Naturalistic Driving Data for a Smart Cloud-based Abnormal Driving Detector

***Abstract*– This paper examines a possible implementation of video and radar data into smart abnormal driving behavior detecting systems. A prevailing research is *SafeDrive*, which positions the application in a successful trajectory in identifying driving anomalies to improve societal transportation safety. However, there are limitations of road and environment data that could allow mobile phone-based applications to accurately evaluate driving abnormality. In this paper we propose utilizing external research from the Second Strategic Highway Research Program (SHRP2), which can provide video and radar data for the improvement of detecting and evaluating a driver's atypical behaviors. The collection of data allows a deep study on safety, renewal, capacity, and reliability in a vehicle's actions that could permit transportation organizations to efficiently assess security operatives. Additionally, the evaluation of gathered data encompasses tools that determine and discuss features of both a vehicle and its driver, which influence driving styles linked to the near-crash or crash events. The Naturalistic Driving Data (NDD), supplies detailed information and examination of a human-vehicle-interaction to better comprehend abnormal driving behavior and critical factors that lead to most traffic accidents.**

Keywords—Naturalistic Driving Studies; SHRP2; Driving Behavior; Human-Vehicle Interaction; Big-Data; Conditional Logistic Regression.

I. INTRODUCTION

Automotive technology research has skyrocketed, influencing the development of technologies applied, which address the issues in regards to transportation and driving behaviors. The assessment of abnormal driving behavior has motivated much research on collision related events. There are several applications modeled to work on a real-world cloud-based Internet of Vehicles (IoV) platform for the detection of abnormal driving, as it enables acquiring data from various vehicles' parameters [1]. However, the data from past research has been limited. First, it does not consider key elements that examine a driver's driving style extensively, such as the analysis on a driver in moments previous to a crash and the environment in which the vehicle is at [1]. Second, the lack of such data is also needed for a more precise ascertainment of abnormal driving behaviors that may be erroneously categorized as "normal" driving styles. Therefore, to ameliorate the distinction of the source of anomalous driving and unsafe aspects of the environment it would require streaming video data and radar data. Thirdly, the analysis on large amounts of data has faced difficulty in achieving useful conclusions due to the labor required to compute methods of sampling, integration, and analysis [7]. Thus, it is important to further probe into all factors involved before and after a near-collision or a collision.

In general, Naturalistic Driving Studies (NDS) permit a face-to-face approach, as the data on driving behavior is reviewed on a daily basis and in their natural settings. In order to successfully take advantage of NDS, project Second Strategic Highway Research Program (SHRP2) has enhanced the Naturalistic Driving Data (NDD), as it narrows the data and determines elements that are of priority in research related to the issues that concern driving safety. The implementation of this type of data utilizes the usage of two focus areas for the research–which are *safety* and *reliability*–for the necessary data needed to improve poor driving behaviors. Also, this type of NDS examines large volumes of data collected from pre-collisions and collisions of the tested cars, benefiting evaluation for prevention and detection.

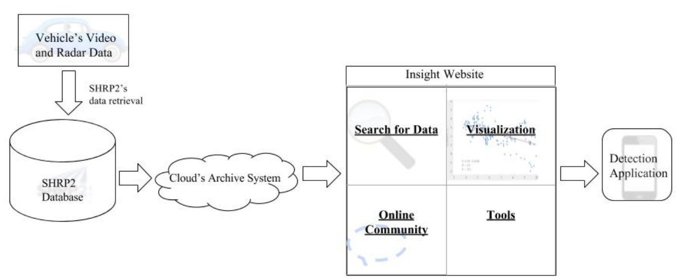

Figure 1.1 Overview of implementation of NDS into an abnormal event detector application

The leading big-scale naturalistic driving study has based its studies on the observation of volunteer drivers on a quotidian basis. The gathered information has been able to assess questions regarding speeding, tailgating, cell phone use, and even alcohol intake–as individuals portrayed their actions in real life conditions [4]. So far, SHRP2 has recorded two years of data from more than 3,100 volunteers with ages ranging from 16 to 80 years old; they were recruited from around six states that include New York, North Carolina, Florida, California, Washington, and Indiana [3]. Additionally, there were various types of vehicles studied for more accurate specifications: passenger cars, minivans, SUVs, and pickup trucks. In 2014 it was claimed that its (NDS-DS) dataset contained over 33,000,000 travel miles originating from 3,800 vehicle-years of driving [2]. Therefore, the broad geography covered by NDS has enabled depicting various driving styles in relation to a variety of environments.

