Report 15-UT-030

NSTSCE

National Surface Transportation Safety Center for Excellence

Using Functional Classification to Enhance Naturalistic Driving Data Crash/Near Crash Algorithms

Final Report

Jeremy Sudweeks

Submitted: January 20, 2015

Technology

Impairment

Mobility

Infrastructure

Housed at the Virginia Tech Transportation Institute
3500 Transportation Research Plaza • Blacksburg, Virginia 24061

ACKNOWLEDGMENTS

The authors of this report would like to acknowledge the support of the stakeholders of the National Surface Transportation Safety Center for Excellence (NSTSCE): Tom Dingus from the Virginia Tech Transportation Institute, John Capp from General Motors Corporation, Lincoln Cobb from the Federal Highway Administration, Chris Hayes from Travelers Insurance, Martin Walker from the Federal Motor Carrier Safety Administration, and Cathy McGhee from the Virginia Department of Transportation and the Virginia Center for Transportation Innovation and Research.

The NSTSCE stakeholders have jointly funded this research for the purpose of developing and disseminating advanced transportation safety techniques and innovations.

The Government and others acting on its behalf shall have unlimited rights to obtain, reproduce, publish or otherwise use the data developed in the performance of this cooperative agreement pursuant to 49 CFR Part 19.36.

EXECUTIVE SUMMARY

A persistent challenge in using naturalistic driving data is identifying events of interest from a large data set in a cost-effective manner. A common approach to this problem is to develop kinematic thresholds against which kinematic data is compared to identify potential events or kinematic triggers. Trained video analysts are then used to determine if any of the kinematic triggers have successfully identified events of interest.

Video validation for a large number of kinematic triggers is time consuming, expensive, and possibly prone to error. The use of video analysis to review a large number of kinematic triggers is due to an inability to effectively discriminate between innocuous driving situations and safety-relevant events in an automated manner. A potential solution to this problem is the development of classification models that would reduce the number of kinematic triggers submitted for video validation through a process of pre-validation trigger classification.

A functional yaw rate classifier was developed that retains a majority of safety relevant events (92% of crashes, 81% of near-crashes) while reducing the number of invalid or erroneous yaw rate triggers by 42%. For large-scale studies such a reduction in the number of invalid triggers submitted for video validation allows video analysis resources to be focused on the aspect of driving research in which it add the greatest value: providing contextual information that cannot be derived from kinematic and parametric data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

DAS	Data Acquisition System
TTC	Time-to-Collision
FDA	Functional Data Analysis
LDA	Linear Discriminant Analysis
PDA	Penalized Discriminant Analysis
CCA	Canonical Correlation Analysis
IQR	Interquartile Range

CHAPTER 1. INTRODUCTION

A persistent challenge in using naturalistic driving data is identifying events of interest from a large data set in a cost-effective manner. A common approach to identifying these events is to define thresholds for kinematic measures collected by the data acquisition system (DAS) that are believed to indicate the occurrence of an event. Continuously collected naturalistic data are then programmatically compared against these thresholds, either individually or in meaningful combinations, and any data points that exceed the thresholds are marked for video validation. The collection of data points that exceed the threshold values within a pre-defined temporal window are referred to as kinematic triggers, or simply triggers. Trained video analysts determine if a given kinematic trigger is valid with the aid of pre-defined operational definitions of events. The process of defining threshold parameters, conducting programmatic comparisons, and validating kinematic triggers with video review is iterative, as threshold parameters often need to be tuned. This iterative approach can be seen in Figure 1.

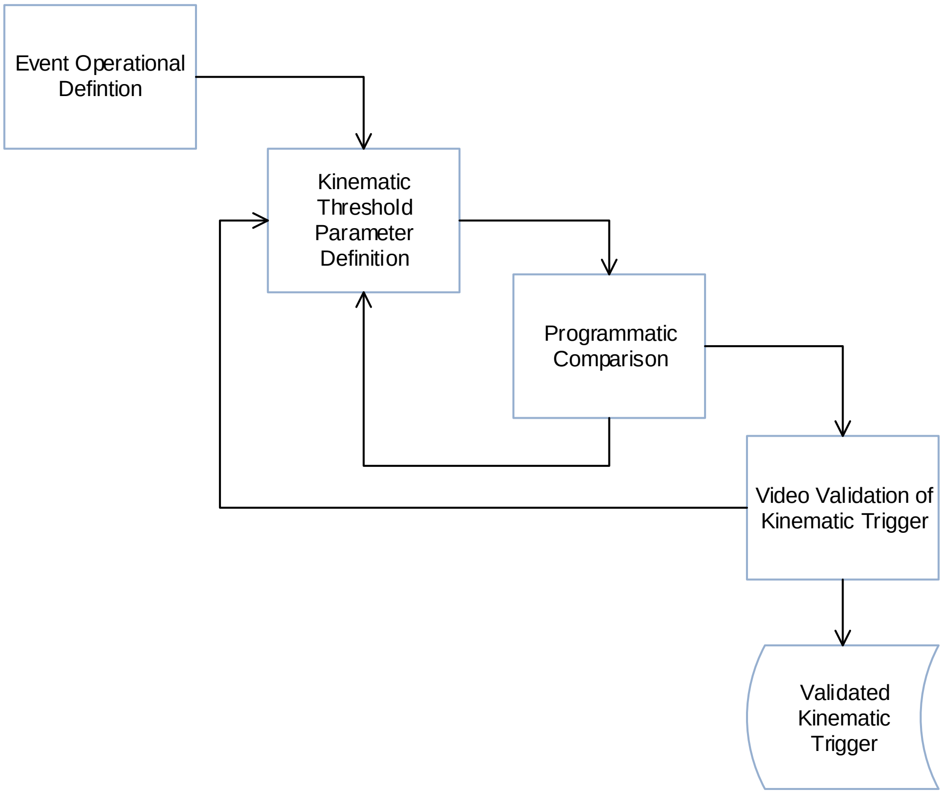
Event Operational Defintion
Kinematic Threshold Parameter Definition
Programmatic Comparison
Video Validation of Kinematic Trigger
Validated Kinematic Trigger

Figure 1. Diagram. Iterative kinematic threshold triggering.

As an example, consider the search for safety-relevant driving events in the 100-Car data set. Approximately 20 kinematic thresholds were defined to identify safety-relevant driving events in the 100-Car data set. From over 120,000 total kinematic triggers, approximately 11,500 validated triggers were associated with 9,123 safety-relevant events. These events were classified according to severity as crashes, near-crashes, or critical incidents.(1) The event severity distribution for 100-Car events is shown in Figure 2.

