movements (e.g., naturalistic data collection). Furthermore, it could be fruitful to separately observe weekends as a perilous period from the point of view of driving speed. Another potential approach is to observe speeding by class, based on the amount of speeding, and to observe cases inside and outside the settlement separately since the results show a higher possibility of speeding outside the inhabited settlement; however, the sample size presented in this study is one of its key advantages, enhancing the reliability and generalizability of the findings, as it reduces the likelihood of sampling bias and increases the precision of estimates. Furthermore, the sample size allows for detecting subtle relationships between the predictors and the likelihood of speeding and for a more granular analysis of subgroups, such as different vehicle types. In this context, the presented research is a base point for directing further examination, employing a more analytical approach.

Based on everything presented, there is a significant potential for further expansion of individual models so that they can ultimately be applied. By applying the results of those models, road authorities and law enforcement offices could precisely predict locations for the implementation of traffic calming measures or the installation of speed cameras. This can contribute to lower costs and increased traffic safety.

6. Conclusions

This study aimed to identify the key factors influencing speeding behavior on Croatian state roads using various statistical and machine-learning methods. By analyzing data collected from traffic counters on rural roads over two years, the research explored the impact of vehicle type, road characteristics, time of day, and other variables on speeding occurrences. Among the models, the random forest demonstrated superior performance, achieving an accuracy of 76.8% and a robust discriminative power, indicating its effective-

ness in predicting speeding behavior; however, binary logistic regression could be one of the most useful models because of its favorable interpretability. The findings consistently highlighted the speed limit as the most significant predictor of speeding, with lower speed limits strongly associated with increased speeding likelihood. The distance to the closest intersection and the width of the roadway also emerged as influential factors. Vehicle type, time of day, and day of the week further contributed to speeding behavior, with motorcycles exhibiting the highest average speeding and speeding more likely to occur during nighttime and weekends.

This study uniquely contributes to road safety research by comprehensively analyzing speeding factors on rural roads in Croatia, a region underrepresented in previous studies. Combining traditional and advanced modeling techniques offers a more robust analytical approach and examines a distinctive set of factors, including some rarely explored in this context. These insights provide practical recommendations for targeted interventions, enhancing the understanding of speeding behavior in specific geographical and infrastructural settings. Overall, this research contributes valuable insights into the multifaceted factors influencing speeding behavior, thereby supporting the development of targeted interventions and policies aimed at enhancing road safety. Some general suggestions for road authorities can be provided:

- Utilize Intelligent Transportation Systems (ITS): Implement advanced traffic management systems that leverage real-time data and analytics to optimize traffic flow and enhance enforcement measures, such as strategically placed speed cameras.
- Focus on Eco-Friendly Road Design: Implement road-narrowing techniques at transition zones, such as residential areas and intersections, to effectively reduce vehicle speeds. By designing roadways that physically guide drivers to slow down, these measures enhance safety while also promoting eco-friendly practices through the use of sustainable materials and designs that minimize environmental impact.
- Enhance Road Visibility and Safety: Improve the visibility and frequency of speed limit signs by incorporating digital displays that adapt to real-time conditions, ensuring drivers are consistently informed of speed regulations. Additionally, perceptual road markings should be utilized to create a visual narrowing effect, which can psychologically encourage drivers to reduce their speed. This combination of advanced signage and perceptual techniques not only enhances visibility but also significantly improves safety by promoting more cautious driving behavior in critical areas.
- **Data-Driven Speed Limit Adjustments:** Regularly review and establish realistic speed limits based on comprehensive analyses of road types, traffic volumes, and crash histories, ensuring that limits are both safe and enforceable.
- **Tech-Enhanced Educational Campaigns:** Initiate campaigns that leverage digital platforms to inform drivers about the significance of adhering to speed limits and the dangers linked to speeding, thereby promoting a culture of safety on the roads.