We validate the NDS as it operates on two analytical methods that include and examine the previously mentioned data to assess driver behavior in relation to the driver's environment and its revolving characteristics. Taking into consideration the limitation on historical data from other studies, our proposed usage of NDD could supply missing data needed to accurately sense anomalous driving. Furthermore, the adoption of a Data Acquisition System (DAS) provided by the American Computer Development Inc., documents assembled data as soon as the studied volunteer start operating the vehicle for real-status-aware evaluation [3].

Our implementation of NDS's data into a cloud-based IoV platform, has formulated "normal" driving behavior to recognize abnormal driving. We introduce the NDS of *SHRP2*, as it contains video and radar data that could aid the event-detector from vehicular information. The purpose of this paper is to advocate current abnormal driving detection applications reinforced by the second Strategic Highway Research Program, as it offers information that supplements research to promulgate efficiently the detection of driving anomalies for safer roadways.

In summary, the main contributions are the following:

- The paper presents an NDS that is able to combine vehicular, video, and radar data to enhance discovery of abnormal driving detection.
- We utilize conditional logistic regression as a method to develop models of the SCE's risk as a function in relation to the driver and the vehicle's environment.
- We create a casual model represented in a directed graph for crash-related events, and a grading conflict severity scale, modeled after Henrich's Triangle hypothesis.
- Finally, simulation results show the benefits of a database with an online archive system for low-cost data mining, preservation of its digital assets, and support towards promoting research on transportation security.

The paper is organized as follows. Section II displays the overview of NDS Database. Section III displays the analysis of data elements. Section IV presents the use of mathematical methods for anomalous driving analysis. Section V elucidates the evaluation of data. Section VI concludes this paper.

II. Design Overview of NDS Database

While several research projects have created detection techniques to cover anomalous behaviors, the modeling of such has faced challenges in need of improvement. The more accurate parameters are set to detect the causation of crash and near-crash events, a high density of data needs to complex evaluation with various tools to diminish incorrect assessments of driving styles and detect the source of anomaly. As previously mentioned, the use of an NDS database can assist applications with real-world video and radar data collected from highways, and state inventory (from the RID) that also identifies and emphasizes pre-crash events. For that reason, the philosophy behind the NDS, which is customer-oriented, and focused on the agent's driving behavior primarily to shape safety technology techniques, is favorable for any application. Specifically, *SHRP2* is supported through a Cloud Archive System that acquires NDD to be implemented into its Insight Website. The overview of acquisition of NDD from an NDS database, is displayed in Figure 1.1, where the model addresses the gathering of data into SHRP2 database, which is then used for the project's purposes that are utilized in the improvement for abnormal driving detection.

A. Database's Focus Areas

In general, the meeting objectives of SHRP2 revolve around the sharing of information to benefit projects analyzing risks and transportation, designing data tools that can facilitate data sharing, to improve driving behavior, as well as to ease road congestion.

B. Cloud's Archive System

1)Realiability Archive

The advantage of the preserved data is that it consists on raw data–mostly from cameras, vehichle instrumentation and roadway data [8]–all of which will soon sustain worldwide engagement. The end users benefit from the unlimited access to *SHRP2's Archive*, since most data originates from the project and public branch with the goal to maximize the usage of metadata [13].

2) Online Community

Furthermore, their Archival Portal features the "Web 2.0," an uncomplicated system that can be smoothly implemented and applies guiding principles for the user interface [11]. It aims to be systematically open for operating users that differ in platforms, languages, and network infrastructures. The Archival Portal has an online community of users that can interact with each other and independently navigate. The information displayed from SHRP*2* and other Safety Projects offer results and allow refinements, customer support and administration, study locations, and others–which mostly may include FAQ and Help interfaces [13]. Particularly, S*HRP2's* InsightWebsite would provide *SafeDrive* with the participant's information, demographics, questionnaires, vehicle information, summary variables, and interactive maps with the driven roads of the study [10].