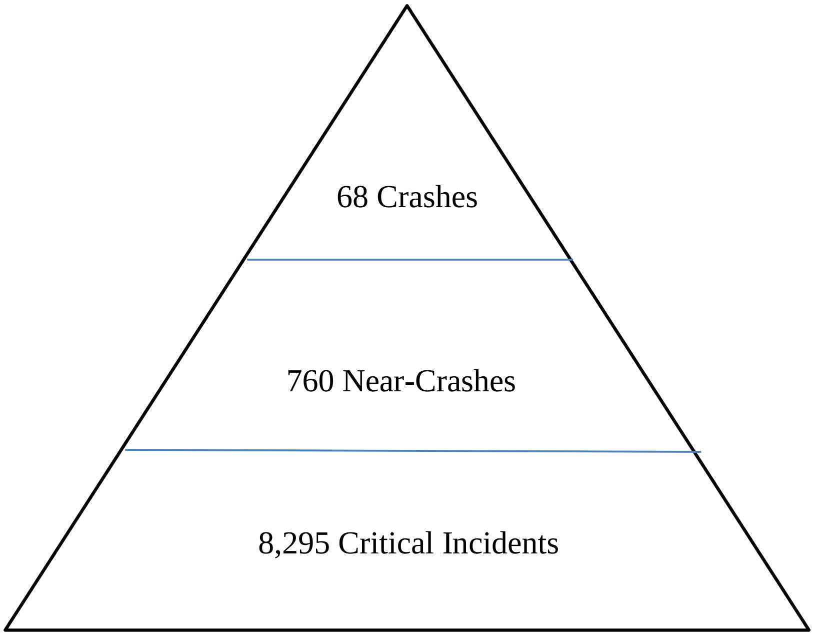

Figure 2. Chart. Event severity distribution for 100-Car events.

While the safety-relevant events discovered in the 100-Car data set were very useful in furthering the understanding of driving behavior the process to identify them was time consuming and expensive.

VALIDATION RATES FOR 100-CAR KINEMATIC TRIGGERS

Although approximately 20 kinematic thresholds were defined to identify safety-relevant events in the 100-Car data set, the efficiency with which valid events were identified varied significantly across thresholds. Threshold performance is examined in Table 1 through Table 3.

Table 1 shows performance, as a percentage of invalid and valid trigger status, for a subset of the 100-Car kinematic thresholds. As mentioned, the determination of the validity of a trigger was made by a trained video analyst with the aid of pre-defined operational definitions of events.

Table 1. Performance of kinematic triggers as percentage of validity class.

Kinematic Threshold	Trigger Frequency (Invalid || Valid)	Percentage of Invalid Triggers	Percentage of Valid Triggers
Forward Time-to-Collision (TTC)	25,536 || 5,371	19.30	46.63
Lane Abort	1,206 || 2	0.90	0.02
Lane Solid	1,209 || 8	0.90	0.07
Lateral Acceleration	3,269 || 88	2.47	0.76
Longitudinal Acceleration	7,937 || 3,675	6.00	31.91
Rear TTC	866 || 440	0.66	3.81
Side Blind Spot	1,507 || 4	1.14	0.03
Side Blinker	3,845 || 3	2.91	0.03
Side Cutoff	853 || 261	0.64	2.27
Side Yaw	1,396 || 15	1.06	0.13
Yaw Rate	84,648 || 1,651	64.00	14.33

As can be seen in Table 1, the forward time-to-collision, longitudinal acceleration, and yaw rate triggers are associated with the largest number of total kinematic triggers. The longitudinal acceleration trigger performed well by identifying approximately 32% of valid triggers and only 6% of invalid triggers, while the forward TTC trigger did not do as well, as it accounted for approximately 47% of valid triggers, while also being associated with 19% of invalid triggers. The relative performance of the yaw rate trigger was the worst of the three as it was associated with around 14% of valid triggers, but accounted for 64% of invalid triggers.

Another way of evaluating the relative cost of the kinematic triggers, in terms of video reduction effort, is to consider the number of invalid triggers that must be evaluated for each valid trigger evaluated. This is portrayed in Table 2, which shows the number of invalid triggers per valid trigger for the set of kinematic triggers shown in Table 1.

Table 2. Performance of kinematic triggers as number of valid triggers per invalid trigger.

Threshold Kinematic Trigger	Number of Invalid Triggers per Valid Trigger
Forward TTC	5
Lane Abort	603
Lane Solid	151
Lateral Acceleration	37
Longitudinal Acceleration	2
Rear TTC	2
Side Blind Spot	377
Side Blinker	1,282
Side Cutoff	3
Side Yaw	93
Yaw Rate	51

The relative performance, in terms of the number of invalid triggers per valid trigger, for the forward TTC, longitudinal acceleration and yaw rate kinematic triggers is pronounced. As shown in Table 2, the longitudinal acceleration trigger has an invalid-to-valid ratio of 2, while the forward TTC trigger ratio is 5, and the yaw rate trigger ratio is 51. For the yaw rate trigger in particular, video analysts were required to view a significant number of invalid triggers in order to find the valid triggers that identified events of interest.

It is informative to consider the performance of the kinematic thresholds in uniquely identify safety relevant events. As can be seen in Figure 2 and Table 1, the total number of kinematic triggers exceeds the total number of safety relevant events. This is due to the fact that for some events multiple thresholds, such as longitudinal acceleration and forward TTC, were activated. Table 3 shows the percentage of events that were associated by thresholds individually or in combination.

Table 3. Performance of kinematic triggers as percentage of safety events.

Kinematic Trigger	Percentage of Events (%)
Forward TTC	33.2
Longitudinal Acceleration, Forward TTC	22.7
Yaw Rate	17.1
Longitudinal Acceleration	15.8
Side Cutoff	3.0
Forward TTC, Rearward TTC	1.5
Longitudinal Acceleration, Forward TTC, Rearward TTC	1.5
Rearward TTC	1.2
Other	3.9
TOTAL	100

As can be seen in Table 3, approximately 33% of valid events were associated with only the forward TTC trigger, around 16% were associated with only the longitudinal acceleration trigger, while 17% were associated with only the yaw rate trigger. Although the yaw rate threshold trigger had a high number of invalid triggers versus valid triggers, as shown in Table 2, it was important to use as it uniquely identified a significant number of safety-relevant events.

PROBLEM STATEMENT AND OBJECTIVE

Video validation for a large number of kinematic triggers is time consuming, expensive, and possibly prone to error. In addition to the obvious time and expense associated with culling invalid triggers, there is also a potential, although unquantified, cost associated with the degradation of validation quality. As video analysts review large numbers of proposed kinematic triggers, task vigilance is likely to decrease due to the rare nature of safety-relevant events. Video analysis efforts need to be focused on the aspect of driving research where they add the greatest value: providing contextual information that cannot be derived from kinematic and parametric data.

The use of video analysis to review a large number of kinematic triggers is due to an inability to effectively discriminate between innocuous driving situations and safety-relevant events in an automated manner. The use of well-developed classification models may reduce the time and expense associated with validating kinematic triggers via video review. Others have recognized the value of using automated classification models to reduce video annotation time.(2) The objective of this analysis is to evaluate a functional classification method to distinguish between valid and invalid kinematic triggers identified during the 100-Car study. If a functional classification method proves effective at distinguishing between valid and invalid triggers, this method can be applied in future naturalist driving studies so that fewer resources will need to be devoted to video evaluation of kinematic triggers. This, in turn, will enabled video analysts to focus on activities that provide the greatest value to understanding driver behavior.