Further, based on the results presented in this study, some more specific and applicable suggestions for road authorities, policymakers, and law enforcement can be proposed:

- **Speed Limit Review:** Conduct a thorough revision of posted speed limits by analyzing operational speeds at critical locations, assessing the current condition of the infrastructure, and considering additional contextual factors such as land use, pedestrian activity, and road geometry. This approach helps ensure that speed limits are appropriate for the environment and encourages safer driving behavior.
- Road Safety Equipment Installation: Strategically install appropriate road safety equipment, such as crash barriers and guardrails, in high-risk areas to reduce the severity of crashes. These physical measures not only protect drivers but also act as visual cues, encouraging speed reduction and caution in known crash-prone zones.
- Increased Police Presence: Implement more frequent and targeted police patrols, especially during high-risk periods such as weekends and holidays when speeding and dangerous driving behaviors are more prevalent. A visible law enforcement presence is a deterrent to speeding and reckless driving.

- **Motorcycle Safety Focus:** Pay special attention to areas with high motorcycle traffic by identifying potential hazards and launching targeted safety campaigns. Educate motorcyclists and other road users about safe practices and implement infrastructure improvements to enhance motorcycle safety.
- **Traffic Surveillance Enhancement:** Install traffic monitoring cameras at locations where errant drivers, particularly motorcyclists, are frequently observed. These cameras help enforce traffic laws in cases where it is challenging for police to apprehend offenders, such as motorcyclists fleeing from officers or concealing license plates.
- **Crash and Speeding Monitoring:** Continuously track and analyze data related to traffic crashes caused by speeding, including the severity of injuries and fatalities. This ongoing evaluation can inform future road safety measures and adjustments to enforcement strategies, helping to reduce the incidence of speed-related crashes over time.

Implementing measures that address the identified factors, such as integrating advanced road design improvements and stricter enforcement of speed limits, can significantly reduce speeding incidents and associated risks. Through a comprehensive approach that combines enforcement, innovative design, and technology, we can create safer road environments that minimize the dangers of speeding.

Author Contributions: Conceptualization, M.F. and A.P.; methodology, M.F.; validation, M.F., A.P. and D.B. (Dario Babić); formal analysis, M.F.; investigation, M.F.; data curation, M.F.; writing—original draft preparation, M.F.; writing—review and editing, A.P., D.B. (Dario Babić) and D.B. (Darko Babić); visualization, M.F.; supervision, A.P.; project administration, M.F.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Special Research Fund (BOF) of Hasselt University with the BOF number "BOF21BL03".

Data Availability Statement: Data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: Marija Ferko is employed by Smart View Ltd. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- 1. Aarts, L.; van Schagen, I. Driving speed and the risk of road crashes: A review. *Accid. Anal. Prev.* 2006, *38*, 215–224. [CrossRef] [PubMed]
- Van den Berghe, W. Road Safety Thematic Report—Speeding; European Commission: Brussels, Belgium, 2021; Available online: https://road-safety.transport.ec.europa.eu/system/files/2021-07/road_safety_thematic_report_speeding.pdf (accessed on 14 June 2024).
- 3. Adminaité-Fodor, D.; Jost, G. *Reducing Speeding in Europe*; European Transport Safety Council: Brussels, Belgium, 2019; Available online: https://www.etsc.eu/pin (accessed on 14 June 2024).
- 4. Islam, M.; Mannering, F. The role of gender and temporal instability in driver-injury severities in crashes caused by speeds too fast for conditions. *Accid. Anal. Prev.* **2021**, *153*, 106039. [CrossRef] [PubMed]
- 5. Alnawmasi, N.; Mannering, F. The impact of higher speed limits on the frequency and severity of freeway crashes: Accounting for temporal shifts and unobserved heterogeneity. *Anal. Methods Accid. Res.* **2022**, *34*, 100205. [CrossRef]
- 6. Nassiri, H.; Mohammadpour, S.I. Investigating speed-safety association: Considering the unobserved heterogeneity and human factors mediation effects. *PLoS ONE* **2023**, *18*, e0281951. [CrossRef] [PubMed]
- 7. *The Safe System Approach in Action*; International Transport Forum: Paris, France, 2022; Available online: https://www.itf-oecd. org/sites/default/files/docs/safe-system-in-action.pdf (accessed on 14 June 2024).
- Vertlberg, J.L.; Švajda, M.; Jakovljević, M.; Ševrović, M. Operating Vehicles' Speed Prediction Models. *Promet-Traffic Traffico*. 2024, 36, 383–398. [CrossRef]