3) Implementation Assistance

Since 2013, *SHRP2* an Implementation Assistance Program (IAP) has supplied *SHRP2* with financial support, and collaborated with transportation agencies by assisting their technical needs. The focus areas of the IAP are renewal, capacity, and reliability, for the beneficiaries under *SHRP2* [14]. On June 2016, *Round7* Implementation Assistance Program was announced by the FHWA. This is important, as the FHWA has previously operated with highway agencies from all states, D.C., and Puerto Rico–and facilitates provision and managmement of funds [8]. The recipients can expect to administer programs with smart cost operations.

III. Analysis Of Data Elements

The metrics adopted by NDS for reviewing data elements that are gathered from a tested vehicle for evaluation, contribute anomaly driving detecion applications as follows:

- *Validation*: By reviewing the trigger specifications and the eyes-path movement to determine risks from the previous events and at the event of a collision allow a more accurate analysis that could be exhibited in time segments with visual contribution.
- *False alarms:* The inspected triggers reduce the chances of an SCE to go unnoticed, which would complement the application's evaluation. This allows to change the structure of the current graph as needed, since it lacks of historical data and sensing false alarms [1].

A. Safety and Reliability from NDS Database

The two areas that need be implemented to greatly benefit the *SafeDrive* program are reliability and safety. Both offer data obtained from more than 3,000 drivers (over a 1 to 2-year period) stored in two databases that include video-recording and radar data. Figure 4.1 shows some usages from the NDS focus areas for crash and near-crash.

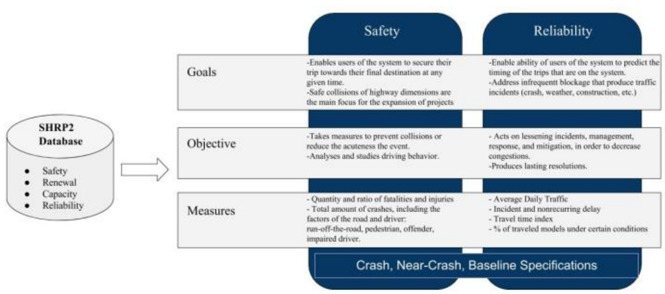

Figure 4.1 Overview of implementation of NDS into an abnormal event detector application

B. Abnormal Behavior Classifications.

The three main constituents that come along collisions and abnormal driving behavior to assess the data's categorization correctly are: crash, near-crash, and baseline definition and identification. Table 4.1 describes them. The following methods expose some of the metrics included in the project, which help constructing the safety framework:

i) Automatic crash notification: DAS through kinematic signature detects uploaded data that could indicate the crash. Followed by video verification to determine if SCE occurred.

ii) Data bank report: data from reported crashes from a site. Also supported with video verification from the DAS.

iii) Identification of data acquisition analysis: video evaluation that helps verify non-identified behaviors.

iv) Active triggers' performance: "triggers" algorithm revolving around the ACN and behavioral signatures from various entries [10].

C. Cloud Archive System

NDS aims to conserve NDD for various decades in a Reliability Archive based on a cloud-system, to provide researchers of transportation, the necessary resources for future developments. In order to employ a feasible data archival system that contains the RID and NDS data altogether, from relevant projects that will contribute the website's raw and aggregate datasets, a cloud archival system is used to link the gathered data from the project's research into the storage for later use by user researchers. Figure 4.2 shows a representation of the method. It utilizes a METS document structure, which is an information package that can handle various categories via XML encoding [13]. The Digital Library Foundation developed it, and the XML serves to transport and store the data in an eye-readable format for the ease of standardizing data and metadata in an organized manner [15].

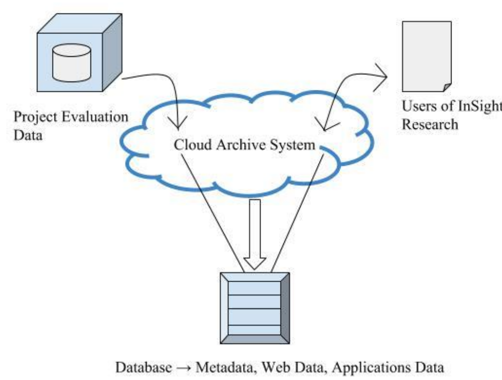

Figure 4.2 Implementation of Cloud Archival System to support the sharing of collected data

D. InsSight Website

The InSight Website supplies the collected data from SHRP2 NDS accessible to researchers. The website provides the assessments and descriptions of the volunteer drivers, a custom query capability constituted of an extent criterion from various datasets, summary of the NDD with visual description with graphs and tables, the description of vehicles evaluated.

a) Environment Awareness Data: The fact that many behaviors correspond to environmental factors, limitates projects (like *Safedrive*), as the applications consider status-aware behaviors without undertaking environmental factors.

