Accident Analysis and Prevention 196 (2024) 107427

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Assessing driver behavior in work zones: A discretized duration approach to predict speeding

Diwas Thapa [a], Sabyasachee Mishra [a,*], Asad Khattak [b], Muhammad Adeel [b]

[a] Department of Civil Engineering, University of Memphis, 3815 Central Avenue, Memphis, TN 38152, United States
[b] Department of Civil and Environmental Engineering, University of Tennessee, 322 John D. Tickle Building, Knoxville, TN 37996, United States

ARTICLE INFO

Keywords:
Speeding
Work zones
Duration framework
Mixed logit
Proactive prediction

ABSTRACT

Higher speeds in work zones have been linked to an increased likelihood of crashes and more severe crash outcomes. To enhance safety, speed limits are often reduced in work zones, aiming to create a steady flow of traffic and safer traffic operations such as merging and flagging. However, this speed reduction can also lead to abrupt speed changes, resulting from sudden braking or acceleration, increasing the risk of crashes. This disruption in speed and flow results increases the likelihood of rear-end crashes. Ensuring driver compliance with the reduced speed limits and traffic flow operations is challenging as work zones may cause frustration and lead to more instances of speeding. Therefore, proactively predicting speeding events in work zones can be crucial for the safety of both workers and road users, as it enables the implementation of speed enforcement measures to maintain and improve driver compliance in advance. In this study, we employ the duration-based prediction framework to forecast speeding occurrences in work zones. The model is used to identify significant predictors of speeding including visibility, number of lanes, posted speed limit, segment length, coefficient of variation in speed, and travel time index. Among these variables, the number of lanes, posted speed limit, and coefficient of variation of speed are positively associated with speeding. On the other hand, visibility, segment length, and travel time index are negatively associated with speeding. Results show the model's predictive accuracy is higher for speeding events with shorter durations between consecutive occurrences. The model predicted speeding within 61% of the actual epoch when speeding events within 5 h of one another were considered for validation. This indicates that the model is more effective for road segments and work zones where speeding occurs more frequently. The prediction framework can be a great asset for agencies to improve work zone safety in real-time by enabling them to proactively implement effective work zone enforcement measures to control speeding and to stay prepared, preventing potential hazards.

1. Introduction

Highway construction and maintenance play a crucial role in enhancing and sustaining transportation infrastructure, which experiences increasing use by travelers every year. During these operations, work zones are established to ensure the safety of workers and road users. These work zones can be noisy, distracting, and confusing due to the presence of heavy equipment and machinery. Consequently, they become more susceptible to safety mishaps, particularly from oncoming traffic. In fact, highway construction work is categorized as one of the most hazardous occupations. For instance, among all road construction sites, work zones involving paving/surfacing equipment operators and maintenance workers have the second and third highest fatality rates (Wang et al., 2018). Most of these fatalities and crashes can be attributed to adverse driver behavior and non-compliance with work zone safety measures. Among several factors influencing work zone crashes, speeding stands out as the most common. According to the Fatality Analysis Reporting System (FARS) database, in 2021, 32% of work zone fatalities were linked to speeding as a contributing factor, with 24% of fatal crashes resulting from rear-end collisions (Federal Highway Administration, 2023).

The Manual of Uniform Traffic Control Devices (MUTCD) classifies work zones based on their location and duration (Federal Highway Administration, 2009). It serves as a comprehensive guide for traffic control and enforcement of safety measures in work zones. The main objective of these safety measures is to ensure smooth traffic flow and

* Corresponding author.
E-mail addresses: dthapa@memphis.edu (D. Thapa), smishra3@memphis.edu (S. Mishra), akhattak@utk.edu (A. Khattak), madeel1@vols.utk.edu (M. Adeel).

https://doi.org/10.1016/j.aap.2023.107427
Received 6 September 2023; Received in revised form 26 November 2023; Accepted 13 December 2023
Available online 22 December 2023

consistent speeds throughout the work zone, thereby avoiding abrupt changes that could lead to crashes. Most work zone crashes, especially rear-end collisions, occur due to inconsistent traffic flow or sudden speed variations. To address this issue, the MUTCD provides guidance on implementing various technologies and strategies to maintain a steady traffic flow and enforce safety measures in work zones. The guide focuses on increasing compliance among road users and eliminating adverse driver behavior using regulatory strategies (such as speed photo radar enforcement and police presence) and warning strategies (including warning signs, dynamic message signs, speed feedback systems, etc.). Despite the implementation of existing safety strategies and work zone enforcement measures, work zone crashes have been on the rise in recent years. Between 2020 and 2021, work zone fatalities witnessed a troubling increase of 10.8% (Federal Highway Administration, 2023). In 2020, 39% of all work zone crash fatalities in the US occurred on interstates, with a slight rise to about 40% in 2021. The higher traffic speed and lower work zone compliance on interstates contribute significantly to the number of fatalities. Extensive evidence in the literature suggests that higher speeds are associated with more severe crashes (Osman, Mishra, et al., 2018; Osman, Paleti, et al., 2018; Shaer et al., 2024).

In this context, the ability to predict speeding can bring significant benefits from both traffic safety and operational perspectives. Having prior knowledge of potential speeding events can assist transportation planners and agencies in preparing in advance and taking necessary steps to prevent such occurrences. Therefore, this study uses a discretized duration framework to model and predict speeding on highway segments with existing work zones. The implemented duration-based framework is specifically designed to incorporate time-varying covariates into the multinomial logit model (MNL) through time-discretization. This enables the calculation of the risk of speeding in real-time, allowing forecasting road users' speeding behavior. The duration-based approach can be utilized to identify conditions related to highway, weather, and traffic flow that elevate the likelihood of speeding. Furthermore, it provides the capability to predict crash probabilities in specific highway segments in real-time. This enables agencies to proactively implement effective work zone enforcement measures to control speeding and to stay prepared, preventing potential hazards. Real-time predictions also empower agencies to strategically allocate limited speed control and regulation resources, such as message signs, barricades, etc., in critical segments such as work zones or those that are prone to speeding.

2. Literature review

In the existing literature, driver behavior within work zones is primarily characterized by compliance with two key factors: (i) the enforced speed limit and (ii) merge behavior. Notably, these two aspects are major contributors to work zone crashes, and as a result, work zone safety measures focus on promoting safer driving behavior by regulating the operating speed limit and merge behavior. Various studies have highlighted the significance of speed compliance in reducing crash risks. Higher speeds have been linked to an increased likelihood of crashes and more severe outcomes (Osman et al., 2016; Osman, Paleti, et al., 2018; Zhang & Hassan, 2019). Additionally, unsafe and aggressive merge behavior, combined with adverse weather and lighting conditions, has been identified as risky driving behavior (Debnath et al., 2015). Interestingly, aggressive driving and merge behavior are also associated with traffic speed. Drivers encountering slower speeds, congestion, and travel delays tend to become frustrated, leading to more aggressive maneuvers on the road. This highlights the interconnectedness between driving behavior and traffic flow within work zones.

2.1. Work zone risk factors and driver behavior

Work zones can lead to sudden disruptions in traffic flow, resulting in slowdowns, queues, lane change maneuvers, traffic conflicts, and speeding, all of which impact driving behavior (Flannagan and Selpi, 2019; Mishra and Zhu, 2015). Researchers have extensively studied the factors that influence driver behavior in work zones. Nearly half of all work zone crashes occur in the vicinity of the activity area (Dissanayake & Akepati, 2009). Among these crashes, approximately 42% are rear-end collisions. The main contributors to work zone crashes include inattentive driving (19%), following too closely (9.7%), and failure to yield right of way (7.5%). Driver behavior also varies based on different work zone types and activity levels. For instance, when navigating through longer work zone closures, drivers tend to travel at higher speeds (Hamdar et al., 2016). The type of barriers used also influences driver headway. Adverse weather, poor lighting conditions, and middle-aged drivers have been associated with risky driving behavior. Workers involved in work zone construction have highlighted the most hazardous conditions they face, such as working in wet weather leading to reduced visibility and skid resistance, driver frustration, aggression towards traffic controllers, and distracted driving due to mobile phone use (Debnath et al., 2015). Similarly, according to Debnath et al. (2015), workers consider non-daylight hours (dawn, dusk, and night) as the most hazardous times for work zone activities, which is attributed to a higher number of drunk drivers and reduced visibility (Debnath et al., 2015). Additionally, workers perceive working on freeways and hilly/curved roads as risky. Regarding speed compliance, workers consider police enforcement, the presence of police cars (even without an officer present), installation of speed bumps, and work zone-oriented driver education as the most effective countermeasures.