- 9. Constantinescu, Z.; Marinoiu, C.; Vladoiu, M. Driving Style Analysis Using Data Mining Techniques. *Int. J. Comput. Commun.* 2010, *5*, 654. [CrossRef]
- 10. Eboli, L.; Guido, G.; Mazzulla, G.; Pungillo, G.; Pungillo, R. Investigating Car Users' Driving Behaviour through Speed Analysis. *Promet-Traffic Traffico.* **2017**, *29*, 193–202. [CrossRef]
- 11. Ju, U.; Wallraven, C. Dynamic measurements of speed and risk perception during driving: Evidence of speed misestimation from continuous ratings and video analysis. *PLoS ONE* **2023**, *18*, e0291043. [CrossRef]
- 12. Familar, R.; Greaves, S.; Ellison, A. Analysis of Speeding Behavior. Transp. Res. Rec. 2011, 2237, 67–77. [CrossRef]
- 13. Atombo, C.; Wu, C.; Zhong, M.; Zhang, H. Investigating the motivational factors influencing drivers intentions to unsafe driving behaviours: Speeding and overtaking violations. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *43*, 104–121. [CrossRef]
- 14. Zhang, W.; Hu, Z.; Feng, Z.; Ma, C.; Wang, K.; Zhang, X. Investigating factors influencing drivers' speed selection behavior under reduced visibility conditions. *Traffic Inj. Prev.* 2018, *19*, 488–494. [CrossRef] [PubMed]
- 15. Alizadeh, M.; Davoodi, S.R.; Shaaban, K. Drivers' Speeding Behavior in Residential Streets: A Structural Equation Modeling Approach. *Infrastructures* **2023**, *8*, 11. [CrossRef]
- 16. Truelove, V.; Watson-Brown, N.; Mills, L.; Freeman, J.; Davey, J. It's not a hard and fast rule: A qualitative investigation into factors influencing speeding among young drivers. *J. Saf. Res.* **2022**, *81*, 36–44. [CrossRef] [PubMed]
- 17. Shandhana Rashmi, B.; Marisamynathan, S. Investigating the contributory factors influencing speeding behavior among long-haul truck drivers traveling across India: Insights from binary logit and machine learning techniques. *Int. J. Transp. Sci. Technol.* 2024; *in press.* [CrossRef]
- 18. Yu, B.; Chen, Y.; Bao, S. Quantifying visual road environment to establish a speeding prediction model: An examination using naturalistic driving data. *Accid. Anal. Prev.* **2019**, *129*, 289–298. [CrossRef]
- 19. Tselentis, D.I.; Gonidi, C.; Yannis, G. Driving speed model development using driving data obtained from smartphone sensors. *Transp. Res. Proc.* **2020**, *48*, 673–686. [CrossRef]
- 20. Kontaxi, A.; Tzoutzoulis, D.-M.; Ziakopoulos, A.; Yannis, G. Exploring speeding behavior using naturalistic car driving data from smartphones. *J. Transp. Eng.* 2023, 10, 1162–1173. [CrossRef]
- Semeida, A.M. Application of artificial neural networks for operating speed prediction at horizontal curves: A case study in Egypt. JMT 2014, 22, 20–29. [CrossRef]
- Cai, Q.; Abdel-Aty, M.; Mahmoud, N.; Ugan, J.; Al-Omari, M.M.A. Developing a grouped random parameter beta model to analyze drivers' speeding behavior on urban and suburban arterials with probe speed data. *Accid. Anal. Prev.* 2021, 161, 106386. [CrossRef]
- 23. Khaddar, S.; Pathivada, B.K.; Perumal, V. Modeling over speeding behavior of vehicles using a random parameter negative binomial approach: A case study of Mumbai, India. *Transp. Res. Interdiscip. Perspect.* **2023**, *18*, 100790. [CrossRef]
- Kutela, B.; Ngeni, F.; Ruseruka, C.; Chengula, T.J.; Novat, N.; Shita, H.; Kinero, A. The influence of roadway characteristics and built environment on the extent of over-speeding: An exploration using mobile automated traffic camera data. *Int. J. Transp. Sci. Technol.* 2024; *in press.* [CrossRef]
- 25. Zhao, G.; Wu, C.; Qiao, C. A Mathematical Model for the Prediction of Speeding with its Validation. *IEEE Trans. Intell. Transp. Syst.* 2013, 14, 828–836. [CrossRef]
- Yadav, A.K.; Velaga, N.R. Investigating the effects of driving environment and driver characteristics on drivers' compliance with speed limits. *Traffic Inj. Prev.* 2021, 22, 201–206. [CrossRef] [PubMed]
- Montella, A.; Calvi, A.; D'Amico, F.; Ferrante, C.; Galante, F.; Mauriello, F.; Rella Riccardi, M.; Scarano, A. A methodology for setting credible speed limits based on numerical analyses and driving simulator experiments. *Transp. Res. Part F Traffic Psychol. Behav.* 2024, 100, 289–307. [CrossRef]
- 28. Brzine Vozila na Hrvatskim Cestama u 2018. Godini; Prometis: Zagreb, Croatia, 2019.
- 29. Malaghan, V.; Pawar, D.S.; Dia, H. Exploring Maximum and Minimum Operating Speed Positions on Road Geometric Elements Using Continuous Speed Data. J. Transp. Eng. A Syst. 2021, 147, 04021039. [CrossRef]
- Figueroa Medina, A.M.; Tarko, A.P. Speed Changes in the Vicinity of Horizontal Curves on Two-Lane Rural Roads. J. Transp. Eng. 2007, 133, 215–222. [CrossRef]
- Dumitrascu, D.-I. Influence of Road Infrastructure Design over the Traffic Accidents: A Simulated Case Study. *Infrastructures* 2024, 9, 154. [CrossRef]
- 32. Faizi, N.; Alvi, Y. Regression and multivariable analysis. In *Biostatistics Manual for Health Research*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 213–247. [CrossRef]
- Mostoufi, N.; Constantinides, A. Linear and nonlinear regression analysis. In *Applied Numerical Methods for Chemical Engineers*, 1st ed.; Elsevier: Amsterdam, The Netherlands, 2023; pp. 403–476. [CrossRef]
- 34. Fritz, M.; Berger, P.D. Will anybody buy? Logistic regression. In *Improving the User Experience Through Practical Data Analytics*; Elsevier: Amsterdam, The Netherlands, 2015; pp. 271–304. [CrossRef]
- 35. Sarker, I.H. Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. SN Comput. Sci. 2021, 2, 420. [CrossRef]
- Maceiras, C.; Cao-Feijóo, G.; Pérez-Canosa, J.M.; Orosa, J.A. Application of Machine Learning in the Identification and Prediction of Maritime Accident Factors. *Appl. Sci.* 2024, 14, 7239. [CrossRef]

