Bicycle Naturalistic Data Collection

CONNECTED VEHICLE/INFRASTRUCTURE UNIVERSITY TRANSPORTATION CENTER (CVI-UTC)

Bicycle Naturalistic Data Collection

Prepared for the Research and Innovative Technology Administration (RITA);
U.S. Department of Transportation (US DOT)

Grant Project Title:
Advanced Operations Focused on Connected Vehicles/Infrastructure (CVI-UTC)

Consortium Members:
Virginia Tech Transportation Institute (VTTI),
University of Virginia (UVA) Center for Transportation Studies,
and Morgan State University (MSU).

Submitted by:

Virginia Tech Transportation Institute
3500 Transportation Research Plaza
Blacksburg, VA 24061

Program Director:
Dr. Thomas Dingus
Program Director, Connected Vehicle/Infrastructure
University Transportation Center
Director, Virginia Tech Transportation Institute
Professor, Department of Biomedical Engineering and Mechanics
at Virginia Tech
tdingus@vtti.vt.edu
(540) 231–1501

Report Authors:
Mohammed Elhenawy
Postdoctoral, Center for Sustainable Mobility
Virginia Tech Transportation Institute
elhenawy@vt.edu
(540) 231-0278

Arash Jahangiri
arashj@vt.edu
(540) 200-7561

Hesham Rakha
Samuel Reynolds Pritchard Professor of Engineering, Dept. of Civil & Environmental Engineering
Courtesy Professor, Bradley Dept. of Electrical and Computer Engineering
Director, Center for Sustainable Mobility
Virginia Tech Transportation Institute
hrakha@vt.edu
(540) 231-1505

DUNS: 0031370150000 EIN: 54-6001805

Grant Funding Period: January 2012 – July 2016
Final Research Reports

June 15, 2016

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Connected Vehicle/Infrastructure UTC

The mission statement of the Connected Vehicle/Infrastructure University Transportation Center (CVI-UTC) is to conduct research that will advance surface transportation through the application of innovative research and using connected-vehicle and infrastructure technologies to improve safety, state of good repair, economic competitiveness, livable communities, and environmental sustainability.

The goals of the Connected Vehicle/Infrastructure University Transportation Center (CVI-UTC) are:

- Increased understanding and awareness of transportation issues
- Improved body of knowledge
- Improved processes, techniques and skills in addressing transportation issues
- Enlarged pool of trained transportation professionals
- Greater adoption of new technology

Abstract

Recently, bicycling has drawn more attention as a sustainable and eco-friendly mode of transportation. Between 2000 and 2011, bicycle commuting rates in the United States rose by 80% in large bicycle friendly cities (BFCs), by 32% in non-BFCs, and overall by 47%. On the other hand, about 700 cyclists are killed and nearly 50,000 are injured annually in bicycle–motor vehicle crashes in recent years in the United States.

More than 30% of cyclist fatalities in the United States from 2008 to 2012 occurred at intersections, and up to 16% of bicycle-related crashes were due to cyclists' violations at intersections. In light of these statistics, this project focused on investigating factors that affect cyclist behavior and predicting cyclist violations at intersections. Naturalistic cycling data were used to assess the feasibility of developing cyclist violation prediction models. Mixed-effects generalized regression model is used to analyze the data and identify the significant factor affecting the probability of violations by cyclists. At signalized intersections, right turn, side traffic and opposing traffic are statistically significant factors affecting the probability of red light violation. At stop-controlled intersections, the presence of other road users, left turn, right turn and warm weather are statistically significant factors affecting the probability of violations.

Violation prediction models were developed for stop-controlled intersections based on kinetic data measured as cyclists approached the intersection. Prediction error rates were 0% to 10%, depending on how far from the intersection the prediction task was conducted. An error rate of 6% was obtained when the violating cyclist was at a time-to-intersection of about 2 seconds, which is sufficient for most motor vehicle drivers to respond.

Acknowledgments

The authors recognize the support that was provided by a grant from the U.S. Department of Transportation – Research and Innovative Technology Administration, University Transportation Centers Program, and the Virginia Tech Transportation Institute.

Table of Contents

List of Figures

List of Tables

Background

Cycling is a growing, eco-friendly mode of travel that will likely attract more participants in the future. From 2000 to 2011, bicycle commuting rates in the U.S. increased by 80% in large bicycle friendly cities (BFCs), by 32% in non-BFCs, and by a national average of 47% [1]. Where BFC is one that welcomes cyclists with trails, bike lanes, share the road campaigns, organized rides, Bike to Work Day events, and so much more. A rich matrix of options that recognizes an area's unique resources, the BFC application evaluates how a community encourages people to bike for transportation and recreation through the "Five Es": engineering, education, encouragement, enforcement, and evaluation. Due to this growth of cycling, there is clearly a need to address bicycling safety. According to a National Highway Traffic Safety Administration (NHTSA) report, 742 cyclists were killed and about 48,000 were injured in crashes between bicycles and motor vehicles in the United States in 2013 [2]. Bicycle fatalities represent 2% of total traffic fatalities, but trips made by bicycles constitute only 1% of all trips. Such disproportionate representation of cyclists in the crash and injury statistics merits concern, and as cycling grows in popularity, more knowledge regarding the behaviors and factors that contribute to this risk is needed.

Research shows that intersections are critical for cyclists. From 2008 to 2012 in the U.S., on average more than 30% of cyclists' fatalities occurred at intersections [3]. Cyclist violations at intersections resulted in up to 16% of bicycle-related crashes [4, 5].

Many different factors influence cyclist violation behavior at intersections, such as age [6-9], gender [6, 7, 9, 10], direction of travel [7, 11], the presence of other road users [6, 7, 11], signal timing [8], intersection type [8], helmet use [8, 9], detector failure [7], design characteristics [12], and alcohol or drug consumption [9]. Violation rates can also be significantly different for different countries (Table 1).

Table 1. Violation Rates in Different Countries

Country	Violation rate	Reference
Australia	7%-9%	[11]
China	56%	[7]
Taiwan	21%	[8]
Brazil (self-reported)	38.4%	[11]

Because drivers and cyclists may fail to obey traffic rules at both signalized and stop-controlled intersections, the problem is how to prevent or mitigate intersection-related crashes that involve bicycles. Failures to comply need to be identified before they occur so actions can be taken to alleviate the consequences. In this research, naturalistic cycling data were used to assess the feasibility of developing cyclist violation prediction models at intersections.

The present research continues previous work that proposed a system architecture that incorporates naturalistic cycling data for developing cyclist violation prediction models (Figure 1) [13]. This figure shows system architecture for developing cyclist violation prediction models using naturalistic data in a connected vehicle environment. Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I) technology has been a highly active area of research. However, the focus has been more on passenger cars and thus less attention has been given to bicycles as an important transportation mode. Figure 1 illustrates that bicycles instrumented with a bicycle data collection system (MiniDAS) were also incorporated into the connected vehicle environment. The MiniDAS was used to collect and extract required variables to develop violation prediction models (the MiniDAS and model development is discussed later in this report). After the violation prediction model is constructed, bicycles and other transportation modes would send their sensor data (e.g. speed, acceleration, location) to the infrastructure as depicted in Figure 1. In addition, signal phase and timing, known as SPaT would be obtained as another input to the prediction model. Violation prediction models for different modes of transport should constantly be monitoring individuals approaching the intersection. When a potential threat is predicted, different actions can be taken depending on the situation; in situations where the endangered driver (or rider) has sufficient time, a warning can be issued and sent from roadside equipment (RSE) to the driver (or rider) to respond. In cases where not enough time is available, the infrastructure can take appropriate actions by changing the signal control setting through the traffic light (e.g. providing all red clearance) as shown in Figure 1.

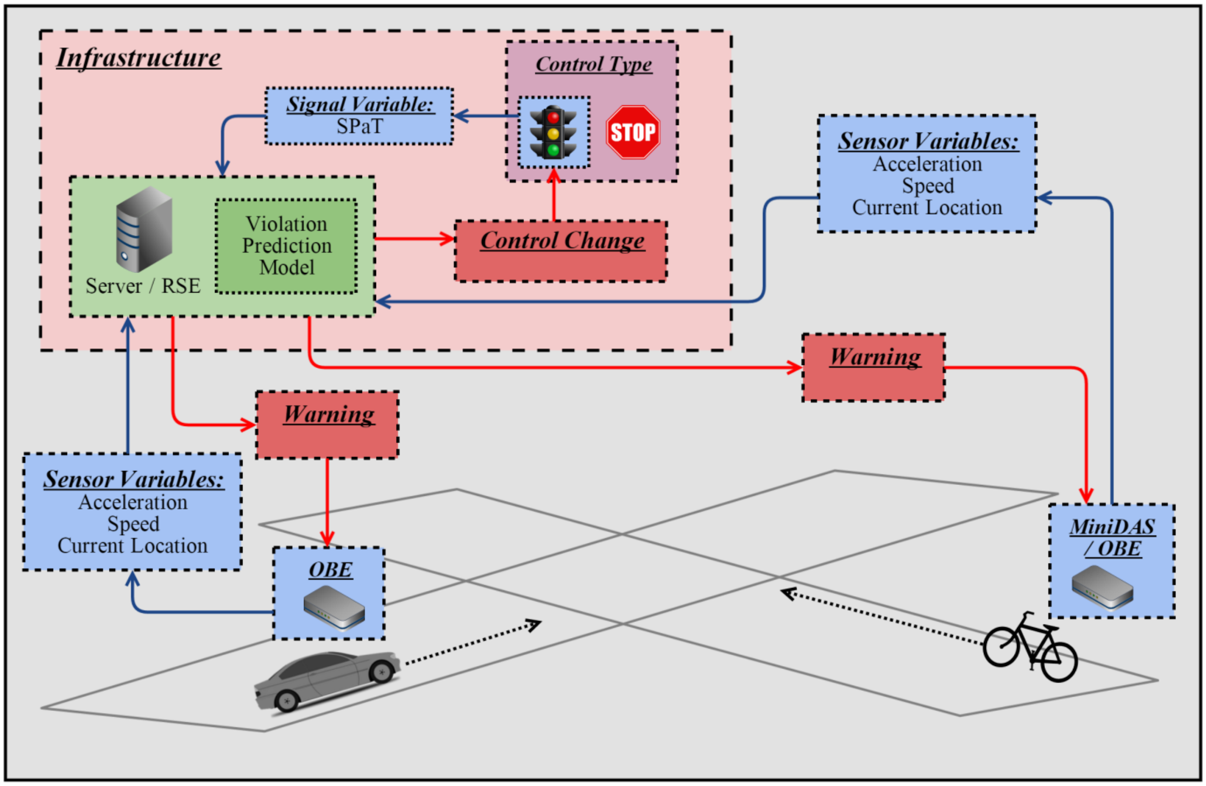


Figure 1. System architecture for intersection bicycle-car crash prediction.