2.2. Predicting driving behavior and traffic flow

As previously mentioned, driving behavior and traffic flow are mutually dependent. Many studies examining driver behavior under various circumstances, such as the implementation of new work zone enforcement measures, have employed three main approaches: i) Field observation and analysis, ii) Traffic micro and macrosimulation, for instance, studies conducted by (Berthaume, 2015; Gan et al., 2021; Hou & Chen, 2019), and iii) Driving simulator experiments, as demonstrated in research conducted by (Algomaiah & Li, 2022; Bashir & Zlatkovic, 2021). The first approach, field observation and analysis, is beneficial when there is no prevalent risk or when adequate safety for road users can be ensured, as seen in previous studies (Benekohal et al., 2010; Mishra et al., 2021; Thapa & Mishra, 2021). On the other hand, the latter two approaches, traffic micro and macrosimulation, and driving simulator experiments, are preferred to avoid hazardous conditions and provide controlled environments for studying driver behavior in work zones.

Understanding the impact of work zones and driving behavior on traffic flow is crucial from an Intelligent Transportation Systems (ITS) perspective. Real-time and accurate traffic data play a vital role in various ITS applications, including traffic planning and management, incident detection and management, travel time estimation, traffic predictions, and traffic planning. To achieve these objectives, researchers have focused on accurately forecasting traffic flow, for missing data and future conditions.

Numerous research approaches have been explored in this area, including time series and regression analysis, Kalman filter, machine learning techniques such as neural networks and support vector machines, as well as deep learning techniques like convolutional neural networks, long short-term memory, and graphical convolutional networks. For a detailed description of these methods and relevant literature, readers are encouraged to refer to studies conducted by (Medina-Salgado et al., 2022) and (Kashyap et al., 2022). In summary, the primary goal of these methods is to forecast traffic flow conditions rather than focusing on driving behavior, contributing to the advancement of ITS applications and traffic management.

2.3. Predicting speeding behavior

Various approaches have been employed in the existing literature to predict driving intention and behavior related to violating traffic laws. The theory of planned behavior has been widely utilized in multiple studies (e.g., (Dinh & Kubota, 2013; Elliott & Thomson, 2010; Forward, 2009; Jovanović et al., 2017; Scott-Parker et al., 2013)). For instance, Cestac et al. (2011) investigated young drivers and found that different latent constructs influenced speeding behavior in different driver groups (Cestac et al., 2011). Novice drivers were influenced by thrill-seeking, beginners by subjective norms, and experienced drivers by the feeling of being in control. In another study, researchers reported that as young drivers are more likely to speed as they gain confidence in their driving abilities (Simons-Morton et al., 2012). Risky peer influences were found to be significant predictors of speeding among novice teenage drivers.

Several studies have utilized naturalistic driving data to understand and predict speeding behavior. For instance, Yu et al. (2019) used naturalistic driving data to develop a speeding prediction model (Yu et al., 2019). The study emphasized the role of driver's visual perception as a major factor in speeding. The prediction model was built based on visual road information, environmental variables, vehicle kinematics, and driver characteristics, utilizing a Random Forest algorithm to achieve an accurate prediction rate of 85%. In a similar vein, Kong et al. (2020) investigated hidden rules governing speeding duration and patterns using naturalistic driving data to understand speeding behavior (Kong et al., 2020). Through classification-based association, they found that moderating speeding was associated with shorter trips, absence of median, and lower functional classes. Conversely, longer trips and higher functional classes were linked to longer speeding events. Perez et al. (2021) employed naturalistic driving data to investigate factors influencing speeding behavior (Perez et al., 2021). They summarized the likelihood of speeding using a beta binomial regression and a driver questionnaire. The authors reported that the odds of younger drivers, aged 16–24 years, engaging in speeding were 1.5 times higher than that of 80-year-old drivers. Additionally, the odds of speeding at lower speed limits (10–20 mph) were 9.5 times higher compared to speeding when speed limits were over 60 mph. Researchers utilized the Strategic Highway Research Program 2 (SHRP 2) naturalistic driving data to study the duration of speeding events, aiming to better understand driving behavior. They reported that driving 10 mph over the speed limit was a common occurrence, with 99.8% of drivers speeding at least once within their trips. The average number of speeding events reported in the study was 2.75 per trip (Richard et al., 2020).

Zhao et al., (2013) developed a mathematical model to predict intentional and non-intentional speeding (Zhao et al., 2013). The model utilized in-vehicle sensor data and driver characteristics to calculate speeding probabilities. The experiments were conducted using a driving simulator, and the authors reported an average prediction accuracy of over 80%.

In another study, Cheng et al., (2019) adopted a two-step approach to identify and predict speed violations (Cheng et al., 2019). They used a binary logit model to identify variables contributing to speeding violations and then applied a decision tree method to predict specific types of speeding violations, such as "foreign license plate" and "intersection" among others. The study found that country roads had a higher incidence of speeding violations compared to urban roads, primarily due to the lower presence of traffic control infrastructure and lower traffic flow. Higher and more intense rainfall was associated with increased speeding violations, while local drivers were less likely to violate speed limits.

3. Study contributions

This study contributes to the literature in three major ways:

i. First, no studies have attempted to develop or implement an econometric framework for proactively predicting speeding, especially in work zones. Based on the literature review, numerous studies have examined speeding behavior using the theory of planned behavior. Additionally, a separate body of literature focuses on predicting speeding at the individual driver level, utilizing environmental and in-vehicle data. Furthermore, another set of studies has applied machine learning, and deep learning techniques to forecast traffic flow and speed, enabling various actions such as crash and congestion prevention, emergency messaging for traffic diversion, rerouting, and queue management, particularly in situations with insufficient or missing disaggregated data. Despite the wealth of research in these areas, we are not aware of any previous study attempting to forecast the likelihood of speeding in the future using historical data and time-varying covariates through a parametric approach. This study aims to fill this research gap by providing insights into predicting speeding behavior using historical data, and time-varying covariates with a parametric approach.

ii. While many prediction models rely on modern data-driven black-box machine learning and artificial intelligence algorithms, our approach is based on exponential models (survival model and MNL). These parametric methods offer the advantage of providing causal inferences through variable effects, including coefficients and marginal/elasticity effects. The enables researchers to gain deeper insights into the relationships between the predictors and speeding.

iii. Besides Thapa et al. (2022), there have been no implementations of the duration-based model) (Thapa et al., 2022). Notably, the original study focused on investigating traffic crashes but overlooked the presence of segment-specific mixed effects despite repeated observations across segments. This research distinguishes itself from the original study as the first to implement the framework for predicting speeding in a work zone. It showcases the integration of real-time weather, traffic flow, and congestion data alongside static covariates like highway characteristics. As mentioned previously, to our knowledge, this is the only study directed at understanding and predicting speeding in an active work zone. Additionally, current research considers the presence of unobserved heterogeneity resulting from multiple speeding events occurring in the same highway segment through a mixed model.

4. Methodology

The description of the duration-based framework here is taken largely from Thapa, et al. (2022) (Thapa et al., 2022; Thapa et al., 2024). Utilizing the duration-based framework, we can determine the likelihood of speeding at a particular time-interval *t*, considering that no speeding has been observed in previous time-intervals. This probability is represented by the hazard function *h(t)*, which can be formulated using a constant hazard rate, *h*.