YAW RATE KINEMATIC THRESHOLD

The kinematic threshold of interest in this analysis is the yaw rate threshold. This threshold was developed to identify situations in which a driver performed a sudden steering maneuver. As was shown in Table 1 through Table 3, this kinematic threshold is a good candidate for use with a classification model as yaw rate triggers identified a large number of valid events but also had many invalid triggers or false alarms.

The yaw rate threshold was defined as follows:[(1)]

- yaw rate oscillation was in excess of 4 degrees per second within a 3-second window;
- the vehicle returned to its direction of travel prior to the steering maneuver;
- minimum speed at trigger onset was 15 miles per hour

Figure 3 shows yaw rate plots for 5 yaw rate triggers. The x-axis is shown in 1/10 second (100 ms) increments, which is the rate at which yaw rate data were recorded on the 100-Car DAS. The y-axis is the yaw rate, reported in degrees per second. The black dashed lines on the plot represent the trigger parameters of −4 degrees per second, 0 degrees per second, and 4 degrees per second.

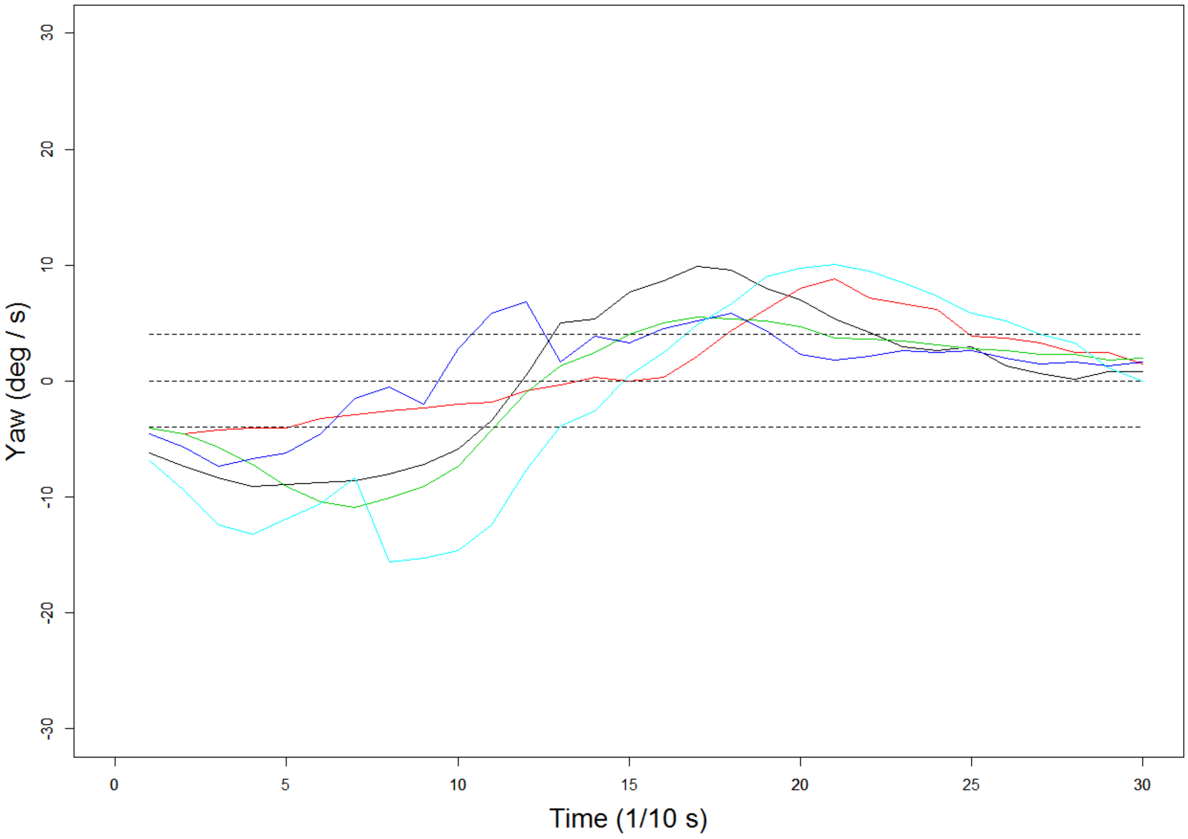

Figure 3. Plot. Yaw rate data for 5 yaw rate kinematic triggers.

As expected from the yaw rate trigger definition, the yaw rate plots in Figure 3 show a roughly sinusoidal shape. As the number of yaw rate triggers plotted increases characteristics of individual curves are difficult to distinguish, but general trends become apparent. Figure 4 shows yaw rate plots for approximately 11,000 yaw rate triggers.