The remainder of this report is organized as follows. First, a literature review summarizes relevant work on bicycle safety at intersections. Second, details are provided on the naturalistic cycling study that served as the foundation for model development. Third, models are presented that were considered and the results of the modeling. Finally, a discussion of results and conclusions are presented.

Literature Review

Research into bicycle safety at intersections can be divided into studies that investigated contributing factors and those that examined countermeasures.

Investigating Contributing Factors

Many studies assessed factors and conditions that influence the crossing behavior of cyclists at intersections. These studies are categorized below based on the data collection method employed: naturalistic data through intersection cameras, naturalistic data through instrumented bicycles, police-reported data, and surveys and interviews.

Naturalistic Data Collection through Video Cameras Installed at Intersections

Video cameras installed at intersections are able to capture how cyclists approach intersections. Since cyclists are unaware of the cameras, the behavior of cyclists is considered to be naturalistic. However, the results may not be generalizable if the data are only collected at limited locations.

Table 2 lists studies that used intersection cameras to capture the crossing behavior of cyclists. Using a single binary logistic regression analysis, Johnson et al. [11] found that the most important factor to predict red light runners was the direction of travel, in this case turning left (in Australia, traffic travels on the left side, so turning left is equivalent to turning right in the United States). By applying logistic regression, Wu, Yao, and Zhang [6] found that the following factors increase the probability of running red lights: when the rider was younger, when the rider was alone, when there were fewer riders waiting, and when there were other riders already violating the red light. Johnson Charlton, and Oxley [10] found gender to be an important factor, with males more likely to violate intersection right-of-way rules. Pai and Jou [8] used a mixed logit model to examine different factors. Factors that increased the probability of a crash were intersections with a short red-light duration, T/Y[1] intersections, riders who were pupils in uniform, riders on electric bicycles, and riders who did not wear a helmet.

Table 2. Data Collection through Video Cameras at Intersections – Summary of Past Studies

Reference	Observations	Behavior classification of riders
Johnson et al. [11]	4,225	6.9% violation rate, no further classification
Wu, Yao, and Zhang [6]	229	56% violation rate Risk-taking[a] (28%); law-obeying[b] (49%); opportunistic[c] (23%)
Johnson, Charlton, and Oxley [10]	5,420	3% morning, 11% afternoon violation rates Racers[d] (25%); runners[e] (42%); impatients[f] (33%)
Pai and Jou [8]	11,410	Risk-taking (4.7%); law-obeying (85.8%); opportunistic (9.5%)

[b] **Law-obeying:** Cyclists who would stop by obeying the red light.
[a] **Risk-taking:** Cyclists who would ignore the red light and travel through the intersection without stopping (but might slow down).
[c] **Opportunistic:** Cyclists who would first wait at red lights but would not be patient enough to wait until the green light and subsequently crossed the intersection as they found gaps between crossing traffic.
[d] **Racers:** Cyclists who encountered an amber light, accelerated, but failed to pass before the light turned to red.
[e] **Impatients:** Cyclists who initially stopped, but then could not wait until the end of the red phase.
[f] **Runners:** Cyclists who rode through the red phase without stopping.

The abovementioned terms were defined by authors of different papers as presented in Table 2. However, there are some similarities as follows; Opportunistic and Impatient riders appear to be the same. Also, it appears that the behavior of runners and racers can both be considered as risk-taking.

Naturalistic Data Collection through Instrumented Bicycles

Some studies instrument bicycles with cameras and/or sensors to collect data for cyclists' entire trips. No special instructions concerning how to ride, when to ride, and where to ride are given to the participants. Thus, the data collected not only reflect realistic cyclist behavior, but also lead to better generalizations because many different locations are included. A few studies have

[1] T intersections have three legs, usually at a 90-degree angle with one another; Y intersections have three legs, usually with one leg at an angle less than 90 degrees, much like the letter "Y".

adopted naturalistic cycling data collection techniques similar to the 100-Car Naturalistic Driving Study [17], Integrated Vehicle-Based Safety System (IVBSS) [18], and euroFOT [19]. Table 3 summarizes these studies.

Table 3. Naturalistic Cycling Data Collection – Summary of Past Studies

Reference	Equipment	Participant criteria
Dozza, Werneke, and Fernandez [20]; Dozza and Fernandez [21]	2 cameras, Global Positioning System (GPS), 3-axis accelerometer, 3-axis gyroscope, 3-axis magnetometer, 2 pressure brake sensors, speed sensor	Age between 25 and 70, rode more than 40 minutes per day on weekdays, bicycle was transportation mode used for commuting, and participants were asked not to carry children on the bicycle during the study
Johnson et al. [22]	Video cameras	Age over 18, regularly commuted by cycling to and from work, rode the majority (70%) of trips on paved roads during commutes, could collect 12 hours of data over 4 weeks per bicycle
Gustafsson and Archer [23]	GPS and cameras	Commuter cyclists rode 17 different major cycle routes. Participants were required to ride the major part of their trips during morning (07:00–09:00) and afternoon (16:00–18:00) peak hours.

To identify critical events such as crashes and near-crashes, Dozza, Werneke, and Fernandez [20] and Dozza and Fernandez [21] designed and installed a human-machine interface on the handlebar so that the cyclist could record the time of critical events using a push button. By analyzing the coherence between video and collected signals, the study showed that naturalistic cycling data can capture safety-critical situations in the same way that naturalistic driving data can. This would contribute to modeling and quantifying the events leading to a critical situation. However, regarding the method applied to capture critical events (i.e. using push buttons), giving participants extra tasks beyond riding could impact riding behavior.

Johnson et al. [22] identified sideswipe as the most frequent event type (40.7%). The majority of events occurred at intersections or intersection-related locations[2]. Motor vehicle drivers were determined to be the violators in most of the cases. However, Johnson et al. [22] did not use sensors in their data collection system to obtain and analyze kinetic data.

Gustafsson and Archer [23] defined safety problems as hard-braking, swerving, or accelerating to avoid a collision. For these safety problems, factors such as involvement of other road users, event location, responsible party, and event frequency were identified. Although it was shown that the most unsafe cycle-car conflicts took place at intersections, no violations at intersections

[2] *at-intersection accident:* An at-intersection accident is a traffic accident in which the first harmful event occurs within the limits of an intersection; *intersection-related accident:* An intersection-related accident is a traffic accident in which the first harmful event (1) occurs on an approach to or exit from an intersection, and (2) results from an activity, behavior, or control related to the movement of traffic units through the intersection.

were reported. Also, the participants were given strict instructions to obey traffic rules at intersections, which negatively impacts the collection of realistic riding behavior data.

Police-Reported Data

Schepers et al. [12] used police-reported data to study two types of bicycle-motorist crashes at stop-controlled intersections. Crashes in which the cyclist had the priority were classified as type I, and the crashes in which the motorist had the priority were classified as type II. They employed negative binomial regression models to study factors such as intersections with two-way bicycle tracks, well-marked, reddish-colored bicycle crossings, the presence of raised bicycle crossings (e.g., speed hump), and other-speed reducing measures. Results from their negative binomial regression models showed that more type I crashes were seen at intersections with two-way bicycle tracks, and well-marked, reddish-colored bicycle crossings. Also, the presence of raised bicycle crossings (e.g. speed hump) and other speed-reducing measures were associated with less type I crashes. Further, intersections with bicycle track approaches deflected between 2 and 5 meters away from the main road corresponded to less type I crashes. However, there were no road factors significantly affecting type II cashes.

Martínez-Ruiz et al. [9] applied logistic regression and multinomial regression analyses to study 19,007 collisions between a bicycle and another vehicle using police-reported crash data. Younger cyclists (age 10 to 19 years), male gender, alcohol or drug consumption, and non-helmet use were identified as factors increasing the risk of crashes.

Surveys and Interviews

A survey study was carried out by Lacherez et al. [24] to investigate visibility factors that impact crashes. One hundred eighty-four cyclists who had been involved in crashes with motor vehicles were surveyed. Stop-controlled intersections and signalized intersections were found to be the third (~17%) and fourth (~9%) most common crash sites.

Red light infringement was examined in another survey study [7]. Out of 2,061 cyclists in the study, 37.3% reported that they ran against a red light. Participants provided the following reasons for their violations: turning left (32%); inductive loop detector failed to detect their bike (24.2%); absence of other road users (16.6%); at a pedestrian crossing (10.7%); and "other" (16.5%). Based on a multinomial logistic regression analysis, males and younger participants were associated with higher violation probability.

Examining Countermeasures

A limited number of studies focused on introducing countermeasures to reduce bicycle-related crashes at intersections. Phillips et al. [14] examined the impacts of having a cycle path at a cycle-road intersection based on the change in the number of yielding and conflict events after introducing the path. In their study at a Norwegian road–cycle path intersection, 57 hours of video data were taken to examine the effects of a cycle path. Yielding and conflict events were assessed 2 months, 4 years, and 10 years following the introduction of the road-cycle path, which resulted in a significant decrease in overall conflict levels after 4 years and further decrease after 10 years. However, the conflict reduction between 4 and 10 years was small. In other words, car-bicycle yielding events were increased that led to fewer number of conflict events. Zhang and Wu [15] adopted logistic regression and analysis of variance to assess the effects of having sunshields for cyclists at two intersections across the city of Hangzhou, China. This was an observational study in which two video cameras were used to examine the crossing behavior of cyclists; 2,477 riders were captured from the video recordings. Logistic regression and analysis of variance were applied to understand how the sunshield as a factor influenced red light running behavior. It was found that the red light infringement was reduced when having the sunshield both on sunny and cloudy days with the positive effect larger during sunny weather compared to cloudy weather. More specifically, the results showed that bicyclists were 1.376 times more likely to violate a red light when sunshields were not in place at an intersection than with sunshields in place. Räsänen, Koivisto, and Summala [16] conducted a before-after study to evaluate the impact of a regulation change on bicycle-related crashes. They found that cyclists' movement direction and their location are significant factors that affect cyclists' crossing behavior. As a result, depending on these characteristics, changes in priority regulations have different effects.