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{he^{-ht}}{1 - (1 - e^{-ht})} = h \tag{1}$$

In the given equation, we represent the probability distribution function and probability density function related to a continuous random variable for time *T* as *f(t)* and *F(t)*, respectively. The probability density function, in this context, indicates the likelihood of observing speeding by time *t*. This is expressed by equation 2 as follows.

$$F(t) = Pr(T \leq t) \tag{2}$$

Assuming that the time duration between consecutive speeding events is discretized into *n* time-intervals, each having a duration of *dt*, we can express the probability of observing the next speeding event at a

specific interval n since the occurrence of the last speeding event as follows:

$$\begin{aligned} Pr(T = ndt) &= Pr(T \le ndt) - Pr(T \le (n-1)dt) \\ &= F(ndt) - F((n-1)dt) \\ &= exp(-h(n-1)dt) - exp(-hndt) \end{aligned}$$

$$= \frac{exp(-h(n-1)dt)}{1/(1-exp(-hdt))} \tag{3}$$

Using a Taylor series expansion, i.e., $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots \infty \; -1 < x < 1$ in the denominator produces equation 4.

$$\begin{aligned} Pr(T = ndt) &= \frac{exp(-h(n-1)dt)}{1+exp(-hdt)+exp(-2hdt)+exp(-3hdt)+\cdots\infty} \\ &= \frac{exp(U_n)}{exp(U_1)+exp(U_2)+exp(U_3)+\cdots\infty}, \text{where } U_n = -h(n-1)dt \end{aligned}$$

$$= \frac{exp(U_n)}{\sum_1^\infty exp(U_c)} \tag{4}$$

Simplifying equation 4 makes it evident that the probability of speeding at the n^{th} interval can be represented as MNL model with infinite alternatives for n. The utility equation for the alternatives can be expressed as $U_n = -h(n-1)dt$. As a result, the utility equation can be modified to accommodate non-linear hazard profiles, as demonstrated in equation 5. However, when disregarding all higher-order polynomial terms, equation 5 reduces to a simple MNL model.

$$U_n = \beta_1(n-1)dt + \beta_2[(n-1)dt]^2 + \beta_2[(n-1)dt]^3 + \cdots \tag{5}$$

For example, consider the time between consecutive speeding events observed at a specific segment, denoted as s, is discretized into epochs e, with C number of time-intervals, each lasting for dt duration. An illustrative example of this discretization is presented in Table 1, where two speeding events are observed, 4 h apart, with $e = 1$ h and $dt = 15$ min. Consequently, each epoch contains four distinct time-intervals, indexed as i={*1,2,3,4*}, and each time-interval provides the corresponding time elapsed since the previous speeding event. This relationship allows us to pinpoint the exact time-interval when the subsequent speeding event occurred. For example, consider Table 1, which denotes speeding was observed at the fourth time-interval of the fourth epoch, denoted as 1, otherwise 0. As a result, the time elapsed between the two speeding events can be calculated as $t_{e,i} = (e-1)Cdt + (i-1)dt = (4-1)*4*0.25 + (4-1)*0.25 = 3.75 hours$. This relationship enables us to construct the utility function considering the time-intervals as choice alternatives in the MNL model. The utility function includes the duration dynamics as the first element and a vector of time-varying covariates as the last element. Notably, even static variables that do not change with time, such as the number of lanes associated with the highway segment, were transformed into dynamic covariates by multiplying them with the corresponding value of $t_{e,i}$ to account for the effect of time elapsed.

$$U_{s,e,i} = \beta_1 t_{e,i} + \cdots + r' X_{s,e,i} \tag{6}$$

Alongside the four time-intervals, there is a fifth alternative to consider, signifying whether the next speeding event will be observed in the current epoch (0) or the next epoch (1). This particular choice alternative serves as the base with an intercept term and can be represented as follows.

$$U_{s,e,C+1} = \beta_{C+1} \tag{7}$$

Since speeding at any time-interval is conditional upon no prior speeding, the conditional probability for any time-interval can be expressed as follows.

$$Pr(T_s = t_{e,i} | T_s > (e-1)Cdt) = \frac{exp(U_{s,e,i})}{\sum_{c=1}^{C} exp(U_{s,e,c}) + exp(U_{s,e,C+1})} \tag{8}$$

We can obtain it by multiplying the conditional probability with the product of probabilities for the fifth alternative, as demonstrated in equation 9.

$$\begin{aligned} &Pr(T_s = t_{e,i}) \\ &= \frac{exp(U_{s,e,i})}{\sum_{c=1}^{C} exp(U_{s,e,c}) + exp(U_{s,e,C+1})} \\ &\quad \times \prod_{e^*=1}^{e-1} \frac{exp(U_{s,e^*,C+1})}{\sum_{c=1}^{C} exp(U_{s,e^*,c}) + exp(U_{s,e^*,C+1})} \end{aligned} \tag{9}$$

Estimating the model parameters, represented as the vector $n = (\beta_1, \cdots, r, \beta_{C+1})'$, involves maximizing the likelihood function associated with the probabilities in equation 9 across all speeding events and segments. It's essential to note that the data, after time discretization, takes the form of panel data with multiple speeding events observed at each segment. Thus, it becomes necessary to consider unobserved heterogeneity at the segment level. To address this, the vector of parameters for any segment is assumed to follow a multivariate normal distribution. The resulting mixed logit model is then estimated by integrating the vector of parameters over this distribution. The random parameters for the mixed logit model are obtained as Cholesky parameters by estimating the elements of the unconstrained lower triangular Cholesky matrix, represented as Γ. This estimation is performed in relation to their variance–covariance matrix, denoted as Σ, such that $\Gamma\Gamma' = \Sigma$.

5. Data

5.1. Speeding events

This research focused on speeding incidents observed in a work zone set up on I-65 in Robertson County, Tennessee. The I-65 segments within the county are currently undergoing lane expansion in both North and Southbound lanes. The specific location of these interstate segments within Robertson County can be seen in Fig. 1. Based on the data obtained from Google Maps Street View, it was determined that the work zone has been active since July 2022. Therefore, the study period considered for this research spans from July 1, 2022, to May 31, 2023. The work zone consists of 14 INRIX Traffic Management Center (TMC) segments, in Table 2. To identify speeding events during the study period, speed data at 15-minute intervals was collected for these 14 TMCs, along with the reference speed for the highway segments. In the context of INRIX, the reference speed represents the average speed of vehicles over the study period. For this study, speeding was identified for any given time-interval whenever the average speed exceeded the reference speed by 10 mph. Our rationale for this choice was to ensure a high level of confidence in identifying speeding, as many vehicles tend to drive at or above the posted speed limit. Therefore, we took into account the average speed observed in the segment and added 10 mph as a threshold to confidently identify instances of speeding. Employing this method, a total of 2,444 speeding events were identified. It is essential to note that each speeding event corresponds to a specific 15-minute time interval during which the average vehicle speed exceeded the reference speed by 10 mph.

Table 1
Example demonstrating the discretization of duration between speeding events.

Segment	Time to next speeding (hours)	Epoch	First 15-min	Second 15-min	Third 15-min	Fourth 15-min	Next epoch
A	4	1	0	0	0	0	1
A	4	2	0	0	0	0	1
A	4	3	0	0	0	0	1
A	4	4	0	0	0	1	0

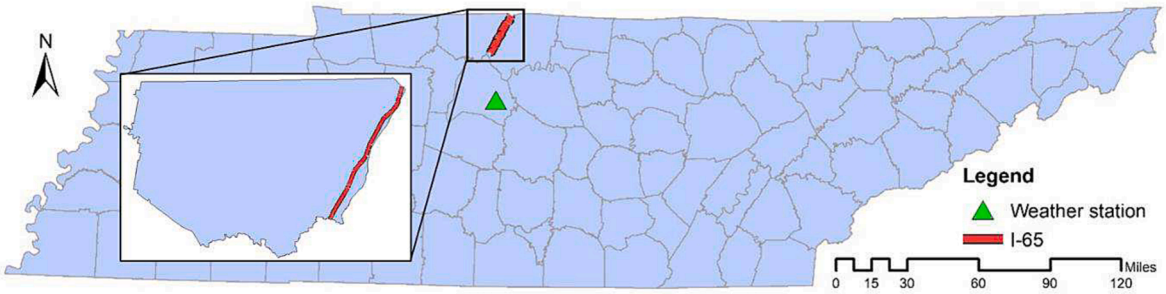

Fig. 1. Location of TMC segments within Robertson County and weather station.