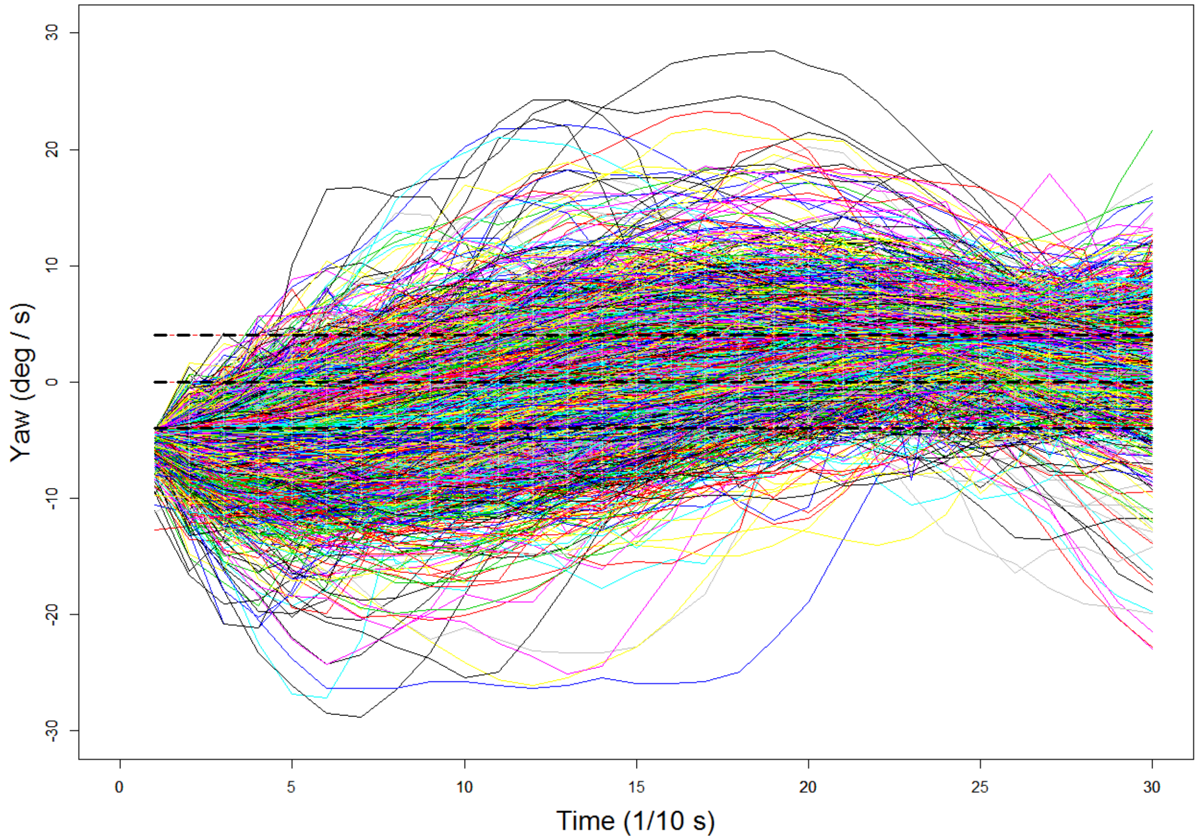

Figure 4. Plot. Yaw rate data for approximately 11,000 yaw rate kinematic triggers.

As with Figure 3, Figure 4 shows the roughly sinusoidal shape of the yaw rate triggers. There also appears to be significant variation along both the yaw scale (amplitude variation) and the time scale (phase variation). The amplitude variation is seen in the differing size or heights of the curves. The phase variation can be seen in the misalignment of curves. Although each curve on the plot represents roughly the same driving maneuver, the features of the curves, such as peaks and valleys, are not aligned. Curve alignment will be discussed more in a subsequent section of this report.

DATA PREPARATION AND CLEANING

There were several driving scenarios in which a yaw rate trigger was recorded in the 100-Car data set. As it seems likely that the causal mechanisms for evasive maneuvers differ under various scenarios, it is possible that scenario identification prior to classification would be beneficial. That is, sources of variation in a relatively homogenous set of triggers for a given driving scenario are more likely to be closely associated with trigger validity than sources of variation from a set of triggers from several driving scenarios. In fact, such a refinement of scenario identification showed some classification improvement in previous efforts.[1] For this investigation, the following criteria were used to define a scenario of interest:

- the yaw rate trigger was activated in isolation from other kinematic triggers;
- the yaw rate trigger was activated for 2 to 3 seconds;
- only yaw rate triggers from light vehicles were considered.

Radar-based measures would be helpful to further develop scenarios, particularly in distinguishing between situations of conflict with a lead vehicle and single vehicle conflicts. However, radar data processing and target classification were considered beyond the scope of this effort.

As can be seen in Figure 4, the initial direction of movement has been disregarded. That is, rather than partitioning the data set into left movement and right movement components, the sign of the yaw rate data was reversed, if necessary, in order to have directionally similar initial movement along the yaw axis. This decision prevents splitting the relatively rare valid yaw rate triggers and was based on the assumption that the direction of initial movement is not as important to classification as the shape of the yaw rate curve.

Based on exploratory analysis and plotting of the approximately 85,000 yaw rate triggers, two primary data cleaning tasks were deemed necessary. The first was to address errors in the yaw rate data stream. This consisted of correcting an apparent non-zero default constant value along with the identification and mitigation of yaw rate outliers.[3] The few yaw rate triggers for which this method was unable to appropriately replace outliers were removed from further consideration in classification.

The second data cleaning task was to determine if particular vehicles had a significantly higher rate of yaw rate triggers than other vehicles. This was accomplished by plotting the number of yaw rate triggers for each vehicle, as shown in Figure 5. It should be noted that a single vehicle accounted for approximately 65,000 yaw rate triggers and was removed from consideration before producing Figure 5.

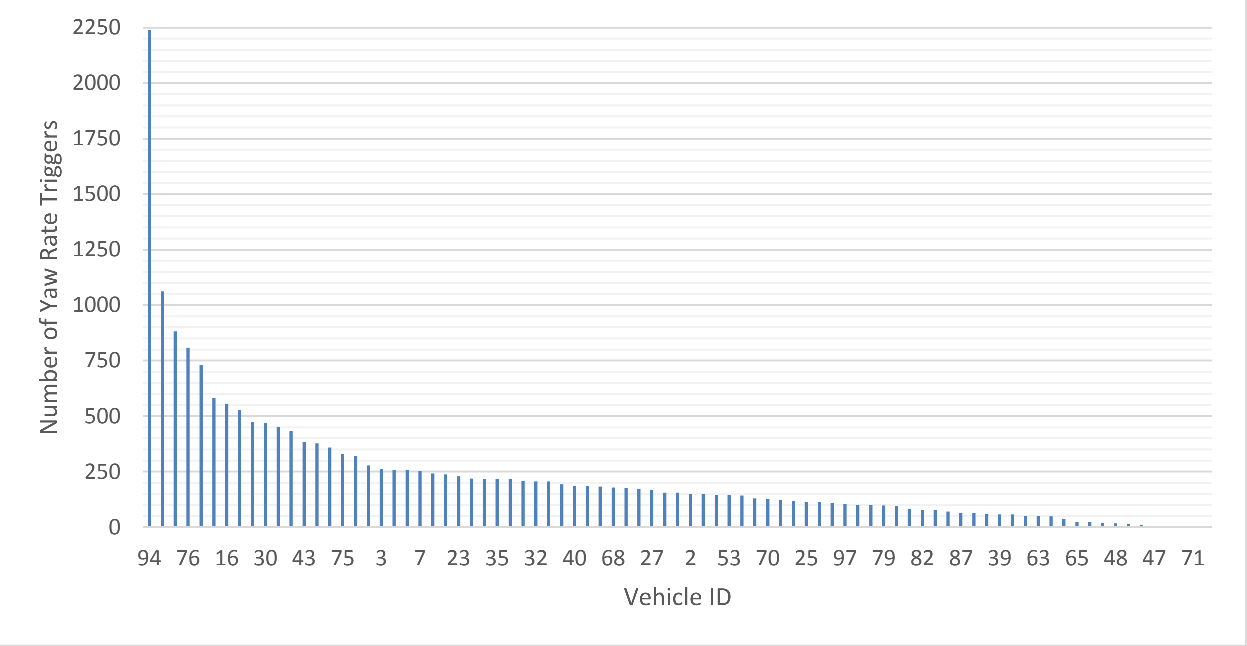

Figure 5. Plot. Number of yaw rate triggers by vehicle.

Based on a review of this plot (Figure 5), it was decided to only consider light vehicles with less than 800 yaw rate triggers for further analysis. Because the purpose of the yaw rate trigger was to identify situations when a participant performed a sudden steering maneuver, establishing an upper threshold on the number of yaw rate triggers seems reasonable, as an excessive number of yaw rate triggers is likely an indication of hardware or data abnormalities rather than risky driving.