Naturalistic Cycling Experiment

The present research used naturalistic data collected from instrumented bicycles. In a manner similar to naturalistic driving studies performed with motor vehicles (e.g., the "100-Car study" [17]), bicycles were instrumented and given to the participants to ride. The participants were asked to ride the bicycles as they normally would without being provided any special instructions.

The naturalistic cycling study was conducted in three steps: (1) a prescreening to identify the participants who encountered the most intersections on their rides; (2) data collection from the instrumented bicycles; and (3) data analysis.

Prescreening

Because the objective of the present study was to investigate cyclist violation behavior at intersections, potential participants were prescreened to understand their weekly trip patterns by bicycle. To identify the cyclists who encountered the most intersections on their rides, a series of questions, which are shown in Appendix A, were asked over the phone. Other factors used to identify eligible participants included the following:

- Participants must make most of their trips on paved roads rather than on sidewalks and bicycle trails.
- Participants must commute or travel by bicycle at least three times per week on average in the Blacksburg, Christiansburg, and/or Radford, Virginia, areas.

- Participants were not allowed to transport children by the instrumented bicycle.
- Participants must be 18–30 or 45–65 years of age from both genders.

Equipment and Data Collection

The bicycles were instrumented with the MiniDAS data acquisition system (DAS) developed by the Virginia Tech Transportation Institute (VTTI) (Figure 2). The MiniDAS had two cameras (one for forward roadway view from bike and the other for rider view) and sensors such as an accelerometer, gyroscope, and GPS. A removable battery, which needed to be charged by the participants using a battery charger, was placed in the water bottle cage. The bicycles were hybrid models (Trek 7.2 FX) available in three sizes: small (15 in.), medium (17.5 in.), and large (20 in.). Participants were responsible for making sure the battery was charged and that the MiniDAS was turned on while riding.

Initial analysis of sample data showed that the acceleration data had too much noise (due to possible DAS vibration when riding) and was difficult to work with. Therefore, a speed sensor was added. The speed sensor measured distance as the wheels rolled, and the speed and acceleration data were then derived from the distance data.

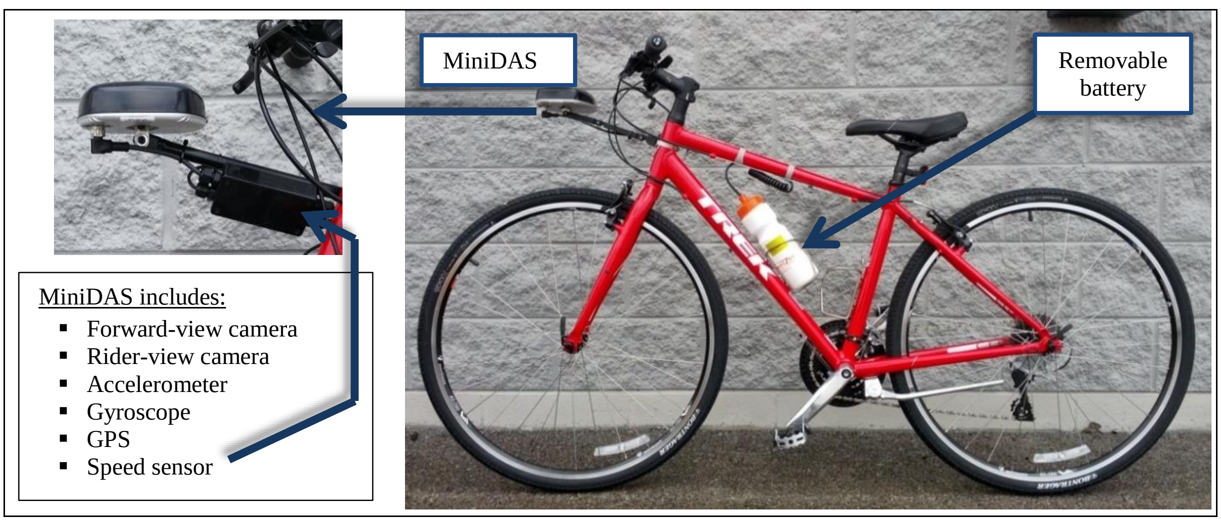


Figure 2. Naturalistic cycling data collection system.

Data Reduction

Hawkeye software, a data visualization tool, was used for data reduction. An "event" was defined as crossing an intersection. Multiple variables were extracted for each event. Table 4 lists the variables obtained through data reduction for all events. The data set used in the present research included data from 20 participants.

Table 4. Variables Obtained from Data Reduction

Variable
Time 1 (morning/noon/evening)
Time 2 (weekend/weekday)
Road slope (uphill/downhill/flat)
Movement type (right/through/left)
Presence of other road users (side/opposing/front/adjacent)
Weather 1 (warm/cool)
Weather 2 (cloudy/rainy/clear)
Yellow onset (signal stage)
Red onset (signal stage)

Results

After conducting data reduction for 20 participants, the naturalized data set was separated into two subsets: 251 crossings at signal-controlled intersections and 2,024 crossings at stop-controlled intersections.

A mixed-effects generalized regression model was used to figure out the significant factors that explain the variation in cyclist behavior at signal-controlled and stop-controlled intersections. A mixed-effects generalized regression model was used because the analyzed data set had many observations from each participant, which violates the independence assumption of a fixed-effect logistic regression.

The results for signal-controlled intersections and stop-controlled intersections are presented in separate sections below.

Signal-Controlled Intersections

A total of seven different violation types were identified for signal-controlled intersections, as shown in Table 5 and Table 6. Based on traffic rules for cyclists in Virginia as presented on the Virginia Department of Transportation (VDOT) website,

> "*Bicyclists must obey all traffic signs, signals, lights, and markings. However, under certain circumstances bicyclists, motorcycles, and mopeds may proceed through a solid red light at an intersection. A driver or rider may proceed through the intersection on a steady red light only if the driver or rider complies with all five provisions listed below:*
>
> > *1. Comes to a full and complete stop at the intersection for two complete cycles of the traffic light or for two minutes, whichever is shorter*
> >
> > *2. Exercises due care as provided by law*
> >
> > *3. Otherwise treats the traffic control device as a stop sign*
> >
> > *4. Determines that it is safe to proceed*

5. Yields the right of way to the driver of any vehicle approaching on such other highway from either direction".

In almost all observations the study participants did not wait two minutes, and thus were considered "violators". In few cases where the participant waited longer (i.e. close to two minutes), the participants were also categorized as "violators". This categorization of cyclist behavior occurred because it was not feasible to observe the video data from each participant to ensure that they had complied with all five VDOT traffic rules for cyclists.

Table 5. Cyclist Violation Types Part 1

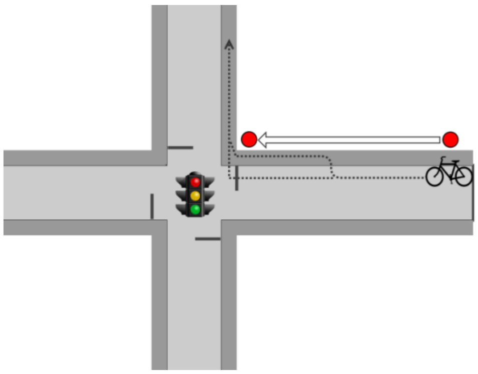
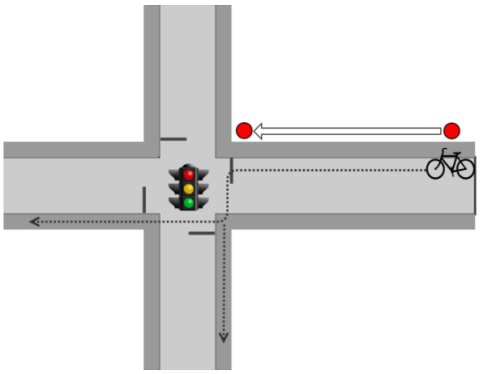
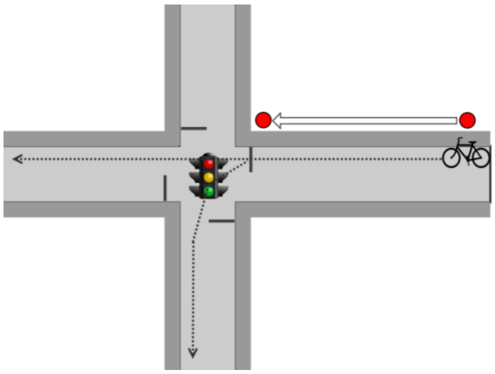
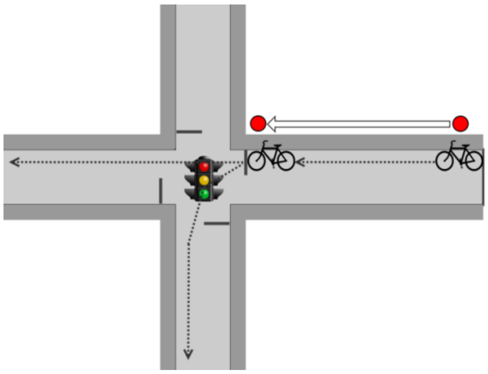
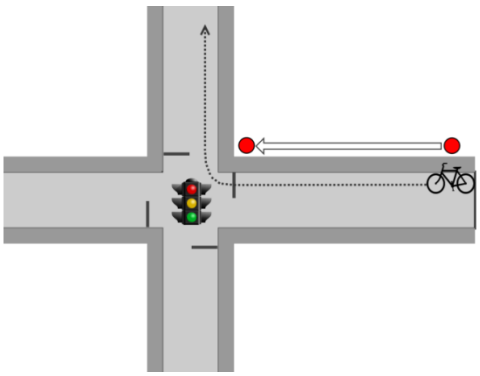
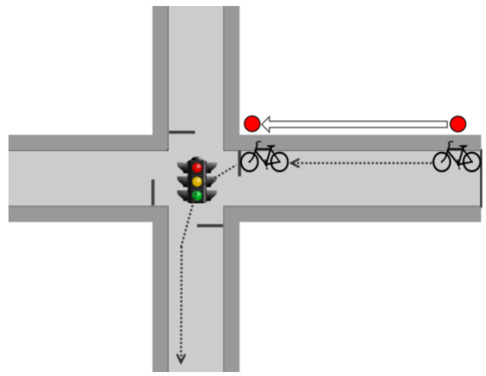
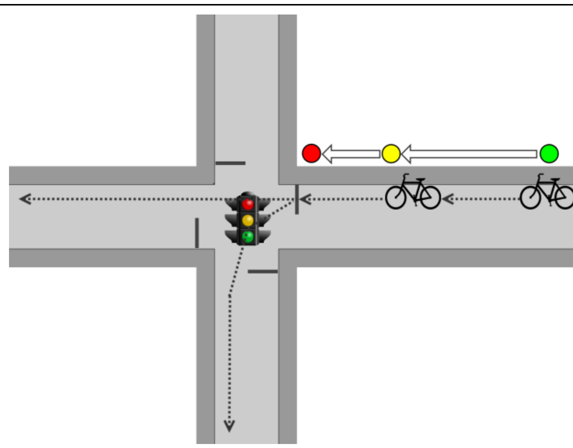
In the following violation types, the traffic light is already red when the cyclist approaches the intersection (i.e., the cyclist does not see the yellow and red onsets or the cyclist sees the red onset when he/she is too far from the intersection).	
The cyclist goes on the sidewalk to the right side of the street and makes a right turn on the sidewalk to avoid waiting at the traffic light (13.75%).	The cyclist goes to the pedestrian crossing and makes a left turn to the sidewalk. Subsequently, continues on the sidewalk or makes a right turn on the pedestrian crossing (5%).
	