Table 2
TMC segments within the work zone.

Northbound TMC segments	Sequence	Length (mi)	Southbound TMC segments	Sequence	Length (mi)
121 + 04668	1	3.49	121–04670	1	3.48
121P04668	2	0.59	121 N04670	2	0.55
121 + 04669	3	4.15	121–04669	3	4.48
121P04669	4	0.59	121 N04669	4	0.53
121 + 04670	5	4.42	121–04668	5	4.07
121P04670	6	0.56	121 N04668	6	0.63
121 + 04671	7	3.39	121–04667	7	3.46

Note: TMC segments with the character "P" or "N" represents segments with ramps.

5.2. Data aggregation

As the INRIX TMC data only provides traffic flow information, additional datasets were utilized to obtain covariates that could effectively represent the highway and weather conditions in the segments during the study period and at the time of each speeding event. The following datasets were used for this purpose:

i. *Traffic flow data*: Traffic flow data was sourced from the TMC data provided by INRIX, as mentioned earlier. This data contained disaggregated traffic speed information, recorded at 5-minute intervals along with respective date and time for the study period. This data was aggregated to 15-minute intervals for the purposes of model development and speeding identification. Moreover, the Coefficient of Variation (COV) was calculated from the 5-minute intervals for each 15-minute interval to factor in variations in speed. Additionally, to address the presence of congestion, values of the travel time index was obtained at 15-minute intervals for the study period. Additionally, yearly averages for traffic composition were also collected for each segment. All of this data represented the traffic characteristics and conditions affecting speeding events during the study period.

ii. *Highway characteristics*: The necessary highway characteristics were collected from the Enhanced Tennessee Roadway Information Management System (ETRIMS), a query-based highway information system managed by the Tennessee Department of Transportation. Among the variables obtained, terrain type, lighting conditions, roadway conditions, and illumination were found to be consistent across all segments and, therefore, were not included in the analysis. The retained variables were the number of lanes and the posted speed limit, as they varied across the segments and were considered essential for the study. A geospatial proximity approach was applied using road inventory data to integrate these highway characteristics with the traffic flow data. This ensured that the relevant highway characteristics were properly aligned with each segment's corresponding traffic flow data.

iii. *Weather conditions*: Weather data for the freeway segments was collected from the Local Climatological Data (LCD) recorded by the nearest weather station situated at Nashville International Airport. The hourly weather data was then merged with the traffic flow data based on the corresponding date and time. Fig. 1 displays the weather station's location relative to the interstate segments. Given that our prediction model's smallest temporal resolution was 15 min with 1-hour epochs, we integrated the hourly weather conditions to account for the impact of changing weather conditions. To achieve this, we multiplied the hourly weather conditions for the corresponding epoch with the time elapsed for each time interval. It is important to note that the hourly weather data included multiple conditions observed within an hour, such as cloudy, rainy, and clear weather. For this study, we categorized the data into two main categories: "Clear" when no adverse weather conditions were observed in a particular hour, and "Other" when at least one adverse weather condition was noted. Additionally, hourly visibility data in miles was obtained as part of the weather conditions dataset. The descriptive statistics for the variables obtained from these datasets are presented in Table 3.

5.3. Training and testing data

After identifying the speeding events, 90% of them (2,200 events) were randomly selected to create forecasting epochs for generating the training data, as explained in the Methodology section. This selection process is referred to as Epoch level sampling (Thapa et al., 2022). All the MNL and mixed logit models were estimated using this training data. The remaining 10% of the speeding events were used to create the testing data for model validation. This data was used to assess the performance and accuracy of the models.

6. Results

The model estimation process comprised two primary steps. In the initial step, a fixed parameter MNL model was estimated by maximizing the log-likelihood associated with the probabilities in equation 9. The parameters acquired during this step served as initial values for the subsequent estimation of the mixed model, as described in the last paragraph of the Methodology section. In this second step, the mixed model was estimated, including the estimation of mean parameters and the parametrization of the covariance matrix of the random parameters using Cholesky decomposition. The mean parameters from our model estimations are shown in Table 4. The mean parameters exhibit similar values between the models; however, the mixed MNL model stands out as a superior fit due to its lower log-likelihood value at convergence. The *t*-stat values for the mixed model are notably large suggesting low values of standard error for the respective estimates. Table 5 presents the Cholesky parameters for the mixed model with respective *t*-stat within parenthesis. The reader will note that the significant digits in the value of *t*-stat are reduced in the table to accommodate the results on the same

Table 3
Descriptive statistics.

Categorical variables	Frequency of speeding			Relative abundance		
Time of day when speeding was observed						
Early morning (6 a.m. to 9 a.m.)	133			5.44 %		
Late morning (9 a.m. to 12p.m.)	8			0.33 %		
Early afternoon (12p.m. to 3p.m.)	14			0.57 %		
Late afternoon (3p.m. to 6p.m.)	10			0.41 %		
Evening (6p.m. to 12 a.m.)	453			18.54 %		
Night (12 a.m. to 6 a.m.)	1,826			74.71 %		
Weather condition						
Clear	2,130			87.15 %		
Other	314			12.85 %		
Continuous variables	**Min**	**Q1**	**Median**	**Mean**	**Q3**	**Max**
Time between speeding (hours)	0.25	1.75	18.62	44.38	44.25	1,559
Driving conditions						
Hourly visibility (miles)	0.121	9.940	10	9.229	10	10
Highway characteristics						
Number of lanes (both directions)	4	4	4	4.58	5	7
Speed limit	55	55	55	61	70	70
Traffic flow characteristics						
Coefficient of variation of speed	0	0.014	0.045	0.055	0.082	0.530
Travel time index	0.7	1	1	1.095	1.1	26.3
Peak hour (%)	7	8	8	7.95	8	9
Passenger vehicles (%)	67	67	69	68.41	69	70
Single unit trucks (%)	3	3	3	3	3	3
Multiple unit trucks (%)	27	28	28	28.59	30	30
% Peak SU trucks	2.31	2.31	3.26	2.87	3.26	3.26
% Peal MU trucks	20.17	20.17	20.17	21.66	23.84	23.84

page. The mean parameters presented in Table 4, such as those for the fixed model, can be utilized to construct the utility function for each alternative following Equation 6 outlined in the Methodology section. This allows us to calculate the utility for each choice alternative,

$U_{s,e,i} = -1.582{*}t_{e,i} - 15.813{*}Hourlyvisbilitymorning + 2.187{*}Numberoflanes\cdots$

The reference category corresponding to the Next epoch, $U_{s,e,C+1} = 4.065$

Based on the adopted methodology and utility equations, the results can be interpreted as follows: a positive coefficient (or a positive correlation) implies that an increase in the variable is associated with a higher likelihood of speeding in the current epoch compared to subsequent epochs. Conversely, a negative coefficient (or a positive correlation) implies that speeding is more likely in the subsequent epochs. The same interpretation applies to categorical variables: a positive coefficient suggests an increased likelihood of speeding in the current epoch for the specific category compared to the base category, while a negative coefficient suggests a decreased likelihood. We subsequently discuss our findings as follows. The findings indicate that higher visibility is negatively correlated with speeding events. As the number of lanes increases, the likelihood of speeding events also increases. As anticipated, a higher travel time index, which reflects congestion, is negatively linked to speeding events. Speeding is most likely during late morning hours. It is worth noting that daytime variables that were statistically insignificant in the fixed parameter model were statistically significant in the mixed model. The effect of posted speed limit was positive in the fixed model but changed to negative in the mixed model (β = 0.789, t-stat = 9.67 versus β = -1.036, t-stat = -285.47). This suggests a strong presence of heterogeneity at the segment level. Segments with a higher peak hour percentage are positively associated with speeding. Surprisingly, clear weather conditions were found to be negatively associated with speeding, suggesting that drivers may be more cautious in clear weather compared to adverse weather conditions.

Table 4
Mean parameters for fixed and mixed models.