As a result of the scenario definition and data cleaning tasks, approximately 11,000 yaw rate triggers were available for classification. A decomposition of the triggers by validity and severity is shown in Table 4.

Table 4. Yaw rate trigger frequency by validity and severity.

Validity	Severity	Frequency
Valid	Crash	5
Valid	Near Crash	51
Valid	Critical Incident	564
Invalid	NA	10,330
TOTAL		**10,950**

As was alluded to in the review of Table 1 through Table 3, valid events are quite rare relative to invalid yaw rate triggers. This is particularly true as event severity increases. The rare nature of valid events will be considered further in the discussion of classification methods.

FUNCTIONAL DATA REPRESENTATION AND SMOOTHING

As the yaw rate data collected by the DAS is a discretization of a continuous signal, functional data analysis (FDA) techniques were used to represent the data for analysis.(4-7) In this approach, the data points collected were used to create an object that estimates the smooth underlying function from which the data points were observed. This approach made use of a B-spline basis system and roughness penalty smoothing and was estimated with the R package 'fda.' The results of functional representation and smoothing are shown in Figure 6. Note that these are the same curves that were plotted as raw data in Figure 4.

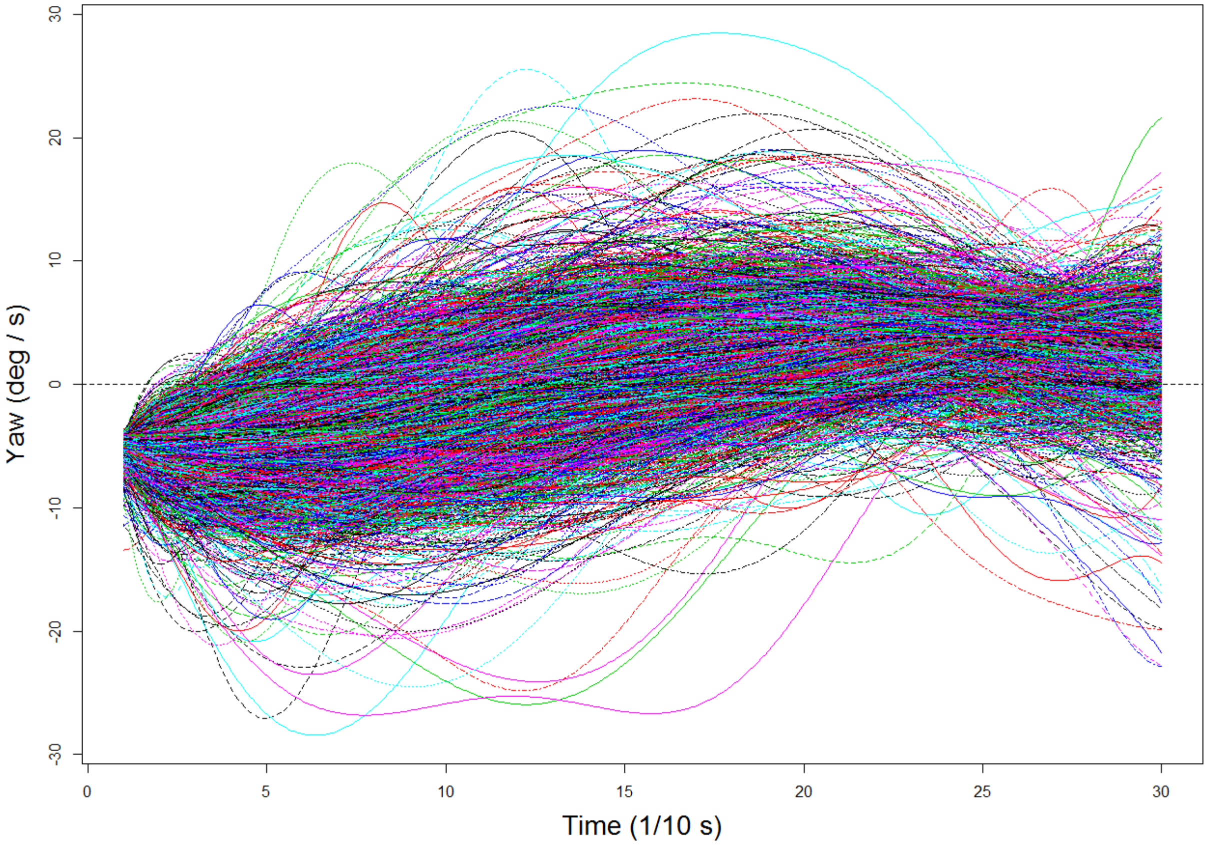

Figure 6. Plot. Smoothed yaw rate data for approximately 11,000 yaw rate triggers.

The roughness penalty approach to smoothing seeks a trade-off between data fit and curve smoothness. As can be seen in a comparison of Figure 4 and Figure 6, the degree of smoothing was relatively light as the raw data were not particularly noisy. After smoothing there is still significant variation along both the time scale (phase variation) and along the yaw scale (amplitude variation). The time scale, or phase variation, is problematic for additional analysis as most statistical methods are designed to deal solely with amplitude variation.(8)

CURVE REGISTRATION

In order to mitigate or eliminate the effects of phase variation, a time transformation is sought that will bring all the curves under consideration onto a common time scale. This process of aligning curves to a common time scale is referred to as registration.(5,8) Two phases of curve registration were used: landmark registration, followed by continuous registration.

Landmark Registration

Landmark registration estimates a common time scale by using fixed end points and any number of known landmark features identifiable on each curve.(5,8) The combination of known points is used with linear interpolation to estimate a common time scale and bring the curves into rough alignment. For the yaw rate curves, the first crossing of zero yaw rate was used as the landmark feature for registration. If a zero yaw rate crossing value could not be determined for a trigger, it was removed from further analysis. The result of this registration can be seen in Figure 7. The vertical dashed line on the plot is the average time for the first zero crossing for all of the curves.