- Van Efferen, L.; Ali-Eldin, A.M.T. A multi-layer perceptron approach for flow-based anomaly detection. In Proceedings of the 2017 International Symposium on Networks, Computers and Communications (ISNCC), Marrakech, Morocco, 16–18 May 2017; IEEE: New York, NY, USA, 2017; pp. 1–6. [CrossRef]
- Udurume, M.; Shakhov, V.; Koo, I. Comparative Analysis of Deep Convolutional Neural Network—Bidirectional Long Short-Term Memory and Machine Learning Methods in Intrusion Detection Systems. *Appl. Sci.* 2024, 14, 6967. [CrossRef]
- 39. Breiman, L. Random Forests. Mach. Learn. 2001, 45, 5–32. [CrossRef]
- 40. Sarker, I.H.; Salah, K. AppsPred: Predicting context-aware smartphone apps using random forest learning. *Internet Things J.* 2019, *8*, 100106. [CrossRef]
- 41. Liu, C.-Y.; Ku, C.-Y.; Wu, T.-Y.; Ku, Y.-C. An Advanced Soil Classification Method Employing the Random Forest Technique in Machine Learning. *Appl. Sci.* 2024, 14, 7202. [CrossRef]
- Dutta, P.; Paul, S.; Kumar, A. Comparative analysis of various supervised machine learning techniques for diagnosis of COVID-19. In *Electronic Devices, Circuits, and Systems for Biomedical Applications*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 521–540. [CrossRef]
- 43. Melo, F. Area under the ROC Curve. In Encyclopedia of Systems Biology; Springer: New York, NY, USA, 2013; pp. 38–39. [CrossRef]
- 44. Song, Y.-Y.; Lu, Y. Decision tree methods: Applications for classification and prediction. *Shanghai Arch. Psychiatry* **2015**, *27*, 130–135. [CrossRef]
- 45. Milanović, M.; Stamenković, M. CHAID Decision Tree: Methodological Frame and Application. *Econ. Themes* **2016**, *54*, 563–586. [CrossRef]
- Onoja, A.A.; Babasola, O.L.; Ojiambo, V. Chi-Square Automatic Interaction Detection Modeling of the Effects of Social Media Networks on Students' Academic Performance. *IOSR-JBM* 2018, 20, 43–51. Available online: https://www.iosrjournals.org/iosrjbm/papers/Vol20-issue7/Version-2/H2007024351.pdf (accessed on 3 December 2024).
- 47. Azam, M.; Arshad, A.; Aslam, M.; Gulzar, S. Application of classification methods to analyze chemicals in drinking water quality. *Accredit. Qual. Assur.* **2019**, *24*, 227–235. [CrossRef]
- Baldock, M.R.J.; Kloeden, C.N.; Lydon, M.; Ponte, G.; Raftery, S. Motorcycling in Victoria: Preliminary findings of the evaluation of the Community Education and Policing Project. In Proceedings of the 2010 Australasian Road Safety Research, Policing and Education Conference, Canberra, Australia, 31 August–3 September 2010; Australian Transport Council: Canberra, Australia, 2010.
- 49. Walton, D.; Buchanan, J. Motorcycle and scooter speeds approaching urban intersections. *Accid. Anal. Prev.* **2012**, *48*, 335–340. [CrossRef]
- 50. Jevtić, V.; Vujanić, M.; Lipovac, K.; Jovanović, D.; Pešić, D. The relationship between the travelling speed and motorcycle styles in urban settings: A case study in Belgrade. *Accid. Anal. Prev.* 2015, *75*, 77–85. [CrossRef]