The cyclist goes through the intersection without stopping when no other road users with a potential conflict are present at the intersection (17.5%).	The cyclist stops first, but after a while, he/she proceeds while the light is still red (6.25%).
	

Table 6. Cyclist Violation Types Part 2

<table>
<tr><th colspan="2">The remaining violation types occurred as a result of being in the dilemma zone as follows.</th></tr>
<tr><td>The cyclist makes a right turn without stopping (40%).</td><td>The cyclist intends to turn left, stops, and waits for the light to change to green. However, the cyclist is not detected by the camera at the intersection and thus the light stays red. Consequently, the cyclist violates the red light as he/she finds a gap (7.5%).</td></tr>
<tr><td></td><td></td></tr>
<tr><td colspan="2">The cyclist sees the yellow onset and decides to proceed. However, the light turns red before reaching the stop bar and the cyclist goes through without stopping (10%).</td></tr>
<tr><td colspan="2"></td></tr>
</table>

In order to find the significant factors that explain cyclist behavior at signalized intersections, a mixed-effects generalized regression model was used. Based on the parameter estimation of the fixed effect, right turns and the presence of other users (i.e., to the side and front) were found to be significant factors (Table 7). This shows that a cyclist is more likely to violate a red light when making right turns. The results also show that the probability of red light violation decreases when there is side traffic or opposing traffic.

Table 7. Mixed-Effects Generalized Regression Model – Signalized Intersections

Fixed effect	Estimate	Std Error	Wald Chi Square	Prob > Chi Square
Intercept	-0.428588	0.1479789	8.3884173	0.0038*
Morning peak	0.5919818	4.1069203	0.0207771	0.8854
Noon peak	-2.123522	3.4130388	0.3871072	0.5338
Evening peak	-1.71912	3.818989	0.2026356	0.6526
Weekend	-0.658993	2.6015948	0.0641626	0.8000
Uphill	-1.682128	2.9916978	0.3161424	0.5739
Downhill	-4.43822	3.2734462	1.8382609	0.1752
Right turn	12.040572	2.5549167	22.2096	<.0001*
Left turn	1.5544359	3.0739644	0.2557101	0.6131
Warm weather	-0.141255	3.1585401	0.002	0.9643
Cold weather	-0.386963	2.1783913	0.0315548	0.8590
Gender	-2.1769	8.0621641	0.0729077	0.7871
Cloudy	-1.182723	2.6363766	0.201257	0.6537
Rainy	1.8662414	2.3841726	0.6127174	0.4338
Snowy	-2.166636	3.3099932	0.4284674	0.5127
Age	2.2903596	14.589569	0.0246446	0.8753
Side traffic	-6.887538	2.4547243	7.8726798	0.0050*
Opposing traffic	-7.787741	2.6207319	8.8303466	0.0030*
Front traffic	-0.560723	2.4781053	0.0511986	0.8210
Adjacent traffic	-0.010431	2.6166964	0.0000159	0.9968

Stop-Controlled Intersections

Based on the parameter estimation of the fixed effect, the following factors are significant for stop-controlled intersections: right turns, left turns, warm weather, and the presence of other users (i.e., to side, front, opposing, or adjacent). Table 8 shows the results of the statistical analysis for all factors. The time of the day, the day of the week, age, and gender were found to be not significant.

Table 8. Mixed-Effects Generalized Regression Model – Stop Intersections

Fixed effect	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare
Intercept	-1.285871	0.066046	379.05473	<.0001*
Morning peak	4.7119668	4.1252868	1.3046564	0.2534
Noon peak	-1.713809	3.5965811	0.2270624	0.6337
Evening peak	0.632353	4.0342465	0.0245694	0.8754
Weekend	-2.323843	2.4601217	0.8922781	0.3449
Uphill	-3.700012	2.9009361	1.626784	0.2021
Downhill	-4.799859	2.7700881	3.002408	0.0831
Right turn	6.9134651	3.1063536	4.9532446	0.0260*
Left turn	-5.590305	2.8285343	3.9061433	0.0481*
Warm weather	-8.712035	4.0593821	4.6059514	0.0319*
Cold weather	-1.691295	3.3912124	0.2487302	0.6180
Gender	-6.186167	11.748152	0.2772708	0.5985
Cloudy	2.518305	2.8853412	0.7617673	0.3828
Rainy	2.5431327	2.8934541	0.7725113	0.3794
Snowy	4.3065705	4.3765515	0.9682757	0.3251
Age	8.8147867	17.665864	0.2489739	0.6178
Side traffic	-25.54735	2.5191416	102.84577	<.0001*
Opposing traffic	-6.723773	2.4651386	7.4394945	0.0064*
Front traffic	-9.873646	2.2828026	18.707615	<.0001*
Adjacent traffic	-5.465376	2.4721507	4.8875396	0.0271*

Model Development

The second objective of this research was to test the feasibility of developing models to predict intersection violations by cyclists. Several machine learning algorithms were tested: multivariate logistic regression (MLR), random forest (RF), K-nearest neighbors (K-NN), and an artificial neural network (ANN). The violation prediction models were behavioral classifiers with binary responses (in this case, 1 as violation, 0 as compliance). In addition, differing values of the time window (monitoring period) during which the kinematics-related variables were computed were tested.

Multivariate Logistic Regression (MLR)

MLR is used for predicting a binary response using multiple variables. The logistic regression model calculates the probability that an observation belongs to a particular response (Equation 1). In the case of the present research, an observation is explained with multiple variables extracted from the data reduction. The binary response in this case was defined as cyclist violation (denoted by 1) versus cyclist compliance (denoted by 0) at intersections. The goal was to develop a relationship model to investigate variable impacts on the violation behavior of cyclists at intersections.

$$p(Y=1|X)=\frac{e^{\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n}}{1+e^{\beta_0+\beta_1X_1+\beta_2X_2+\cdots+\beta_nX_n}} \qquad \text{Equation 1}$$

Where,

X	$(X_1,\dots,X_n)$
$X_1,\dots,X_n$	n variables
$\beta_0,\dots,\beta_n$	Model coefficients
Y	Binary response
$p(Y=1\|X)$	Probability that a response is 1 given X

Random Forest (RF)

RF as a supervised machine-learning technique was applied to predict cyclists' violations. RF was introduced by Breiman [25], and is considered to be an ensemble learning approach based on the decision-tree method. RF creates a number of decision trees, and the outcome is obtained from averaging the results from each tree in the case of regression problems or taking the majority votes in the case of classification problems. To grow each tree, data are divided into two parts in several steps until a desirable separation between classes (the two classes in the case of the present research are violation and compliance) is achieved. Data separation in each step is carried out by a recursive binary splitting method in which different criteria can be used to split the data. The Gini index, one of the recommended approaches, was employed in this study (Equation 2) [26].

$$G=\sum_{k=1}^{K}P_k^m(1-P_k^m) \qquad \text{Equation 2}$$

Where,

$$P_k^m=\frac{1}{N^m}\sum_{x_i^m}I(y_i^m=k)$$

N^m	Number of observations received at node m
y_i^m	The response value corresponding to the observation i at node m
x_i^m	The feature vector corresponding to the observation i at node m
k	class

K-Nearest Neighbors Algorithm (K-NN)

K-NN is a naive technique that is good for classification [27]. The advantage of K-NN is that it does not require training to build the model because the training data represent the model. The simplest version of a K-NN model has the training portion of the algorithm simply store the data points of the raw training data $\mathrm{T}=\{\mathrm{x}_1,\dots,\mathrm{x}_\mathrm{n}\}$ without any processing. To classify an unseen instance, x_t , the algorithm calculates the distances from x_t to all instances in the data set, then it finds the K smallest distances (nearest neighbors). Finally, the algorithm assigns x_t to the class that has the majority of K nearest neighbors. This algorithm has two important parameters, the distance metric and the number K. A typical distance metric $d(x_t, x_i)$ is the Euclidean distance, where the value of K is typically between 5 and 20 and is set up experimentally as,

$$d(x_t, x_i) = \sqrt{(x_t - x_i)^T (x_t - x_i)} \qquad \text{Equation 3}$$

However, when the training data set is very large, the K-NN algorithm has a memory problem. There is another version of K-NN called agglomerative nearest neighbor (A-NN) that is used to overcome memory problems in K-NN algorithms. During the training phase, the A-NN algorithm clusters the training instances that have the same label using any of the well-known clustering algorithms and saves the cluster centroids for each class $C_j = \{c_{j1}, \dots, c_{jm}\}$. To classify unseen instances, x_t, the A-NN algorithm calculates the distances from x_t to all instances in C_j $\forall$ j, and then finds the K smallest distances (nearest neighbors). Finally, the algorithm assigns x_t to the class that has the majority of centroids in K. A-NN was not required because data from only 20 bicyclists was used and thus there was no memory problem.