Collision Categories based on their Severity	
Crash	**Level 1 (Severe):** airbag deployment, injury of driver or pedestrian, vehicle rollover, towing required. *Impact*: change in speed > 20 mph. **Level 2 (Moderate):** sign strikes, animal strikes. *Impact:* worth approximately $1,500 for damage repair (video estimated). **Level 3 (Minor):** small anima strikes, curb and tire strike during traffic. *Impact:* minimal to no damage through physical contact or road departure. **Level 4 (Low):** struck tire, clipping a curb during right turn. *Impact:* little or no risk
Near-Crash / Crash-Relevant Conflict	**Not a crash:** no contact with any object, no road departure. **Not premeditated:** maneuver that is not planned by subject, like rapid evasive maneuver by only one car, if more are involved (aggressive lane change = unseen vehicle in adjacent lane). **Evasion Required:** evasive maneuver to avoid collision (steering, braking, accelerating). **Rapidity Required:** sudden response within a time frame among the subject's reaction and the possible strike.
Non-Participant Conflict	Video recorded event where the participant driver is only a third party to the event: crash-relevant, near-crash, crash.
Non-Conflict	A behavior that could be considered as "normal driving" taking into account the timing of the "flagged event" from the time-series of the data. The driver's response to a circumstance; there is neither evasion nor risk.

Table 4.1. SCEs & Severities Definitions

Therefore, the improper judgement affects the assessment of situations, so to prevent estimating false alarms, deep examination of environmental factors are essential. Acquiring video and radar data to evaluate in detail the information from the collected data, leads to acknowledging the missing environmental elements. In addition, this allows an upgrade when evaluating data with online abnormal detectors, which are limite by this data and need to enhace the tructure of a graph-system that relies on states measurd in different times.

b) Safety Critical Events and Trigger Types. The metrics from the NDS in this case consists on determining what is "normal" driving behavior, which requires samples from baseline events. In this paper, the analysis conducted remarks these two scenarios as safety-critical events (SCEs): crash and near-crash. SHRP2 identifies SCEs based on **trigger specification, validation,** and avoiding **false alarms**.

E. Framework and Methods

a) Coded Trigger Specifications. The uniqueness of the project comes from the combination of SHRP2 and Virginia Tech Transortation Institute (VTTI) data and trigger coding. NDS gainings mostly have as frame of reference event-trigger interpretations [13]. The performances of the tracked triggers are used to flag errors with the aid of a dozen trigger algorithms employed, along with the video examination, with the goal to improve the system that aloccated the driver's physical realistic conditions in real time [13]. The monitored data help *SafeDrive* on its state graph model, which analyses and correlates the behaviors of acceleration and deceleration of a vehicle in relation to the gear position [1].

Similar to SafeDrive's time-stamps of online stream data measurements, SHRP2's trigger specifications also revolve around the timing in relation to deceleration, acceleration, and swerving. Their analysis employs timestamps at their commencement. In fact, each of the trigger types have specific characteristics, in which their threshold is exceeded for at least one timestamp. Also, if multiple triggers are within a 2 second window, then they are typically incorporated with similar potential events. Here, some trigger types are presented, and whose performance is constantly monitored to mold the amount of SCEs, identified by each kind of trigger:

i. Longitudinal Deceleration
ii. Longitudinal Acceleration
iii. Freeway Deceleration
iv. Lateral Acceleration
v. Swerve
vi. Yaw Rate
vii. Longitudinal Jerk
viii. Steering Evasive Maneuver
ix. Advanced Safety System

Through the finding of SCEs by trigger type, a set of cases can be examined in accordance to the threshold of the speed. In fact, studies made under this concept have actually shown that near-crashes are detected due to high increase of velocity, and crashed through the reported information on a location, as well as with the Automatic Crash Notification (ACN), which is patented by the VTTI [13]. Especially, since the trigger algorithm with unknown values that is used is also fit for *SafeDrive's* real-time data extraction, could also be used for SG comparison [1].