- 51. Kontaxi, A.; Ziakopoulos, A.; Yannis, G. Investigation of the speeding behavior of motorcyclists through an innovative smartphone application. *Traffic Inj. Prev.* 2021, 22, 460–466. [CrossRef]
- 52. Broughton, P.S.; Fuller, R.; Stradling, S.; Gormley, M.; Kinnear, N.; O'dolan, C.; Hannigan, B. Conditions for speeding behaviour: A comparison of car drivers and powered two wheeled riders. *Transp. Res. Part F Traffic Psychol. Behav.* 2009, 12, 417–427. [CrossRef]
- 53. Garber, N.J.; Gadiraju, R. Factors affecting speed variance and its influence on accidents. *Transp. Res. Rec.* **1989**, *1213*, 64–71.
- 54. Mahmud, M.S.; Gupta, N.; Safaei, B.; Jashami, H.; Gates, T.J.; Savolainen, P.T.; Kassens-Noor, E. Evaluating the Impacts of Speed Limit Increases on Rural Two-Lane Highways Using Quantile Regression. *Transp. Res. Rec.* **2021**, 2675, 740–753. [CrossRef]
- Gupta, N.; Mahmud, M.S.; Jashami, H.; Savolainen, P.T.; Gates, T.J. Evaluating the Impacts of Freeway Speed Limit Increases on Various Speed Measures: Comparisons Between Spot-Speed, Permanent Traffic Recorder, and Probe Vehicle Data. *Transp. Res. Rec.* 2023, 2677, 357–371. [CrossRef]
- Alonso, F.; Esteban, C.; Calatayud, C.; Sanmartin, J. Speed and road accidents: Behaviors, motives, and assessment of the effectiveness of penalties for speeding. *Am. J. Appl. Psychol.* 2013, *1*, 58–64. Available online: https://pubs.sciepub.com/ajap/1/3/5/ (accessed on 14 June 2024).
- 57. Perez, M.A.; Sears, E.; Valente, J.T.; Huang, W.; Sudweeks, J. Factors modifying the likelihood of speeding behaviors based on naturalistic driving data. *Accid. Anal. Prev.* 2021, 159, 106267. [CrossRef]
- 58. Zhao, D.; Han, F.; Meng, M.; Ma, J.; Yang, Q. Exploring the influence of traffic enforcement on speeding behavior on low-speed limit roads. *Adv. Mech. Eng.* 2019, *11*, 168781401989157. [CrossRef]
- Fitzpatrick, K.; Das, S.; Gates, T.; Dixon, K.K.; Park, E.S. Considering Roadway Context in Setting Posted Speed Limits. *Transp. Res. Rec.* 2021, 2675, 590–602. [CrossRef]
- 60. Lobo, A.; Amorim, M.; Rodrigues, C.; Couto, A. Modelling the Operating Speed in Segments of Two-Lane Highways from Probe Vehicle Data: A Stochastic Frontier Approach. *J. Adv. Transp.* **2018**, 2018, 3540785. [CrossRef]
- 61. Olmez, S.; Douglas-Mann, L.; Manley, E.; Suchak, K.; Heppenstall, A.; Birks, D.; Whipp, A. Exploring the Impact of Driver Adherence to Speed Limits and the Interdependence of Roadside Collisions in an Urban Environment: An Agent-Based Modelling Approach. *Appl. Sci.* **2021**, *11*, 5336. [CrossRef]
- 62. Jurecki, R.S.; Stańczyk, T.L. A Methodology for Evaluating Driving Styles in Various Road Conditions. *Energies* **2021**, *14*, 3570. [CrossRef]
- 63. Høye, A. Speeding and impaired driving in fatal crashes—Results from in-depth investigations. *Traffic Inj. Prev.* 2020, 21, 425–430. [CrossRef] [PubMed]