Variable groups	Variables	Coeff. (t-stat) Fixed model	Mixed model
Intercept	Intercept	4.065 (26.300)	3.863 (117.901)
Duration dynamic	Time since speeding	−1.582 (-6.228)	−1.813 (-196.957)
Driving condition	Hourly visibility (miles)	−15.813 (-6.233)	−18.131 (-209.859)
Highway characteristics	Number of lanes	2.187 (2.876)	2.298 (110.530)
	Posted speed limit (mph)	0.789 (9.667)	−1.036 (-285.468)
	Segment length (mi)	−0.242 (-2.800)	−0.762 (-41.076)
Traffic flow characteristics	Coefficient of variation (Speed)	1.357 (3.276)	0.327 (5.739)
	Travel time index	−0.664 (-4.152)	−1.862 (-25.511)
	DHV %	6.464 (5.404)	3.930 (38.579)
	SU Tracks %	−4.746 (-6.23)	−5.441 (-208.491)
	MU Trucks %	14.383 (6.059)	4.601 (11.322)
	% Peak SU Trucks	−9.645 (-6.158)	−9.275 (-330.679)
	% Peak MU Trucks	−14.564 (-6.406)	–23.581 (-60.451)
Time of day	Early morning (6 a.m. to 9 a.m.)	0.172 (0.274)*	0.320 (37.412)
(base = Night (12 a.m. to 6 a.m.))	Late morning (9 a.m. to 12p.m.)	9.644 (4.798)	9.713 (2,042.216)
	Early afternoon (12p.m. to 3p.m.)	3.989 (1.746)*	4.027 (1,074.780)
	Late afternoon (3p.m. to 6p.m.)	1.587 (0.839)*	1.622 (427.829)
	Evening (6p.m. to 12 a.m.)	0.028 (0.113)*	−1.093 (-18.650)
Weather condition (base = Other)	Clear	−1.582 (-6.228)	−1.862 (-25.511)
Model fit measures	Observations	97,346	
	Average initial LL	−0.381	−0.381
	Average final LL	−0.133	−0.124
	McFadden's R-squared	0.651	0.674

*Indicates the variables were not statistically significant at 5% level of significance.

7. Validation

The duration-based approach offers the capability to predict crash probabilities in specific highway segments in real time. Noting the practicality of the model, it is essential to demonstrate its predictive capability to establish its validity. For this purpose, validation of the prediction model was performed at two levels, first, the ability of the model to predict the epoch where speeding was observed and second, the accuracy of model's prediction regarding the time-interval at which speeding was observed. These predictions were derived from the speeding probabilities obtained through the model predictions. To elaborate, the estimated model was utilized to calculate the probabilities of speeding for forecasted epochs and time-intervals. The specific time-

Table 5
Correlated random parameter model (Cholesky parameters).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	-1.311 (-24)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	-0.534 (-11)	0.225 (11)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1.245 (23)	0.604 (42)	0.232 (11)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0.759 (41)	0.001 (0)	-1.066 (-16)	-1.856 (-24)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0.028 (1)	0.039 (11)	-1.01 (-18)	1.789 (26)	0.546 (16)	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	-0.089 (-3)	0.035 (8)	-0.555 (-13)	-0.905 (-17)	-0.397 (-18)	0.658 (14)	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0.04 (1)	0.049 (10)	0.131 (17)	-0.564 (-16)	-0.107 (-12)	0.627 (14)	0.146 (14)	0	0	0	0	0	0	0	0	0	0	0	0
8	1.31 (30)	0.098 (16)	0.083 (8)	-3.105 (-19)	-0.215 (-15)	0.856 (17)	0.292 (19)	4.086 (18)	0	0	0	0	0	0	0	0	0	0	0
9	1.242 (18)	0.008 (3)	-0.831 (-18)	-1.623 (-28)	-0.225 (-17)	0.45 (32)	0.025 (6)	1.72 (30)	-0.503 (-18)	0	0	0	0	0	0	0	0	0	0
10	-0.072 (-4)	-0.013 (-4)	0.068 (5)	1.064 (17)	0.325 (14)	-0.177 (-19)	-0.04 (-10)	-0.504 (-18)	-0.028 (-6)	-0.328 (-22)	0	0	0	0	0	0	0	0	0
11	1.116 (27)	0.02 (6)	-0.817 (-17)	-1.651 (-20)	-0.122 (-14)	0.147 (13)	0.06 (13)	1.248 (19)	-0.115 (-17)	1.105 (19)	0.892 (20)	0	0	0	0	0	0	0	0
12	1.343 (24)	-0.033 (-8)	-0.874 (-14)	1.265 (13)	0.06 (13)	-0.098 (-9)	-0.098 (-13)	-1.339 (-14)	-0.006 (-2)	-1.045 (-13)	0.202 (48)	-0.09 (-12)	0	0	0	0	0	0	0
13	2.697 (32)	0.037 (8)	0.804 (16)	-3.76 (-17)	-0.464 (-16)	0.834 (14)	0.11 (11)	3.04 (15)	-0.569 (-20)	3.395 (16)	-0.141 (-28)	-0.221 (-29)	-0.045 (-11)	0	0	0	0	0	0
14	0.199 (3)	0.023 (6)	0.464 (16)	2.713 (31)	0.874 (16)	-0.286 (-19)	0.068 (12)	-0.534 (-30)	0.463 (19)	-1.047 (-24)	0.779 (19)	-0.274 (-19)	0.034 (9)	0.015 (4)	0	0	0	0	0
15	1.248 (19)	0.01 (3)	-0.836 (-13)	0.302 (9)	-0.166 (-14)	-0.078 (-9)	0.03 (9)	-0.143 (-6)	0.233 (18)	-0.57 (-14)	1.495 (17)	0.048 (14)	-0.004 (-1)	-0.005 (-1)	0.008 (3)	0	0	0	0
16	0.018 (0)	0.079 (13)	-0.223 (-11)	0.314 (16)	0.16 (18)	0.221 (14)	0.237 (16)	1.318 (15)	0.681 (17)	-0.039 (-2)	-0.035 (-2)	0.127 (10)	0 (0)	0.004 (1)	0.003 (1)	-0.202 (-32)	0	0	0
17	-0.095 (-1)	0.048 (10)	0.016 (1)	2.896 (46)	0.477 (16)	0.357 (11)	0.144 (13)	-0.028 (-1)	0.61 (20)	-0.785 (-42)	0.228 (14)	0.105 (21)	0.07 (13)	0.042 (11)	0.044 (11)	-0.427 (-21)	0.479 (13)	0	0
18	-0.273 (-3)	0.056 (12)	-0.584 (-23)	1.368 (28)	0.794 (16)	0.429 (17)	0.168 (16)	0.748 (17)	0.465 (17)	0.036 (3)	1.487 (19)	0.011 (3)	0.021 (6)	0.004 (1)	0.005 (2)	-1.019 (-15)	0.559 (17)	0.056 (12)	0
19	-0.272 (-8)	0.021 (5)	-0.689 (-14)	-1.154 (-12)	-0.047 (-11)	0.37 (13)	0.062 (9)	1.207 (11)	-0.17 (-13)	1.329 (12)	-0.802 (-18)	-0.01 (-3)	0.002 (1)	0.004 (1)	0.008 (3)	-0.781 (-27)	0.205 (10)	0.021 (5)	-1.283 (-21)

Note: 1 = Intercept, 2 = Time since speeding, 3 = COV (speed), 4 = Posted speed limit, 5 = Number of lanes, 6 = DHV %, 7 = SU Tracks %, 8 = MU Trucks %, 9=% Peak SU Trucks, 10=% Peak MU Trucks, 11 = Segment length (mi), 12 = Early morning (6 a.m. to 9 a.m.), 13 = Late morning (9 a.m. to 12p.m.), 14 = Early afternoon (12p.m. to 3p.m.), 15 = Late afternoon (3p.m. to 6p.m.), 16 = Evening (6p.m. to 12 a.m.), 17 = Visibility (miles), 18 = Clear weather, 19 = Travel time index.

interval and its associated epoch with the highest probability of speeding were identified as the exact time when speeding was anticipated to occur. Subsequently, the corresponding epoch and time-interval were employed in the assessment and validation of the model, as outlined below.

i. Epoch level prediction: At the epoch level, we introduce a new measure called Predicted Temporal Proximity $(PTP) = \left|\frac{Predicted\ overspeeding\ epoch - Actual\ overspeeding\ epoch}{Actual\ overspeeding\ epoch}\right| * 100\%$ to assess the model's performance in predicting the epoch when speeding was observed. A lower value of PTP indicates a more accurate prediction of the epoch of speeding. To evaluate the model's predictive ability across different temporal ranges, we created subsets of the test data by removing speeding events that occurred at higher numbers of epochs. We then calculated the average value of PTP for each subset. We created these subsets for validation to explore the impact of the duration between events on model predictions. This is especially relevant as the framework is duration-based, capable of providing insights into the temporal dynamics of events. As discussed in the Methodology section, this is accomplished through addition of dynamic covariates that vary across time-intervals, and static covariates multiplied by $t_{e,i}$ to account for the effect of time elapsed.