b) Road Performance Data. The ability of SHRP2 to target several driving styles is also achieved through their video-data analysis. Since 35% of American drivers drive more aggressively rather than passively, the collected data could establish the difference between what is considered to be "normal" and actually an aggressive driving behavior [1]. Therefore, the data of SafeDrive, which is based on safe driving under certain secure instructions, could be enhanced with the ordinary driving commands of the conductors examined. The detection of abnormal driving behavior is majorly influenced by a person's habits and reactions to a situation. In this case, the triggers activated through a driver's response towards a near crash and after the collision, are components that also influence and has an effect on societal transportation safety.

c) Glance Behavior. In order to determine the trigger distractive actions, the glance behavior that SHRP2 focuses on allows the development of a predictive metric that estimates risk. Since glances occur more frequently than distracting activities, these thus enhance the odds ratios of the distracting activities that were analyzed [14]. The off-path-glance metrics provide indicators of crashes and near crashes, as well as indication of risk and the correlation between distractive activities or surroundings with such events. SHRP2 categorized three different glance metrics:

i. Off3to1: Focuses on 3 seconds until the 1 second before the crash or time of collision.
ii. Mean duration of off-path glance (max.off): Examines 12 seconds before the collision point [14].
iii. Metric of uncertainty (m.uncertainty): Also analyses 12 second previous to the class, but is the mean value of a compound measure that bases on the uncertainty of driving conditions [14].

In fact, the analysis on glances has concluded that crashes need not be based on the duration of an off-the-road glance, as these are rare. In fact, the majority of crashes were in relation with short glances of around 2 seconds associated with high situation change rate. The mechanism for crashed results based on vehicles' kinematics and visual cues reconsidered what a dangerous glance consists of [14].

IV. Mathematical Methods for Anomalous Driving Analysis

To estimate an event-detection and abnormal behavior, we show approaches to estimate conflict probability and the severity grade given the collected data on crash and near crash events.

A. Conditional Logistic Regression

Logistic regression is used throughout the project to compare two possible event scenarios to estimate the probability of the two, which are designated as two different types. Thus, it estimates the function in relation to the probability of a crash or near-crash scenario. The project can calculate an *odds ratio* to forecast the risk of events. By employing **conditional logistic regression,** we specify the distracting activities that may provoke the SCEs, in order to model crash and near crash risks as functions of the distractive actions or things [13]**.**

1) Logistic Regresson for Odds Ratio

This binary method bases its concept on probability of success above failure, where the odds increment as the probability does too, and conversely (see Equation 5.1) [16]. Given that it is a linear model; variables are assigned to create the function of predictors corresponding to the driver behavior. The purpose in this case constitutes the success of events resulting in a crash or not. The odds are determined from the probabilities and ranges from 0 to infinity. For instance, odds of success and failure are mathematically displayed in Equation 5.1.

$$odds(success) = \frac{p}{(1-p)} = \frac{p}{q} \mid odds(failure) = \frac{p}{(1-p)} = \frac{q}{p} \qquad (5.1)$$

where p is the probability of success and q is the probability of failure.

The approach for probability of a collision is as follows: A variable y is assigned value of 1 when resulting in crash, and 0 otherwise. Then, the data set with n predictors that entail a feature depending on the circumstance, driver, or conduct [2]. It can also be expressed in terms of x, as $x_1, x_2, \ldots, x_n,$ where each x designated with a particular attribute of a behavior, driver, or situation (distractive actions, glance pattern, other human factors, etc.) [13]. As previously stated, the goal–which takes into account the predictors–predicts **risk** of an impact. Equation 5.2 shows the formula used for *odds* of crashing. It defines the predicted probability (p(x)) as the probability for crash outcomes, β_0 and β_n are the regression parameters, constrained in lie between 0 and 1, resulting in

$$\beta_s \quad p(x) = \frac{e^{\beta 0 + \beta 1 x}}{1 + e^{\beta 0 + \beta 1 x}} \rightarrow \quad p(x) = \frac{1}{1 + e^{-(\beta 0 + \sum_{i=1}^{n} \beta 1 x n)}} \qquad (5.2)$$

Then, the equation is rewritten for the purposes of predicting the outcome as a linear function of predictors, by defining the odds of crashing as a success ($\frac{p}{(1-pn)}$), and turns the logic to the left side, to employ the natural logarithm (ln) of odds of crashing to establish a linear function from the variables on the right side of the equation (the predictors) [14]. Lastly, it considers M events, starting from $n = 1$. Equation 5.3 shows the equation for M events derived from the first formula but focusing on the linear portion.