Artificial Neural Network (ANN)

In machine learning, ANNs are used to estimate or approximate unknown linear and nonlinear functions that depend on a large number of inputs. ANNs can compute values or return labels using inputs.

An ANN consists of several processing units, called neurons, which are arranged in layers. In this study, a multi-layered, feed-forward ANN was used, which is commonly used for classification analysis. In multi-layered, feed-forward ANNs, the neurons are connected by directed connections, which allows information flow from the input layer to output layer. A neuron k at layer m receives an input x_j from each neuron j at layer $m - 1$. The neuron adds the weighted sum of its inputs to a bias term, then applies the whole thing to a transfer function and passes the result to the downstream layer. In general, the ANN requires a definition of the number of layers, the number of neurons in each layer, and the neuron's transfer function. Given a training data set, the ANN uses a learning algorithm such as back-propagation to learn the weights and biases for each single neuron [28].

Monitoring Period

In a manner similar to the driver violation prediction models developed by Jahangiri, Rakha, and Dingus [29], the monitoring periods were defined so that the naturalistic study data could be used to develop the prediction models.

Figure 3 illustrates the variables used to define a monitoring period for modeling violation prediction models. The violating bicycle on one approach is going to violate the intersection. As a result, the endangered vehicle/bicycle on a conflicting approach is at risk of a right-angle crash. The monitoring period, t_{mon}^{v}, is defined by its start (point a) and end (point b). To exclude superfluous information, the start point should not be selected too early. The end point is restricted by t_{min}^{v} (from point b to point c), which is the minimum time required for the endangered vehicle/bicycle to avoid the possible crash. In other words, t_{min}^{v} is equivalent to the time required for the endangered driver/rider to react ($t_{driver/rider}^{e}$, from point x to point y) and the time required for the endangered vehicle/bicycle to come to a complete stop ($t_{vehicle/bicycle}^{e}$, from point y to point z).

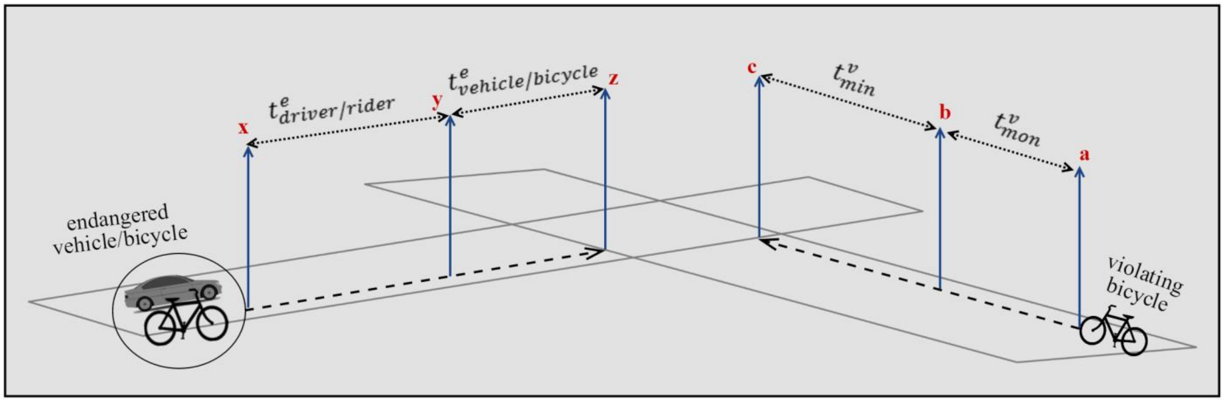


Figure 3. Variables defining monitoring period.

Based on a distribution of the human response time as presented by McLaughlin, Hankey, and Dingus [30], values from 0.5 to 2.5 seconds were chosen for the driver/rider response time ($t_{driver/rider}^{e}$), which corresponds to a response time for 5% to 95% of the population. It should be noted that the distribution of the human response time is for drivers. However, Landis et al. [31] have shown that response times for cyclists are similar to drivers. Values from 1.9 to 3.4 seconds (in the case of passenger cars) and 1.6 to 2.8 seconds (in the case of bicycles) can be considered for the vehicle/bicycle braking time ($t_{vehicle/bicycle}^{e}$), which corresponds to vehicles approaching at velocities from 25 to 45 mph and bicycles approaching at velocities from 20 to 35 mph, respectively. In the case of four-way stop signs, the endangered vehicle/bicycle is going to stop regardless of what the violating bicycle intends to do (i.e., normal crossing at stop signs). Consequently, in this case there is no need to account for the vehicle/bicycle response time and $t_{vehicle/bicycle}^{e}$ would be zero. In the case of two-way stop signs, however, the $t_{vehicle/bicycle}^{e}$ should also be added.

Variable Selection

In addition to the variables extracted during video data reduction, speed and acceleration data for the monitoring period were obtained from the MiniDAS. Subsequently, statistical measures (e.g., mean, range, max, etc.) were applied to create kinematics-related variables to characterize cyclist behavior (i.e., stopping vs. proceeding).

Modeling Results

The effect of varying the end point and the length of the monitoring period on the overall accuracy of the predictive model built using RF was studied. Two model parameters needed for model development, namely the number of decision trees and the number of variables (or factors) considered at each split, were determined to be 20 and all factors, respectively. After determining the best monitoring period, several learning algorithms were adopted to build different predictive models and overall accuracy, true positives, and false alarms were compared. The predictive models at signal-controlled intersections and stop-controlled intersections are based on the kinetic information of the cyclists. Statistical measures were applied to the kinetic data (e.g., speed and acceleration) to create factors for model development. In addition to speed and acceleration variables, time-to-intersection (TTI), distance-to-intersection (DTI), and required deceleration parameter (RDP) were used to create more factors. Table 9 presents the list of all examined factors.

Table 9. List of Factors to Develop the RF Model

Factor
$mean(TTI)$ over the t_{mon}^{v}
$range(TTI)$ over the t_{mon}^{v}
$std(TTI)$ over the t_{mon}^{v}
$max(TTI)$ over the t_{mon}^{v}
$min(TTI)$ over the t_{mon}^{v}
$mean(speed)$ over the t_{mon}^{v}
$range(speed)$ over the t_{mon}^{v}
$std(speed)$ over the t_{mon}^{v}
$max(speed)$ over the t_{mon}^{v}
$min(speed)$ over the t_{mon}^{v}
$mean(acceleraiton)$ over the t_{mon}^{v}
$range(acceleration)$ over the t_{mon}^{v}
$std(acceleration)$ over the t_{mon}^{v}
$max(acceleration)$ over the t_{mon}^{v}
$min(acceleration)$ over the t_{mon}^{v}
$mean(DTI)$ over the t_{mon}^{v}
$range(DTI)$ over the t_{mon}^{v}
$std(DTI)$ over the t_{mon}^{v}
$max(DTI)$ over the t_{mon}^{v}
$min(DTI)$ over the t_{mon}^{v}
$mean(RDP)$ over the t_{mon}^{v}
$range(RDP)$ over the t_{mon}^{v}
$std(RDP)$ over the t_{mon}^{v}
$max(RDP)$ over the t_{mon}^{v}
$min(RDP)$ over the t_{mon}^{v}

Signal-Controlled Intersections

Recall that the end point of the monitoring period is restricted by the reaction time of the endangered driver/rider and the endangered vehicle's stopping time, the sum of $t^{e}_{driver/rider}$ and $t^{e}_{vehicle/bicycle}$. Therefore, $t^{e}_{vehicle/bicycle} = \frac{posted\ speed}{2*a}$, where $a = 3\ \text{m/s}^2$. The other term defining the end point of the monitoring period is the driver/rider response ($\mathrm{t}^{\mathrm{e}}_{\mathrm{driver/rider}}$), which has a range of 0.5 to 2.5 seconds. A sensitivity analysis was conducted to understand how model performance changed with different monitoring periods. For the sake of comparison, driver/rider response was extended to range from 0.25 to 4 seconds. For each end point, four different start points were considered to examine four different monitoring period lengths of 1, 2, 3, and 4 seconds. Figure 4 illustrates the overall accuracy of RF models with different monitoring periods.

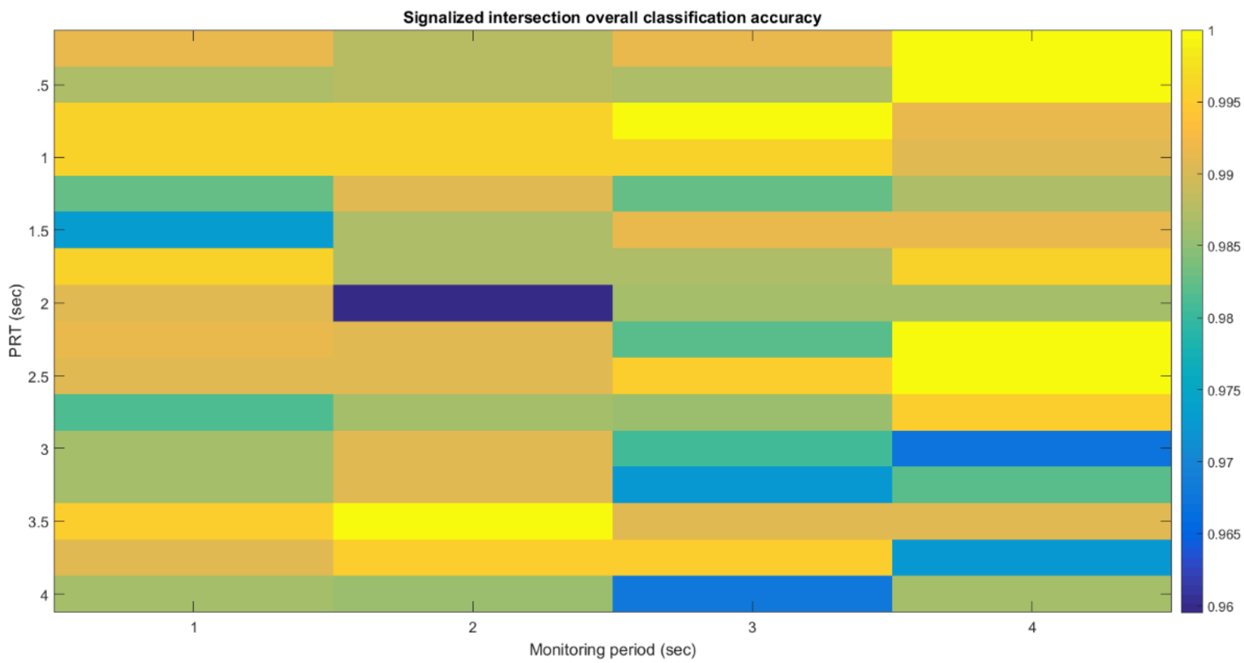


Figure 4. RF models for signalized intersections – overall accuracies.