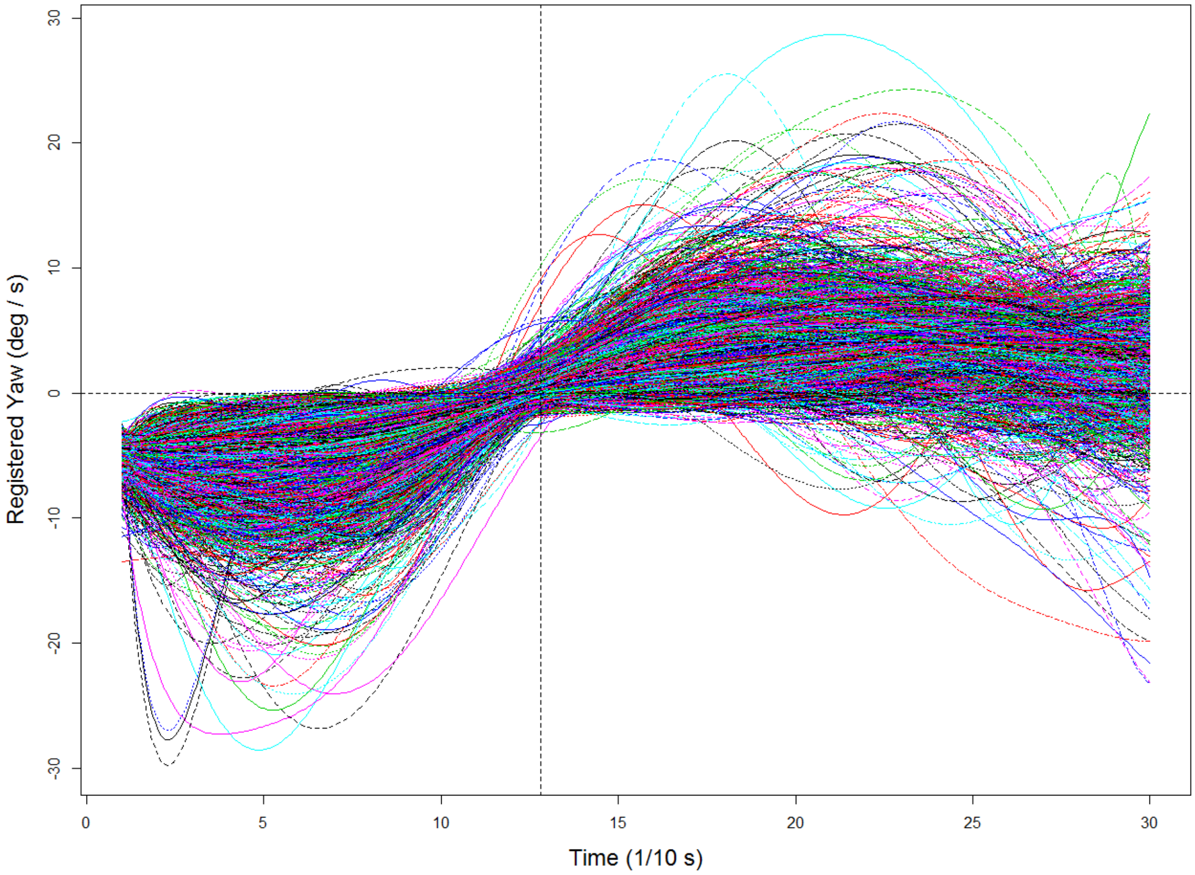

Figure 7. Plot. Landmark registered data for approximately 11,000 yaw rate triggers.

The curve alignment achieved through landmark registration is an improvement over the unaligned curves, as can be seen by comparing Figure 6 and Figure 7. In a few cases, the landmark registration process produced registered yaw rate curves that deviated significantly from the unregistered curve. These yaw rate triggers were excluded from further analysis.

Continuous Registration

In order to further align the curves and reduce the effect of phase variation, a continuous registration method was used, as implemented in the 'fda' R package. Continuous registration methods seek to use information from the entire curve for alignment rather than just a few specified values as in the landmark approach.(5,8) The results of the landmark registration were used as inputs to the continuous registration.(5) The results of the second phase of registration are shown in Figure 8.

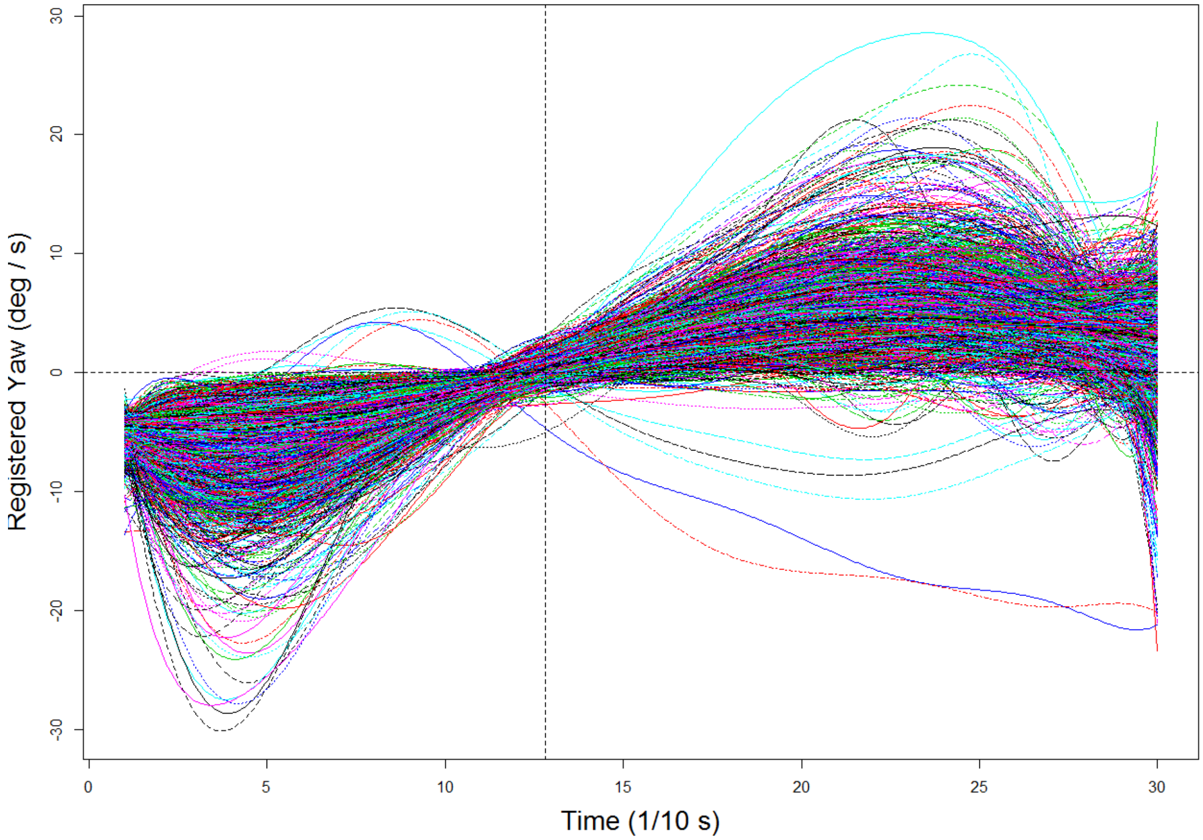

Figure 8. Plot. Continuous registered data (lambda = 2) for approximately 11,000 yaw rate triggers.

As can be seen in Figure 8, the alignment of minimum and maximum values (peaks and valleys) is better across the curves than was observed with just the landmark registration shown in Figure 7. The results from the continuous registration were used in the development of classification models.