The results from the average PTP, as shown in Fig. 2 indicate that speeding events observed within 25 epochs of the last speeding event have notably smaller PTP values. For instance, when considering speeding that occurred within 5 epochs, the average PTP is 61%. However, this average PTP increases to 76% when considering speeding within 25 epochs. These findings suggest that the model's predictions are more accurate for road segments where speeding is more commonly observed. Additionally, the model suggests that predictions become less reliable as the number of epochs increases, indicating a potential decrease in accuracy for predicting speeding events that are farther apart in time.

ii. Time-interval level prediction: The model's predictions at the time-interval level were evaluated by considering the rates of False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN) using two key metrics: Specificity and Sensitivity.

As presented by equation 10, Specificity is defined as the proportion of correctly identified non-speeding intervals (TN) among all the non-speeding intervals (TN + FP). It represents the model's ability to predict the absence of speeding events accurately. Sensitivity, on the other hand, as presented in equation 11 is defined as the proportion of correctly identified speeding intervals (TP) among all the speeding intervals (TP + FN). It measures the model's ability to predict the presence of speeding events correctly.

$$Specificity = \frac{True\ Negatives(TN)}{True\ Negatives(TN) + False\ Positives(FP)} \quad (10)$$

$$Sensitivity = \frac{True\ Positives(TP)}{True\ Positives(TP) + False\ Negatives(FN)} \quad (11)$$

Model predictions results were: TN = 544, TP = 56, FP = 188, FN = 188, Specificity = 0.74, Sensitivity = 0.23. High Specificity indicates low false positives, while low value of Sensitivity suggests a high rate of false negatives. It is important to consider that after reformulating the speeding data, there is a preponderance of 0 s (non-speeding intervals) compared to 1 s (speeding intervals). Given this imbalance, the low Sensitivity is expected. The model may tend to predict non-speeding intervals (TN) more accurately but struggle to identify all the instances of speeding intervals (TP). This is a common challenge in models dealing with imbalanced datasets, and further efforts may be required to improve the Sensitivity while maintaining a high Specificity.

8. Discussion

Utilizing the capability of parametric regression models to derive variable effects, we examined the influence of several variables on speeding probability. The variables considered in this investigation are listed below. The influence of other variables can also be derived similarly.

Effect of time: Fig. 3(a) demonstrates that the effect of time on speeding remains consistent across the four alternatives. As the time since the last speeding event increases, the probability of speeding steadily rises with each time interval. In this analysis, the time since the last speeding event is scaled between a minimum value of 0 and a maximum value of 1 to avoid the undue influence of large differences between the minimum and maximum time values.

Effect of visibility: Fig. 3(b) illustrates that as visibility increases, the likelihood of speeding decreases for the first and second time-intervals, but steadily increases afterward. Furthermore, the rate of increase in the probability of speeding is highest during the final time interval. Conversely, the rate of decrease in the probability of speeding is most pronounced during the first time-interval. This observation indicates that the change in the probability of speeding with variations in visibility is dependent on the specific time interval.

Effect of number of lanes: Fig. 3(c) demonstrates that the increase in the probability of speeding is characterized by an upward curve as the number of lanes increases. However, it is worth noting that the differences between the time intervals themselves are not easily distinguishable. In other words, although there is no clear time-interval-specific differentiation, the probability of speeding exhibits a non-linear increase with an increase in the number of lanes.

Effect of COV: The effect of the coefficient of variation of speed appears to be linear (see Fig. 3(d)). While there is a small difference between the probabilities observed at the second and third time-intervals, the contrast between the first and fourth time-intervals is more noticeable, with the latter being associated with a higher probability. In other words, although there is a notable difference between the probabilities associated with the first and fourth time-intervals, the magnitudes on the y-axis suggest that the value differences are relatively small.

Effect of segment length: As discussed in the literature review section, studies have indicated that work zones with longer lane closures are

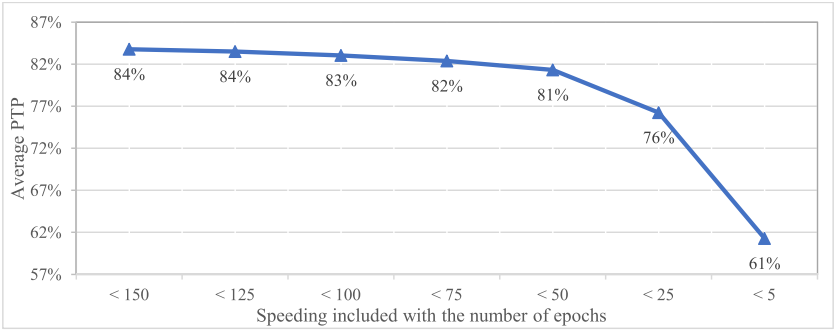

Fig. 2. Value of PTP for different subset of test data.