In order to test the effect of the monitoring period end point and length on the overall accuracy, linear regression was used to model overall accuracy as a linear function of these two factors. As shown in Table 10, the accuracy decreases as $t^{e}_{driver/rider}$ increases. In other words, as the end point of the monitoring period was moved further from the stop line, the overall accuracy of the model got worse. The table also shows that the length of the monitoring period is not a significant factor.

Table 10. Linear Regression Model – Signalized Intersections

	Estimate	*p*-value
Intercept	99.3176	2.58E-100
PRT[a]($t^{e}_{driver/rider}$)	-0.183	0.043195
Monitoring length = 2 sec	9.47E-03	0.974016
Monitoring length = 3 sec	-0.226	0.430154
Monitoring length = 4 sec	-0.041	0.886623

[a] PRT is perception reaction time and it is the sum of two portions. The first is the time a driver needs to perceive sensory signal and decide his or her response. The second portion is the time needed for executing the decided response.

Stop-Controlled Intersections

As mentioned in the monitoring period section, there is no need to account for the vehicle/bicycle response time in the case of four-way stop-controlled intersections. Therefore, only the driver/rider response ($t^{e}_{driver/rider}$) with a range of 0.5 to 2.5 seconds was a factor in determining the monitoring period, which basically dictated the end point of the monitoring period. A sensitivity analysis was conducted to understand how model performance changed with different monitoring periods. For the sake of comparison, the end point of the monitoring period was extended from 2.5 to 4 seconds. For each end point, four different start points were considered to examine four different monitoring period lengths of 1, 2, 3, and 4 seconds. Figure 5 illustrates the overall accuracy of RF models with different monitoring periods. As the end point of the monitoring period was moved closer to the intersection, the classification error decreased, which shows that cyclists' behavior can be predicted with higher accuracies as they near the intersection. However, the prediction accuracies of almost 100% (e.g., when the end point is 0.25 seconds) may not be practical as the time is not sufficient for the endangered rider/driver to respond.

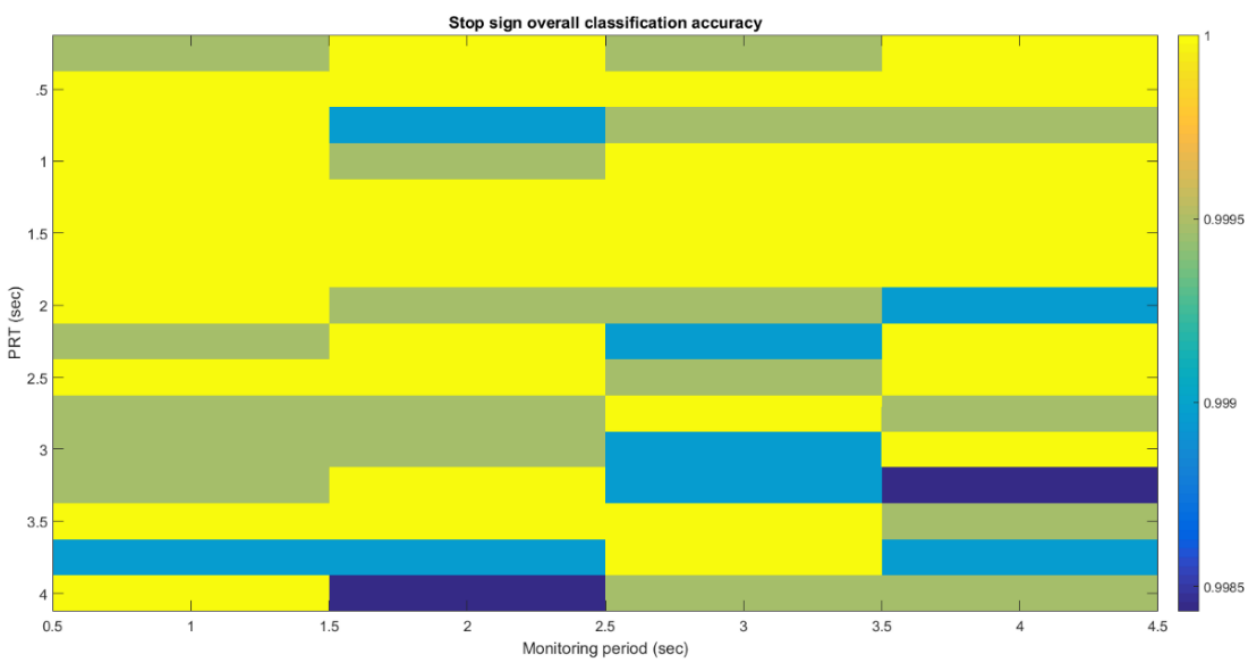


Figure 5. RF models for stop controlled intersections – overall accuracies.

The sensitivity analysis results (overall accuracies) are used to test the effect of the monitoring period end point and length on the overall accuracy. Linear regression was used to model overall accuracy as a linear function of these two factors. As shown in Table 11, the accuracy decreases as $t^{e}_{driver/rider}$ increases. In other words, as the end point of the monitoring period gets further from the stop line, the overall accuracy of the model gets worse. The table also shows that the length of the monitoring period is not a significant factor.

Table 11. Linear Regression Model – Stop Intersections

	estimate	*p*-value
Intercept	100.0055	3.58E-177
PRT($t^{e}_{driver/rider}$)	-0.014	0.002756
Monitoring length = 2 sec	-0.011	0.440016
Monitoring length = 3 sec	-0.012	0.410675
Monitoring length = 4 sec	-0.012	0.409998

Model Comparison

Four different prediction models were compared. Recall that $\mathrm{t}^{\mathrm{e}}_{\mathrm{driver/rider}}$ is one factor that was used to define the end point of the monitoring period. The results reported in this section set $\mathrm{t}^{\mathrm{e}}_{\mathrm{driver/rider}}$ equal to 4 seconds so that the monitoring period ends at least 4 seconds before the intersection. Four seconds is enough time for the intersection bicycle-car crash prediction system to process sensor data and communicate with bicycles and vehicles at the intersection. The learning algorithms vary from the simple K-NN, which does not need training, to the state-of-art RF algorithm. Logistic regression is one of the learning algorithms used in this section because it does not need any fine-tuning. The overall accuracy (ACC), true positive (TP) rates, and false positive (FP) rates for the stop-controlled and signal-controlled intersections are shown in Table 12 and Table 13.

Table 12. ACC, TP Rate, and FP Rate of Signal-Controlled Intersection Data Set

	ACC	TP	FP
RF	99.09	99.47	3.33
KNN (k = 7)	89.44	97.33	41.33
ANN	91.74	97.37	29.83
Logistic regression	91.28	95.37	33.00

Table 13. ACC, TP Rate, and FP Rate of Stop-Controlled Intersection Data Set

	ACC	TP	FP
RF	99.95	99.94	0.91
KNN (k = 7)	95.26	99.72	9.43
ANN	98.28	99.33	10.8
Logistic regression	99.53	99.83	2.43

In this context, the violation is positive. The best prediction model is the one that has the highest TP rate, which means it predicts violations with high accuracy. At the same time, this model should not distract drivers and cyclists with a high FP rate (false alarm). As shown in the table above, the worst prediction model for both data sets was the K-NN. ANN and logistic regression gave quite similar results because ANN is considered a type of logistic regression. RF was the best because it gave a high TP rate and at the same time kept the FP rate very low.

Conclusions and Recommendations

This project investigated cyclist violations at intersections using naturalistic cycling data. Different factors that affect cyclist violations at both signalized and stop-controlled intersections were examined. Mixed-effects generalized regression was carried out to identify statistically significant factors. It was found that a cyclist is more likely to violate a red light when making right turns at signalized intersections. In addition, the probability of a red light violation decreases when there is side traffic at the intersection or when there is opposing traffic to the cyclist. In the case of stop-controlled intersections, it was found that right turn, left turn, and warm weather were significant factors. Moreover, the likelihood of a stop sign violation increases when no other users are present.

Violation prediction models were developed for stop-controlled intersections and signal-controlled intersections using an RF method based on kinetic data. The kinetic data, such as speed and acceleration, were obtained through instrumented bicycles as part of a naturalistic cycling study. Different monitoring periods to extract kinetic data were considered. Different period lengths (from 1 to 4 seconds) were tested. However, the monitoring period length did not change a model's performance significantly. Another factor to define a monitoring period was the end point of the period, which is basically the time at which the computing and prediction task is started. The closer the end point was to the intersection, the higher the prediction accuracy that was achieved significantly. However, the trade-off was that higher accuracies are associated with less time for endangered users to react. The RF was compared with three other learning algorithms and found that RF was the best, having a high TP rate and at the same time keeping the FP rate very low.

Two limitations of this work include: (1) the data collection was conducted in Virginia and most observations were a small, rural town (Blacksburg). To generalize the results, additional data from larger, urban environments (e.g. cities) and collected in other states would be beneficial, as cyclist behavior may differ in dissimilar environments (e.g. big cities vs small towns), especially considering that the topography of Blacksburg includes many hilly roads which may affect cyclists' riding behavior. And (2), since violations are rare events, collecting more data would also be useful to obtain violation behavior for different types of violations. Having more observations of the same violation type would lead to developing models for that particular violation type. Therefore, future direction should focus on collecting more data and in various environments for better generalization. This research would allow models to be developed for different violation types to understand circumstances and factors that would lead to those violation types. Future research should also seek to improve current models to achieve better performance in predicting violations.