- 64. Heydari, S.; Miranda-Moreno, L.F.; Fu, L. Is speeding more likely during weekend night hours? Evidence from sensor-collected data in Montréal. *Can. J. Civ. Eng.* **2020**, *47*, 1046–1049. [CrossRef]
- 65. Islam, S.; Brown, J. A comparative injury severity analysis of motorcycle at-fault crashes on rural and urban roadways in Alabama. *Accid. Anal. Prev.* 2017, *108*, 163–171. [CrossRef] [PubMed]
- 66. Shaheed, M.S.B.; Gkritza, K.; Zhang, W.; Hans, Z. A mixed logit analysis of two-vehicle crash severities involving a motorcycle. *Accid. Anal. Prev.* **2013**, *61*, 119–128. [CrossRef]
- 67. Etika, A.A.; Merat, N.; Carsten, O. Do drivers differ in their attitudes on speed limit compliance between work and private settings? Results from a group of Nigerian drivers. *Transp. Res. Part F Traffic Psychol. Behav.* 2020, 73, 281–291. [CrossRef]
- Liu, J.; Cai, J.; Lin, S.; Zhao, J. Analysis of Factors Affecting a Driver's Driving Speed Selection in Low Illumination. *J. Adv. Transp.* 2020, 2020, 2817801. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.