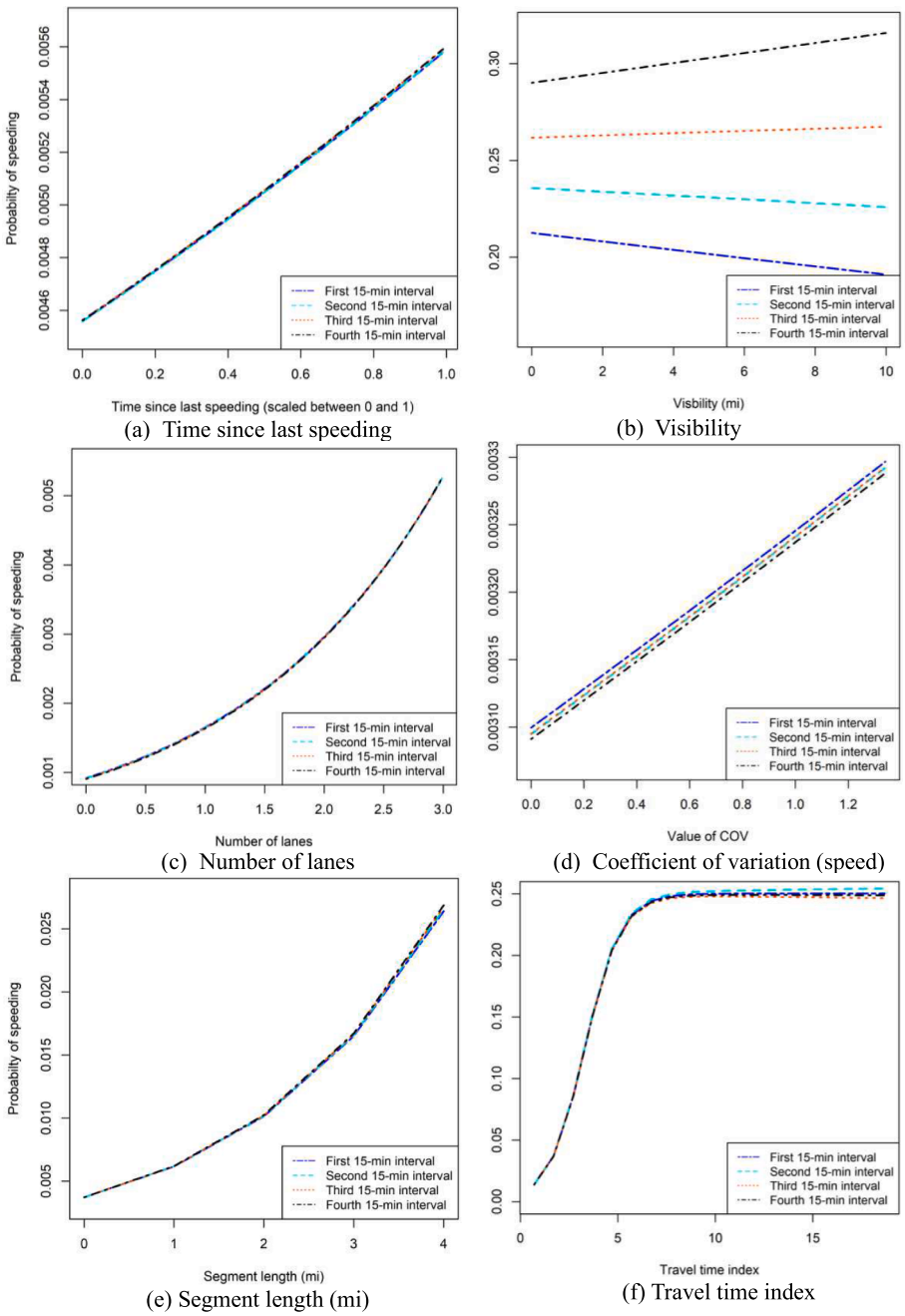

Fig. 3. Change in probability of outcomes with change in variable values.

more prone to speeding (Hamdar et al., 2016). The findings from this study align with those observations, showing that with an increase in highway segment length, the average vehicle speed, and consequently the probability of speeding, also increases, as depicted in Fig. 3(e). The nature of the plot also reveals that the rate of increase in probability intensifies as the segment length increases. For instance, the rate of increase in the probability of speeding observed on segments longer than 3 miles is much steeper than on segments that are shorter than 1 mile.

Effect of travel time index: A travel time index greater than 1 indicates a longer travel time than expected based on the operating speed limit, suggesting a congestion condition. Interestingly, Fig. 3(f) reveals that the probability of speeding rises sharply when the values of the travel time index are less than 6. However, beyond a travel time index of 6, the probability of speeding remains relatively constant. This plot suggests that the probability of speeding increases with higher travel times but only to a certain extent, and then it plateaus. This could be attributed to drivers attempting to recover lost time during mild congestion. However, when the travel time index is much higher—indicating more congestion—the probability of speeding plateaus. This implies that during periods of high congestion, when vehicles are very close together, such as bumper to bumper, drivers can no longer travel at higher speeds.

9. Conclusion

A well-functioning transportation system requires significant highway construction and maintenance to ensure efficiency. However, increased construction activities also expose workers to hazardous traffic conditions, which raises the risk of crashes. Despite implementing safety equipment, measures, laws, and policies, ensuring driver compliance with safety measures remains a concern. Enforcing reduced speed limits in work zones proves challenging due to the inherent nature of such work zones, leading to abrupt speed changes and speed

violations that significantly contribute to crashes. Moreover, it is well-established that higher speeds are linked to more severe crashes, further amplifying the safety risks associated with work zones.

The objective of this study was to introduce a novel approach for predicting speeding, especially in work zones where it poses a significant threat to road users. Discretized duration framework was used, which allows to consider past speeding trends using historical data and real-time factors like weather, traffic flow, and highway characteristics to estimate the likelihood of speeding in discrete time intervals. Using this modeling framework, we can forecast future speeding behavior by treating these time intervals as choice alternatives in a MNL model. We focused on speeding events on I-65 in Robertson County, Tennessee, to test this approach. We utilized traffic speed and speeding identification data from INRIX, highway characteristics from ETRIMS, and weather data from LCD. The model successfully identified major contributors to speeding and demonstrated reasonably accurate predictive abilities, as evaluated using metrics like PTP, Specificity, and Sensitivity. The average value of PTP indicates that the model can predict speeding within 61% of its time of occurrence. With Specificity at 0.74 and Sensitivity at 0.26, the model shows low false-positives and high false negatives. Overall, the model predictions suggest that transportation agencies can implement it to predict speeding events in real-time with a fair degree of accuracy.

While there are data-driven approaches to predicting speeding, such as machine learning, these are not readily transferrable. In contrast, the duration-based approach is grounded in an econometric framework that is readily applicable to any location. Additionally, the framework enables causal analysis through coefficients and marginal/elasticity effects. From a utility perspective, the combined ability of causal and real-time predictive analysis using the duration-based framework can be of great value to transportation planners, agencies, and safety officials. This has become even more relevant in recent years, considering Vision Zero initiatives. The causal analysis can help quantify the impact of various factors on speeding through coefficients, elasticity effects, and marginal effects. Simultaneously, predictive analysis can be of great value to agencies with limited resources for advanced planning and deployment of resources at critical segments. For instance, adverse conditions identified from causal analysis can aid agencies in remaining prepared in advance. Meanwhile, predictive analysis can help identify critical segments and prioritize the deployment of limited speed enforcement measures.

It is also important to highlight some limitations of the current study. First, reformulating speeding events using binary variables for speeding identification resulted in a prevalence of 0 s over 1 s in the data, leading to a low value of Sensitivity. Future research could explore ways to improve model performance by addressing this class imbalance, such as using techniques like Synthetic Minority Oversampling Technique (SMOTE). In addition, the effects of different work zone enforcement techniques and strategies can be further studied. This would be helpful to identify which measures are more effective than others for particular work zone setups. Finally, it is important to note that the effect of predictors might differ with work zone types. For instance, a difference in the model can be expected in left lane closure versus middle lane closure, and stationary versus mobile work zones. These factors must be taken into account in future analyses.

CRediT authorship contribution statement

Diwas Thapa: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Visualization, Writing – review & editing. **Sabyasachee Mishra:** Conceptualization, Supervision, Funding acquisition. **Asad Khattak:** Conceptualization, Supervision, Funding acquisition. **Muhammad Adeel:** Formal analysis, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

This research was funded by the Tennessee Department of Transportation (TDOT), National Science Foundation award # 2222699 and the Center for Transportation Innovations in Education and Research (C-TIER) at the University of Memphis. The views expressed here are solely those of the authors and do not necessarily represent that of the funding agencies.

References

Algomaiah, M., Li, Z., 2022. Enhancing Work Zone Capacity by a Cooperative Late Merge System Using Decentralized and Centralized Control Strategies. J. Transp. Eng., Part a: Systems 148 (2). https://doi.org/10.1061/JTEPBS.0000632.

Bashir, S., Zlatkovic, M., 2021. Assessment of Queue Warning Application on Signalized Intersections for Connected Freight Vehicles. Transp. Res. Record: J. Transp. Res. Board 2675 (10), 1211–1221. https://doi.org/10.1177/03611981211015247.

Benekohal, R. F., Hajbabaie, A., Medina, J. C., Wang, M.-H., & Chitturi, M. V. (2010). SPEED PHOTO-RADAR ENFORCEMENT EVALUATION IN ILLINOIS WORK ZONES (FHWA-ICT-10-064). Illinois Department of Transportation.

Berthaume, A. L. (2015). Microscopic Modeling of Driver Behavior Based on Modifying Field Theory for Work Zone Application [Doctoral Dissertation, University of Massachusetts Amherst]. https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1328&context=dissertations_2.

Cestac, J., Paran, F., Delhomme, P., 2011. Young drivers' sensation seeking, subjective norms, and perceived behavioral control and their roles in predicting speeding intention: How risk-taking motivations evolve with gender and driving experience. Saf. Sci. 49 (3), 424–432. https://doi.org/10.1016/j.ssci.2010.10.007.

Cheng, Z., Lu, J., Zu, Z., Li, Y., 2019. Speeding Violation Type Prediction Based on Decision Tree Method: A Case Study in Wujiang, China. J. Adv. Transp. 2019, 1–10. https://doi.org/10.1155/2019/8650845.