$$\ln\left(\frac{p_n}{1-p_n}\right) = \beta_0 + \beta_{1_{xi}} + \ldots + \beta_{n_{x}i_n} \qquad i=1, 2, \ldots, M \qquad (5.3)$$

From the logistic regression formulation, Equation 5.2, displays the *ln odds* ratio on the left side, so a constant predictor x such as inverse Tau, slope parameters are linked to demonstrate *ln odds* of a crash for the increase of such unit (Tau) at the values already predetermined by other predictors. For a continuous x with inverse Tau, the gradient parameters also equal the ln odd of a collision. Altogether, these create an estimate of β_z , and a new constituent that calculates the blunder part that measures demonstrated instability utilizing any absence of attack of the model to the information.

Algorithm 1 Odds Ratio to Establish Crash or Near Crash
Input: Data set D with n-predictors.
D = p(n) for $n = x_1, x_2, \ldots, x_n$ = p(x)
Output: A concise representation of odds ratio of either success or failure.
(1) odds(success); (2) odds(failure)

```
1: Establish boolean β0 and βn as parameters of p(n) ∈ D
2:  0 → β
3: initialize y and set its elements
5:     if gi = 1, denote yi, if gi = any number ≠ 1, denote yi = 0
6:  Let p1(x; θ) = p(x; θ) and set parameters θ = {β0, βn}
7: // odds formula into probability
8: for each βs p(x) in D
9:         define as odds for success: change to natural logarithm
10:         //contrast parameters
11:         if ln(odds) = y, then define as crash
```

```
        else ln(odds)= z, and define as near-crash
//measure instability
calculate constituent β_z with x and y
Return β_z
end
```

Given the data that uses logistic regression, Algorithm 1 describes the data sequence of given predictors.

B. *Trajectory-Based Reconstruction*

A method undertaken by the NDS calculates variables from the historical data to dictate the driver's circumstances that lead to crash related events, when–in addition to the subject vehicle–one or more vehicles are involved.

Causal Model as a Directed Graph. The set of variables are described in a structural formula that reacts accordingly to the alteration of another model's variable [6]. The graphical model is shown in Figure 5.1. Node *a* represents the environment, node *x* denotes the evasive behavior, and node *y* the action determined by causes regarding *a* and *x* as functions. Therefore, $y(a,x) = 0$, when *a* and *x* do not produce a crash, otherwise, it equals 1.

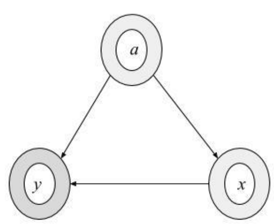

Figure 5.1 Directed Graph for Event Collisions

To equate the prediction, an initial speed and breaking deceleration of the subject car are set as variables v_1 and c_1; similarly, v_2 and c_2 are assigned for a succeeding vehicle. Then, equation 5.4 shows the equation that describes the driver's forwards movement, along with the evasive action, with variables designated f_1 and e_2 respectively [6]. A collision is evaluated when the stopping distance accessible to the next driver is less than what would be needed to stop without actually crashing into the leading car.

$$v_2 e_2 - \frac{v_2^2}{2c_2} > f_2 v_2 - \frac{v_1^2}{2c_1} \tag{5.4}$$

The case where the deceleration of a driver is the evasive action, then the collision adopts all the variables, which become elements of node *a*, and the evasion is represented byc_2 [14]. Thus, the function of the generic collision is depicted in equation 5.5.

$$y(a,\mathrm{x}) = \begin{matrix} 0 - \ if\ v_2 e_2 - \frac{v_2^2}{2c_2} \leq f_2 v_2 - \frac{v_1^2}{2c_1} \\ 1 - if\ v_2 e_2 - \frac{v_2^2}{2c_2} > \ f_2 v_2 - \frac{v_1^2}{2c_1} \end{matrix} \tag{5.5}$$

On account of that, the events leading to crash, where two vehicles are involved, are established through the appliance of such simulation, which utilizes the formulated equation. The application of this method improves and follows up on previous methods employed to detect *rare* events that result in collisions, since most scenarios are usually near-crash situations. Essentially, the trajectory-based reconstruction, narrows down collision data even further.