Appendix A

The Blue Cycle Study-Screening Questionnaire

Eligible: Yes No

Note to Researcher:
Initial contact between participants and researchers may take place over the phone. If this is the case, read the following Introductory Statement, followed by the questionnaire. Regardless of how contact is made, this questionnaire must be administered verbally before a decision is made regarding eligibility for this study. Once this questionnaire is completed, remove the page with their personal information and shred when no longer needed; keeping separately from the screening questions.

Introductory Statement:
After prospective participant calls or you call them, use the following script to guide you through the screening interview.

Hello. My name is _____ and I'm with the Virginia Tech Transportation Institute, here at the Smart Rd, in Blacksburg, VA. We are recruiting participants for a bicycle study that will take place in the New River Valley area of Virginia.

This project will study how people naturally ride their bicycles. You will be provided with an instrumented bicycle owned by VTTI and you will be asked to ride it instead of your own bicycle for about 4 weeks. The bicycles used for this project are equipped with cameras and sensors to observe the events and experiences that cyclists encounter in real world situations. The cameras and sensors are very small and are placed out of the way. The data collected will allow researchers to analyze many aspects of bicycle riding.

The experiment is divided into 2 phases; each phase will consist of a 2-week period. You will be paid $10 dollars per day of successful data collection plus $30 for study completion ($10 to come to VTTI and complete initial paperwork, and $20 to complete a questionnaire at the end of the experiment). Successful data collection means that you charged the system's battery the night before you intend to ride and you have the system turned on while you ride. Payments will be made by check after completion of each phase. Phase 2 will be a couple of months later after phase 1 (next year). It is possible you may not be called back for phase two, if there is insufficient data collected during the first 2 weeks of participation.

Does this sound like something you would be interested in doing?

If they indicated that they are <u>not</u> interested:

Thank you for your time.

If they indicated that they are interested:

I need to go over some screening questions to see if you meet all the eligibility requirements. Any information given to us will be kept secure and confidential. If you are not chosen to participate, we will delete this information, while we will use the information if you are selected to be in the study.

Questions

1. *Do you travel by bicycle in Blacksburg, Christiansburg, or the Radford area?* (Criterion for participation: subject must ride at least 3 days per week in one or more of these areas)
   □ Yes □ No
   If YES,
   How many days do you ride per week? _________ *(on average)*
   (Type 1-commute to work) Do you ride to work? □ No □ Yes, how many times per week?___ *(on average)*
   (Type 2-commute to school) Do you ride to school? □ No □ Yes, how many times per week?___ *(on average)*
   (Type 3-other) How many times per week do you ride to places other than work or school?___ (on average)

2. *Are you able to ride a standard bicycle with gears, without assistive devices or special equipment?*
   □ Yes □ No

3. *Are you willing to wear your helmet during the experiment*? □ Yes □ No

4. *Are you willing to use an instrumented bike instead of your own bike during the course of this experiment*? (it is a hybrid model (Trek 7.2 FX); we have three different sizes: 1. Small (15") 2. Medium (17.5") 3. Large (20") Please note that except the small bicycles, the other two sizes (medium, large) have another place for your water bottle if need be.)
   □ Yes □ No

   If yes:
   a. *Which size(s) would likely suit you?*
      □ Small (15") □ Medium (17.5") □ Large (20") □ None of these (if answer none would work, then ineligible)
   b. *Are you okay with NOT letting children ride along* <u>or any other riders</u> *on our bike while you are in possession of it*?
      □ Yes □ No (if answer no, then ineligible)

(just to remind you, you will still have your own personal bike for use on occasion if necessary)
Comments:____

5. *Are you willing to charge the battery of the instrumented bike every night*? *(you only need to take the battery off the bike and plug it in with a cable that we provide, similar to charging a cell phone)*
   □ Yes □ No

6. *Are you willing to use a bicycle lock that we provide to lock the instrumented bicycle while not in use*?
   □ Yes □ No

7. *Are you willing to keep the instrumented bicycle inside at night* and on days when you leave it at home? (Is okay if riding to a destination at night and leave outside for a short period of time, as long as it is properly locked; when possible, keep indoors at night while at a destination other than home).
   □ Yes □ No

8. *What is your current age?* ______
   (Criterion for participation: subject must be 18-30 or 45-65) *If disqualified for age, and are eligible on Q1 – Q7, ask if they would like to be called if we later change the age criterion. "Although you do not fit into our current age categories, may we call you back later if these age categories change?"*

9. *Are you available to participate at any time during the next 6 months? (note: we may not be asking participants to ride during the main part of winter – over the holidays, etc.)*
   □ Yes □ No

10. *Are you a U.S. Citizen?*
    □ Yes □ No (If No, ask next question)

    If NO, Do you have a green card?

    □ Yes □ No (If no, not eligible)

11. *Are you willing to provide your social security # should you participate, as required by the University?* (if they ask why, explain they will be asked to complete a W-9)
    - Yes
    - No (If No, then they do not qualify)

Please note that for tax recording purposes, the fiscal and accounting services office at Virginia Tech (also known as the Controller's Office) requires that all participants provide their social security number *to receive payment for participation in our studies. Or if a VT employee they may provide their VT employee #.*

12. *Indicate participant's gender :* (Criterion for participation: both are eligible)
    □ Male □ Female

13. *Do you have normal or corrected to normal hearing and vision?* (Criterion for participation: subject must have normal, or corrected to normal hearing and vision)
    □ Yes □ No

We need to ask a few questions about your medical history…

14. *Do you have a history of any of the following medical conditions? If yes, please explain.*

    a. Neck or back pain or injury to these areas
    - Yes____
    - No ___________

    b. Head injury, stroke, or illness or disease affecting the Brain
    - Yes____
    - No ___________

    c. Heart condition
    - Yes____
    - No ____________

    d. Current respiratory disorder or any condition which requires oxygen
    - Yes____
    - No _____________

    e. Epileptic seizures or lapses of consciousness within the past 12 months
    - Yes____
    - No ____________

    f. Chronic migraines or tension headaches (more than 1/month during the past year)
    - Yes____
    - No ____________

    g. Inner ear problems, dizziness, vertigo, or any balance problems (current)
    - Yes____
    - No _____________

    h. Diabetes which requires insulin?
    - Yes____
    - No __________

    i. Have you had major surgery in the past 6 months?
    - Yes____
    - No __________

j. Are you taking any substances on a regular basis which could impair your motor skills or your ability to drive?

- Yes____
- No ___________

15. (Females only) Are you currently pregnant? Yes__________ No____________ *(if "yes," politely inform the participant: while being pregnant does not disqualify you from participating in this study, you are encouraged to talk to your physician about your participation to make sure that you both feel it is safe. If you like, we can send you a copy of the consent form to discuss with your physician. Will they still be able to participate next year, if pregnant now? Answer any questions)*

Note to Researcher:
If a response to any of the questions above does not meet its criterion, read the following:

Unfortunately you are not eligible for this particular study. Thank you for your time. Would you like to be called for future studies? If yes, skip to bottom of survey.

If all the questions above met the criteria, then continue to ask the following questions:

(Now I will be asking some questions about your daily trips by bicycle, please indicate if your answers are different for your commute to work, commute to school, and other trips). Also, please answer the questions based on your two-way trips; for example, if you commute to school, and it takes you 10 minutes to go to school and 15 minutes back home, your answer to a question asking you about your trip length should be 25 minutes (10+15).
(Also, if you use other transportation modes (like your car or bus) as part of your trips, please consider only the part where you ride by bicycle)

Three different trip types are:

Type 1 trip: Commute to Work
Type 2 trip: Commute to School
Type 3 trip: Other Trips by Bicycle

16. *On average, how long does your trip take in minutes?(sum of all trips above 5 minutes could be eligible) (Remind that they should account for the **two-way** trips)*
Type 1**:** ____ *minutes*
Type 2**:** ____ *minutes*
Type 3**:** ____ *minutes*

17. *Approximately what percentage of your trip is on paved roads, bicycle trails, and sidewalks? (NOT eligible if more than 90% of all trips (i.e. CTW, CTH, and OTB) are bicycle trails or on sidewalks, other than that could be eligible) (Remind that they should account for the **two-way** trip)*

Type 1: *(a. ___% on paved roads b. ___% on bicycle trails c. ___% on sidewalks)*
Type 2: *(a. ___% on paved roads b. ___% on bicycle trails c. ___% on sidewalks)*
Type 3: *(a. ___% on paved roads b. ___% on bicycle trails c. ___% on sidewalks)*

18. *What are the main streets/roads you are riding on?(any answer could be eligible)*
Type 1: ________________________
Type 2: ________________________
Type 3: ________________________

19. *Do you encounter intersections in your trips? (must encounter intersections)*
□ Yes □ No (If YES, ask next question)
*If yes, how many of them are signalized and how many are stop signs? (eligible if encounter 3 or more signalized intersections <u>AND</u> 3 or more stop signs - 6 or more in total)(Remind that they should count for the **two-way** trip)*
Type 1: *signalized:*_______________*stop signs:*__________
Type 2: *signalized:*_______________*stop signs:*__________
Type 3: *signalized:*_______________*stop signs:*__________

20. *Is there a place where you experience a lot of delays throughout your trips? Where (e.g. a specific street or intersection)?(any answer could be eligible)*
Type 1: □ No □ Yes, Location: ___________________
Type 2: □ No □ Yes, Location: ___________________
Type 3: □ No □ Yes, Location: ___________________

21. *What time of day do you ride?(any answer could be eligible)*
Type 1: ___________________
Type 2: ___________________
Type 3: ___________________

22. *Are there conditions in which you may decide not to ride and use other modes of transport (conditions like: being sick, rainy day, snowy day, possibility of rain, specific months like in December, etc.)?(any answer could be eligible)*
Type 1: □ No □ Yes, Conditions: ______________________________
Type 2: □ No □ Yes, Conditions: ______________________________
Type 3: □ No □ Yes, Conditions: ______________________________

Comments in response to question 22:____

23. *Do you typically carry children on your trips using your bicycle?*
□ Yes □ No (If Yes, ask next question)

If Yes, How often do you carry children on your trips?