YAW RATE CURVE CLASSIFICATION

Once curve registration was completed, additional steps were required in order to build a model to classify yaw rate curves. The following sections briefly describe the selected classification method, the estimation of classification error, an approach to group imbalance between valid and invalid yaw rate triggers, and an approach to overall classification assessment.

Penalized Discriminant Analysis

Linear discriminant analysis (LDA) is a well-known and widely used classification procedure.(9-13) Although LDA performs well in many contexts, its simplicity causes it to perform poorly in others.(14) In an attempt to overcome the shortcomings of LDA in dealing with many highly correlated predictor variables, such as when trying to classify curves, Hastie et al. proposed penalized discriminant analysis (PDA).(15) PDA casts LDA into a regression context and makes use of the relationship between LDA, canonical correlation analysis (CCA), and optimal scoring. Optimal scoring is a method of translating the categorical variables used to denote class membership into quantitative variables.(5) The interested reader may refer to Hastie et al. for details on PDA.(15)

Cross-Validation

A challenge when assessing the performance of a classification model is proper estimation of classification error. A standard approach to this problem is to divide the analysis data set by random allocation into a training portion on which the classification model will be developed, and a testing portion on which the model will be tested.(10,11) However, in the case of the yaw rate trigger data set, class imbalance between the invalid and valid triggers makes this approach difficult to implement. This is particularly true when event severity is considered. With only five crash events, splitting between a training and a test data set would further dilute this rare and important class.

Another approach when the entire data set cannot be partitioned into training and test sets is to make use of k-fold cross-validation.(10,11) In this approach the data are partitioned into an arbitrary number of groups or folds with training and testing occurring on combinations of the folds. A 5-fold cross-validation scheme was used for error rate assessment in evaluating yaw rate triggers. For each iteration of the process, model training was completed using data from four of the folds and model testing was done using data from the fifth fold. This proceeded until each fold had been withheld from training and used as the testing set. The classification error was then taken as the average of the estimates from the five testing folds.

Down-Sampling

A common problem when building a classification model for a data set with severe class imbalance is that the classifier is trained to essentially ignore the minority class in favor of the majority class.(16) This is particularly problematic as the minority class is often the class of interest. This behavior occurs as classifiers are often designed to optimize overall error rate, so the incorrect classification of a small number of minority cases when compared to the correct classification of a much larger number of majority cases is insignificant. Different methods exist to address class imbalance, including cost-sensitive learning, up-sampling minority cases, and down-sampling majority cases.(16) Down-sampling of majority cases consists of randomly selecting cases from the majority group for inclusion in evaluation. The number of cases selected is typically around the point of a 50% to 50% split between majority and minority classes in the analysis data set.

For the yaw rate trigger classification, down-sampling of the majority class (invalid trigger) was implemented and tested at several different proportions between the majority class and minority class (valid trigger). A configuration of a 42% invalid to 58% valid was selected for use in training the classifier. The testing fold described in the cross-validation section had a configuration similar to what was observed in the complete data set, 94% invalid to 6% valid.

Bootstrap Samples

As sampling was implemented to deal with class imbalance, a method for overall assessment of classifier performance was selected that could take into account sampling variability. Bootstrap sampling accomplishes this by repeatedly sampling, with replacement, from an available data set and then using estimates of the measure of interest from each of the samples to estimate accuracy.(10,11)

For the yaw rate classification problem, 1,000 bootstrap samples were drawn in order to approximate the sampling distribution of the proportion of yaw rate triggers that were correctly classified. In particular, the estimation was done at the event severity level as retaining crashes and near crashes is more importation than primarily retaining critical incidents. The bootstrap sampling was accomplished by completing the process of 5-fold cross-validation, down-sampling, classifying with PDA, and testing 1,000 times. Using the 1,000 bootstrap samples, 95% confidence intervals for the proportion of correct classifications were calculated with estimates of the 2.5th and 97.5th percentiles.

CHAPTER 3. RESULTS

Although several levels of down-sampling were investigated, the results presented here focus on training with 42% invalid and 58% valid, with testing done at the observed 94% invalid and 6% valid. Estimates of the proportion of correct classification, by severity, are shown in Table 5.

Table 5. Proportion of yaw rate triggers correctly classified.

Severity	2.5th Percentile of Proportion	Average Proportion	97.5th Percentile of Proportion
Crash	0.8	0.94	1
Near Crash	0.71	0.83	0.92
Critical Incident	0.76	0.83	0.87
Invalid Trigger	0.34	0.42	0.49

The average or mean proportion of the yaw rate triggers correctly classified as valid is quite high. For those triggers associated with crashes, the mean proportion correctly classified is 0.94. For near crashes and critical incidents, it is 0.83. The mean proportion of correct classification for invalid yaw rate triggers, meaning that they were classified as invalid, is somewhat lower at 0.42. For crashes, 95% of the observed proportion correctly classified are between 0.8 and 1. For near crashes, the percentiles are 0.71 and 0.92, and for critical incidents 0.76 and 0.87. For invalid triggers, 95% of the observed correctly classified proportions are between 0.34 and 0.49. Based on the estimate of the proportion of correctly classified, the average proportion of misclassification for crashes, meaning that they were classified as invalid, is around 0.06. For near crashes and critical incidents, the average proportion of misclassification is around 0.17. The proportion of misclassification for invalid triggers, meaning that they were classified as valid, is around 0.58.

In order to get a sense of the distribution of the proportion of correctly classified yaw rate triggers, side-by-side box plots, by severity, are shown in Figure 9. Box plots are a quick, graphical way to display distribution summary information. The ends of the rectangular box (from the bottom of the plot to the top) are estimates of the 25th and 75th percentiles, respectively. The thick black line is an estimate of the 50th percentile or median. The dashed lines extend to 1.5 times the interquartile range (IQR), which is defined as the difference between the 75th and 25th percentiles. Any observations beyond this range are indicated by a circle and may represent possible outliers.