Debnath, A.K., Blackman, R., Haworth, N., 2015. Common hazards and their mitigating measures in work zones: A qualitative study of worker perceptions. Saf. Sci. 72, 293–301. https://doi.org/10.1016/j.ssci.2014.09.022.

Dinh, D.D., Kubota, H., 2013. Speeding behavior on urban residential streets with a 30km/h speed limit under the framework of the theory of planned behavior. Transp. Policy 29, 199–208. https://doi.org/10.1016/j.tranpol.2013.06.003.

Dissanayake, S., Akepati, S.R., 2009. Identification of Work Zone Crash Characteristics. Federal Highway Administration. https://intrans.iastate.edu/app/uploads/2018/08/Dissanayake_WZCrashChar.pdf.

Elliott, M.A., Thomson, J.A., 2010. The social cognitive determinants of offending drivers' speeding behaviour. Accid. Anal. Prev. 42 (6), 1595–1605. https://doi.org/10.1016/j.aap.2010.03.018.

Federal Highway Administration, 2023. FHWA Work Zone Facts and Statistics. Work Zone Management Program. https://ops.fhwa.dot.gov/wz/resources/facts_stats.htm.

Federal Highway Administration. (2009). Manual on Uniform Traffic Control Devices (MUTCD). https://mutcd.fhwa.dot.gov/.

Flannagan, C. A., Selpi, Baykas, P. B., Leslie, A., Kovaceva, J., & Thomson, R. (2019). Analysis of SHRP2 Data to Understand Normal and Abnormal Driving Behavior in Work Zones (FHWA-HRT-20-010). Federal Highway Administration. https://rosap.ntl.bts.gov/view/dot/48835.

Forward, S.E., 2009. The theory of planned behaviour: The role of descriptive norms and past behaviour in the prediction of drivers' intentions to violate. Transport. Res. F: Traffic Psychol. Behav. 12 (3), 198–207. https://doi.org/10.1016/j.trf.2008.12.002.

Gan, H., Wei, J., Wang, G., 2021. A generic work zone evaluation tool driven by a macroscopic traffic simulation model. Int. J. Mob. Commun. 19 (1), 1. https://doi.org/10.1504/IJMC.2021.111884.

Hamdar, S.H., Khoury, H., Zehtabi, S., 2016. A simulator-based approach for modeling longitudinal driving behavior in construction work zones: Exploration and assessment. SIMULATION 92 (6), 579–594. https://doi.org/10.1177/0037549716644515.

Hou, G., Chen, S., 2019. An Improved Cellular Automaton Model for Work Zone Traffic Simulation Considering Realistic Driving Behavior. J. Phys. Soc. Jpn. 88 (8), 084001 https://doi.org/10.7566/JPSJ.88.084001.

Jovanović, D., Šraml, M., Matović, B., Mićić, S., 2017. An examination of the construct and predictive validity of the self-reported speeding behavior model. Accid. Anal. Prev. 99, 66–76. https://doi.org/10.1016/j.aap.2016.11.015.

Kashyap, A.A., Raviraj, S., Devarakonda, A., Nayak, K., Bhat, S.J., 2022. Traffic flow prediction models – A review of deep learning techniques. Cogent Engineering 9 (1), 2010510. https://doi.org/10.1080/23311916.2021.2010510.

Kong, X., Das, S., Jha, K., Zhang, Y., 2020. Understanding speeding behavior from naturalistic driving data: Applying classification based association rule mining. Accid. Anal. Prev. 144, 105620 https://doi.org/10.1016/j.aap.2020.105620.

Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., Sierra, J.E., 2022. Urban traffic flow prediction techniques: A review. Sustainable Comput. Inf. Syst. 35, 100739 https://doi.org/10.1016/j.suscom.2022.100739.

Mishra, S., Zhu, X., 2015. Corrections of Self-Selection Bias in Crash Causality Study: An Application on All-Red Signal Control. J. Transp. Saf. Secur. 7 (2), 107–123. https://doi.org/10.1080/19439962.2014.929603.

Mishra, S., Golias, M. M., & Thapa, D. (2021). Work Zone Alert Systems. Tennessee Department of Transportation. https://rosap.ntl.bts.gov/view/dot/56274.

Osman, M., Paleti, R., Mishra, S., Golias, M.M., 2016. Analysis of injury severity of large truck crashes in work zones. Accid. Anal. Prev. 97, 261–273. https://doi.org/10.1016/j.aap.2016.10.020.

Osman, M., Mishra, S., Paleti, R., 2018a. Injury severity analysis of commercially-licensed drivers in single-vehicle crashes: Accounting for unobserved heterogeneity and age group differences. Accid. Anal. Prev. 118, 289–300. https://doi.org/10.1016/j.aap.2018.05.004.

Osman, M., Paleti, R., Mishra, S., 2018b. Analysis of passenger-car crash injury severity in different work zone configurations. Accid. Anal. Prev. 111, 161–172. https://doi.org/10.1016/j.aap.2017.11.026.

Perez, M.A., Sears, E., Valente, J.T., Huang, W., Sudweeks, J., 2021. Factors modifying the likelihood of speeding behaviors based on naturalistic driving data. Accid. Anal. Prev. 159, 106267 https://doi.org/10.1016/j.aap.2021.106267.

Richard, C.M., Lee, J., Atkins, R., Brown, J.L., 2020. Using SHRP2 naturalistic driving data to examine driver speeding behavior. J. Saf. Res. 73, 271–281. https://doi.org/10.1016/j.jsr.2020.03.008.

Scott-Parker, B., Hyde, M.K., Watson, B., King, M.J., 2013. Speeding by young novice drivers: What can personal characteristics and psychosocial theory add to our understanding? Accid. Anal. Prev. 50, 242–250. https://doi.org/10.1016/j.aap.2012.04.010.

Shaer, A., Talebian, A., Mishra, S., 2024. Informing the Work Zone Safety Policy Analysis: Reconciling Multivariate Prediction and Artificial Neural Network Modeling. J. Transp. Eng., Part a: Systems 150 (2), 04023137.

Simons-Morton, B.G., Ouimet, M.C., Chen, R., Klauer, S.G., Lee, S.E., Wang, J., Dingus, T. A., 2012. Peer influence predicts speeding prevalence among teenage drivers. J. Saf. Res. 43 (5–6), 397–403. https://doi.org/10.1016/j.jsr.2012.10.002.

Thapa, D., Mishra, S., 2021. Using worker's naturalistic response to determine and analyze work zone crashes in the presence of work zone intrusion alert systems. Accid. Anal. Prev. 156, 106125 https://doi.org/10.1016/j.aap.2021.106125.

Thapa, D., Paleti, R., Mishra, S., 2022. Overcoming challenges in crash prediction modeling using discretized duration approach: An investigation of sampling approaches. Accid. Anal. Prev. 169, 106639 https://doi.org/10.1016/j.aap.2022.106639.

Thapa, D., Mishra, S., Velaga, N.R., Patil, G.R., 2024. Advancing proactive crash prediction: A discretized duration approach for predicting crashes and severity. Accid. Anal. Prev. 195, 107407.

Wang, X., Katz, R., Dong, X.S., 2018. Fatal Injuries at Road Construction Sites among Construction Workers [Quarterly]. Center for Construction Research and Training. https://www.cpwr.com/wp-content/uploads/publications/publications_Quarter2-QDR-2018.pdf.

Yu, B., Chen, Y., Bao, S., 2019. Quantifying visual road environment to establish a speeding prediction model: An examination using naturalistic driving data. Accid. Anal. Prev. 129, 289–298. https://doi.org/10.1016/j.aap.2019.05.011.

Zhang, K., Hassan, M., 2019. Identifying the Factors Contributing to Injury Severity in Work Zone Rear-End Crashes. J. Adv. Transp. 2019, 1–9. https://doi.org/10.1155/2019/4126102.

Zhao, G., Wu, C., Qiao, C., 2013. A Mathematical Model for the Prediction of Speeding with its Validation. IEEE Trans. Intell. Transp. Syst. 14 (2), 828–836. https://doi.org/10.1109/TITS.2013.2257757.