Type 1: ____________
Type 2: ____________
Type 3: ____________

If Yes. You cannot use this bicycle to transport children while in our study, will that be a problem? Ineligible if this is a problem (reminding them, they have access to their own bicycle if necessary during participation)

Comments:

____

____

Note to Researcher:
At this point the questions are finished, please read the following:

Thank you for your time. Depending on the number of people who respond, it is possible not all eligible people will be selected to participate in this study. Appointments will be filled until sufficient numbers have been enrolled.

Criteria For Participation

1. Must commute/travel by bicycle at least 3 times per week on average in the Blacksburg, Christiansburg, and/or Radford areas
2. Must be able to ride a bike with gears without assistive devices or any special equipment
3. Must wear their own helmet while participating/riding our bike
4. Must be willing to use an instrumented bicycle that VTTI provides during the experiment <u>*AND*</u>
   A. Agree they will NOT allow children to ride along or any other riders on the VTTI bicycle
5. Must take the battery off the bike every night and charge it for the next day/ride
6. Must use the bicycle lock that we provide when not using the bicycle
7. Must be able to keep the instrumented bicycle inside at night or while not 'in use' (out riding it)
8. Subject must be 18-30 or 45-65 years of age
9. Must be available to participate during the next 6 months (not moving or scheduled to have major surgery, or anything that may interfere with them participating in part 2 next year.
10. Must be a U.S. citizen or hold a green card
11. Must be willing to complete a W-9 tax form which includes their *SSN or VT ID #*
12. Both males and females are eligible (we need the information later)
13. Must have normal (or corrected to normal) hearing and vision (self-report).

14. Health Questions*:*

   a. *Cannot have a history of neck or back conditions which still limit their ability to participate in certain activities.*
   b. *Cannot have a history of brain damage from stroke, tumor, head injury, recent concussion, or disease or infection of the brain.*
   c. *Cannot have a current heart condition which limits their ability to participate in certain activities.*
   d. *Cannot have current respiratory disorders or disorders requiring oxygen.*
   e. *Cannot have had epileptic seizures or lapses of consciousness within the last 12 months.*
   f. *Cannot have chronic migraines or tension headaches (averages no more than one per month).*
   g. *Cannot have current problems with motion sickness, inner ear problems, dizziness, vertigo, or balance problems.*
   h. *Cannot have diabetes which requires insulin*
   i. *Must not have had any major surgery within the past 6 months.*
   j. *Cannot currently be taking any substances that may interfere with driving ability (cause drowsiness or impair motor abilities).*

15. If pregnant, encourage them to speak with their doctor first. Remind of timeline, that they will be called back next year for phase 2 and see if this will be a problem (would they still be able to participate a few months from now)

If not Eligible: Would you like to be contacted for future studies? Yes: ______ No: _______

Full Name: ____________________________ Year Of Birth: ___________________

Contact Numbers:____

Email:__________________________(

Specialty License(s)?: ____

If Eligible:

Name: ___________________________ Phone: ________________________________

E-mail or mailing address: ____

Availability: ____

Scheduled on (date & time):____

Would you like to be contacted for future studies? Yes: ______No: ______YOB:_________

Primary Vehicle: ___________________________ Specialty License: ____________________

References

1. McLeod, K., D. Flusche, and A. Clarke. Where We Ride: Analysis of Bicycling in American Cities. League of American Bicyclists, 2013.

2. National Highway Traffic Safety Administration. *Bicyclists and Other Cyclists: 2013 Data*. Traffic Safety Facts Report No. DOT HS 812 151. National Center for Statistics and Analysis, US Department of Transportation, 2015.

3. National Highway Traffic Safety Administration. *Fatality Analysis Reporting System*. http://www-fars.nhtsa.dot.gov/People/PeoplePedalcyclists.aspx. Accessed 2015.

4. The University of North Carolina Highway Safety Research Center. *North Carolina Bicycle Crash Types 2005-2009*. 2011. http://www.ncdot.gov/bikeped/download/summary_bike_types05-09.pdf.

5. Tan, C. Crash-Type Manual for Bicyclists. Publication No. FHWA-RD-96-104. Federal Highway Administration, 1996.

6. Wu, C., L. Yao, and K. Zhang. The Red-Light Running Behavior of Electric Bike Riders and Cyclists at Urban Intersections in China: An Observational Study. *Accident Analysis & Prevention*, Vol. 49, 2012, pp. 186-192.

7. Johnson, M., J. Charlton, J. Oxley, and S. Newstead. Why Do Cyclists Infringe at Red Lights? An Investigation of Australian Cyclists' Reasons for Red Light Infringement. *Accident Analysis & Prevention*, Vol. 50, 2013, pp. 840-847.

8. Pai, C.-W. and R.-C. Jou. Cyclists' Red-Light Running Behaviours: An Examination of Risk-Taking, Opportunistic, and Law-Obeying Behaviours. *Accident Analysis & Prevention*, Vol. 62, 2014, pp. 191-198.

9. Martínez-Ruiz, V., P. Lardelli-Claret, E. Jiménez-Mejías, C. Amezcua-Prieto, J. J. Jimenez-Moleon, and Juan de Dios Luna del Castillo. Risk Factors for Causing Road Crashes Involving Cyclists: An Application of a Quasi-Induced Exposure Method. *Accident Analysis & Prevention*, Vol. 51, 2013, pp. 228-237.

10. Johnson, M., J. Charlton, and J. Oxley. Cyclists and Red Lights—A Study of the Behaviour of Commuter Cyclist in Melbourne. *Australasian Road Safety Research, Policing and Education Conference*, Adelaide, 2008.

11. Johnson, M., S. Newstead, J. Charlton, and J. Oxley. Riding through Red Lights: The Rate, Characteristics and Risk Factors of Non-compliant Urban Commuter Cyclists. *Accident Analysis & Prevention*, Vol. 43, No. 1, 2011, pp. 323-328.

12. Schepers, J. P., P. A. Kroeze, W. Sweers, and J. C. Wüst. Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections. *Accident Analysis & Prevention*, Vol. 43, No. 3, 2011, pp. 853-861.

13. Jahangiri, A., H. A. Rakha, and T. A. Dingus. Developing a System Architecture for Cyclist Violation Prediction Models Incorporating Naturalistic Cycling Data. *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE*, Las Vagas, NV, 2015.

14. Phillips, R. O., T. Bjørnskau, R. Hagman, and F. Sagberg. Reduction in Car–Bicycle Conflict at a Road–Cycle Path Intersection: Evidence of Road User Adaptation? *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 14, No. 2, 2011, pp. 87-95.

15. Zhang, Y. and C. Wu. The Effects of Sunshields on Red Light Running Behavior of Cyclists and Electric Bike Riders. *Accident Analysis & Prevention*, Vol. 52, 2013, pp. 210-218.

16. Räsänen, M., I. Koivisto, and H. Summala. Car Driver and Bicyclist Behavior at Bicycle Crossings under Different Priority Regulations. *Journal of Safety Research*, Vol. 30, No. 1, 1999, pp. 67-77.

17. Dingus, T.A., S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. A. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher. Z. R. Doerzaph, J. Jermeland, and R. R. Knipling. The 100-Car Naturalistic Driving Study, Phase II–Results of the 100-Car Field Experiment. Publication DOT HS 810 593. NHTSA, U.S. Department of Transportation, 2006.

18. Sayer, J., D. LeBlanc, S. Bogard, D. Funkhouser, S. Bao, M. L. Buonarosa, and A. Blankespoor. Integrated Vehicle-Based Safety Systems Field Operational Test Final Program Report, 2011.

19. Benmimoun, M., M. L. Aust, F. Faber, G. Saint Pierre, and A. Zlocki. Safety Analysis Method for Assessing the Impacts of Advanced Driver Assistance Systems within the European Large Scale Field Test euroFOT. *8th ITS European Congress*, Lyon, France, 2011.

20. Dozza, M., J. Werneke, and A. Fernandez. Piloting the Naturalistic Methodology on Bicycles. *Proceedings of the International Cycling Safety Conference 2012,* Helmond NL, Nov 7-8, 2012.

21. Dozza, M., and A. Fernandez. Understanding Bicycle Dynamics and Cyclist Behavior From Naturalistic Field Data (November 2012). *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 1, 2014, pp. 376-384.

22. Johnson, M., J. Charlton, J. Oxley, and S. Newstead. Naturalistic Cycling Study: Identifying Risk Factors for On-Road Commuter Cyclists. *Annals of Advances in Automotive Medicine/Annual Scientific Conference,* Association for the Advancement of Automotive Medicine, 2010.

23. Gustafsson, L., and J. Archer. A Naturalistic Study of Commuter Cyclists in the Greater Stockholm Area. *Accident Analysis & Prevention*, Vol. 58, 2013, pp. 286-298.

24. Lacherez, P., J. M. Wood, R. P. Marszalek, and M. J. King. Visibility-Related Characteristics of Crashes Involving Bicyclists and Motor Vehicles–Responses from an Online Questionnaire Study. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 20, 2013, pp. 52-58.

25. Breiman, L. Random Forests. *Machine Learning*, Vol. 45, No. 1, 2001, pp. 5-32.

26. Hastie, T., R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics, Springer, 2009.

27. Friedman, J. H., F. Baskett, and L.J. Shustek. An Algorithm for Finding Nearest Neighbors. *IEEE Transactions on Computers*, Vol. C-24 No. 10., 1975, pp. 1000-1006.

28. Rumelhart, D. E., G. E. Hinton, and R. J. Williams. Learning Representations by Back-Propagating Errors. *Cognitive Modeling,* Vol. 5, No. 3, 1988, p. 1.

29. Jahangiri, A., H. Rakha, and T. A. Dingus. Predicting Red-light Running Violations at Signalized Intersections Using Machine Learning Techniques. Presented at 94th Annual Meeting of the Transportation Research Board, Washington, D.C., 2015.

30. McLaughlin, S. B., J. M. Hankey, and T. A. Dingus. A Method for Evaluating Collision Avoidance Systems using Naturalistic Driving Data. *Accident Analysis & Prevention*, Vol. 40, No. 1, 2008, pp. 8-16.

31. Landis, B., T. Petritsch, H. Huang, and A. Do. Characteristics of Emerging Road and Trail Users and their Safety. In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1878*, 2004, pp. 131-139.