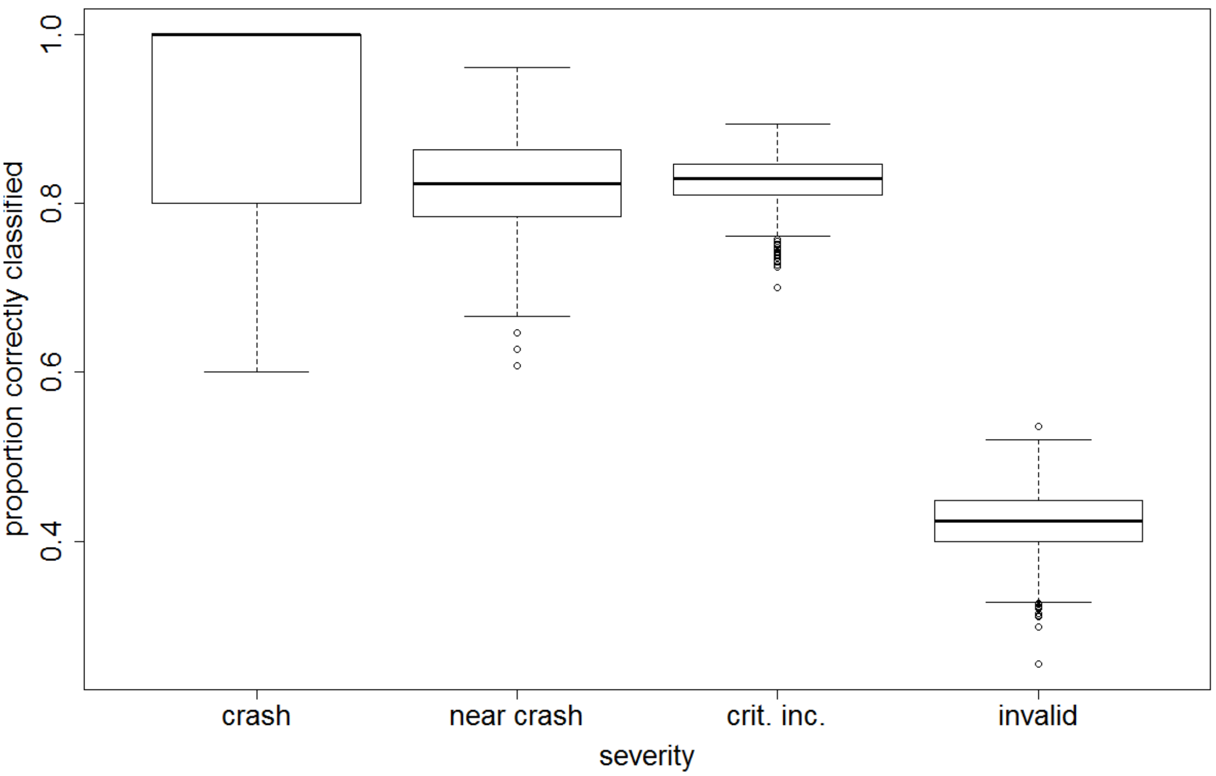

Figure 9. Plot. Side-by-side box plot of proportion of correctly classified yaw rate triggers by severity.

The variability in the proportion estimate for crashes is substantially higher than for near crashes, critical incidents, or invalid triggers, as indicated by the relatively large IQR. This is likely due to the small number of crashes available for testing and training, as shown in Table 4. However, as indicated by the estimates in Table 5, the mass of proportion estimates for crashes exceed 0.8. The range of estimates for the near crash, critical incident, and invalid groups are relatively compact and roughly symmetric.

CHAPTER 4. DISCUSSION

The results in Table 5 demonstrate that a classifier can be developed that retains a large majority of valid yaw rate triggers across different severity levels while also reducing the number of invalid yaw rate triggers considered for video validation. Although the proportion of correct classification for invalid yaw rate triggers is somewhere in the range of 0.34 to 0.49, for larger-scale studies this would result in a significant reduction in the number of triggers presented to video analysts for validation. As an example, consider the set of yaw rate triggers used in this analysis, as described in Table 4. A 42% reduction in the number of invalid triggers (10,330) would mean that approximately 4,300 fewer invalid triggers would have been flagged for video validation. Yaw rate triggers marked by the classifier as valid would be reviewed by video analysts to assign a severity designation for valid events and to cull the invalid triggers that were erroneously identified as valid.

The results of this investigation highlight the trade-off that often has to be made when attempting to retain a substantial number of actual items of interest while also rejecting as many invalid items as possible. In order to retain as many valid yaw rate triggers as possible, a choice was made to also retain some invalid yaw rate triggers. The need to retain some invalid yaw rate triggers in order to capture a large proportion of valid triggers is not surprising given the significant amount of overlap shown in Figure 6 through Figure 8.

The overlap and difficulty in distinguishing between valid and invalid triggers may be due to the scenario definition, which, although useful, is quite generic. Additional elements in the scenario definition, such as the presence or absence of a lead vehicle, might aid in classification by removing variability among curves due to driving scenario differences.

A second possible reason for the significant overlap is the nature of the safety surrogates used in analysis. As near crashes and critical incidents represent cases in which evasive maneuvers were successful in avoiding a crash,(1) it may be difficult to fully distinguish them from invalid triggers based solely on a single kinematic measure.

It should be noted that the registration approach used to align curves is quite important. Classification rates showed improvement with the transition from the landmark method to the more sophisticated continuous method. As continuous registration methods are an active area of research,(8) it is possible that different continuous methods would improve classifier performance.

Follow-on work to this effort could include validating the results against another naturalistic driving study data set, such as 40-Teen, examining the impact of newly developed curve registration methods, and evaluating newly developed functional classifiers.

REFERENCES

1. Dingus, T., Klauer, S., Neale, V., Petersen, A., Lee, S., Sudweeks, J., . . . Gupta, S. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment* (No. HS-810 594). Washington, DC: National Highway Traffic Safety Administration.

2. Schreiner, C., Torkkola, K., Gardner, M., & Zhang, K. (2006). Using machine learning techniques to reduce data annotation time. In *Proceedings of the Human Factors and Ergonomics Society, 50*(22), 2438–2442.

3. Basu, S., & Meckesheimer, M. (2007). Automatic outlier detection for time series: An application to sensor data. *Knowledge & Information Systems, 11*, 137–154.

4. Graves, S., Hooker, G., & Ramsay, J. O. (2009). *Functional data analysis with R and MATLAB*. New York: Springer.

5. Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis* (2nd ed.). New York: Springer.

6. Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies*. New York: Springer.

7. Eubank, R. L., & Hsing, T. (2007). *Functional data analysis*. Paper presented at the 32nd Annual Summer Institute of Applied Statistics, Brigham Young University, Provo, Utah.

8. Ramsay, J. O. (2011). Curve registration. In F. Ferraty & Y. Romain (Eds.), *The Oxford Handbook of Functional Data Analysis* (pp. 235–258). New York: Oxford University Press.

9. Rencher, A. C. (1995). *Methods of multivariate analysis*. New York: John Wiley & Sons, Inc.

10. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (Vol. 2). New York: Springer.

11. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York: Springer.

12. McLachlan, G. J. (1992). *Discriminant analysis and statistical pattern recognition*. New York: John Wiley & Sons, Inc.

13. Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (2nd ed.). New York: John Wiley & Sons, Inc.

14. James, G. M., & Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sample curves. *Journal of the Royal Statistical Society, Series B, 63*(3), 533–550.

15. Hastie, T., Buja, A., & Tibshirani, R. (1995). Penalized discriminant analysis. *The Annals of Statistics, 23*(1), 73–102.

16. McCarthy, K., Zabar, B., & Weiss, G. (2005, August). *Does cost-sensitive learning beat sampling for classifying rare classes?* In *Proceedings of the 1st International Conference on Utility-Based Data Mining* (pp. 69–77). ACM.