C. *Grading Severity*

Heinrich's Triangle. In order rank the seriousness in increasing order of the quarrels that emerge in relation to traffic-conflict methods employed, we employ Henrich's triangle hypothetical model. This one expresses the gravity of neighboring events in a pyramid to portray the scale in increasing order, but in decreasing frequencies, since crashes are more infrequent [6]. Figure 5.2 displays the levels of conflict occasioned by abnormal driving scenarios that lead to near-crash and ultimate collisions, which are related to the evasive actions displayed in vehicle-to-vehicle interaction.

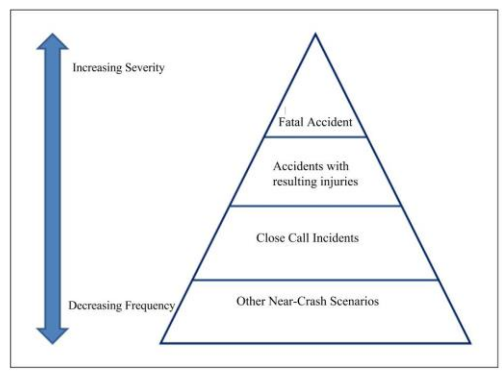

Figure 5.2 Heinrich's Triangle based on severity of the incident and its frequency

The technique used to calculate the severity is based on Hayward's time to collision (TTC) technique, which effectively measure rate of severity under traffic conditions and has also been used to have a biased judgment on normal driving behaviors [5].

$$Severity = f\ /\ (\mathrm{TTC/CS}) \tag{5.6}$$

Equation 5.6 displays the function, where higher values of TTC/CS are the least dangerous. The function is inversely proportional to minimal reduced speed required to stop a vehicle completely. Thus, enhancing providing accurate statistic of crash-related events, in particular, when crashes take place, are important in evaluating driving anomalies.

V. Evaluation Of Data

We evaluate the performance of this work on a real-world IoV system. It is designed as a cloud-based IoV architecture in which driving data are collected with OBD devices plugged in the vehicles. Each OBD device has integrated a wireless communication module to maintain connections with the back-end server and send the collected data to the server with an adjustable time interval. Over 29,000 real vehicles from 60 cities have been connected to the system. This system collects around 0.2 billion data instances daily.

SHRP2 used timing intervals to determine specific combinations of the off-road glances in relation to the kinematics of a collision. In a scenario where short glances were involved in rapid changing situations resulted in a crash. On the other hand, long off-road glances produced a crash when involved in a situation where the changes were rather slow [9]. The pattern among these events was also confirmed through a what-if simulation (process of changing parameters) that aimed to comprehend the mismatch allying the glance and the lead-vehicle road closures rate [14].

Through the examination of the glance allocation, the video recording, and trigger types, a driver's reaction may also be determined and categorized in the analysis of a near crash or crash situation. The evaluation on the pattern reaction mechanism was able to determine that the reaction of drivers in a crash and near crash differ greatly. The reactions that involved braking and steering had a greater proportion of reactions in near crashes, compared to those resulting in crashes. Based on the proportion of a reactions displayed on certain events that resulted in crash or near crash, Figure 6.1 shows the distribution of both in certain situations that resulted in a reaction or no reaction at all.

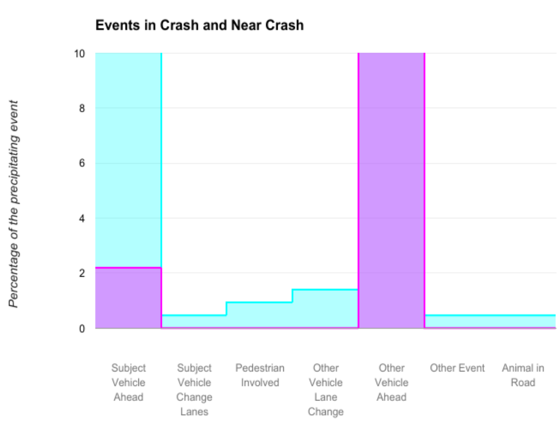

Figure 6.1 Near crash and crash – Percentage of precipitating events

Overall, crashes displayed a larger proportion of no reaction (17%), and it the greatest proportion was displayed mostly when another subject vehicle was involved, in cases of breaking with lockup consisted on 15%. In near-crash-related events, braking without lockup consisted on 75% of the cases, while braking and steering 18%. Given that *SafeDrive* examines sequences of acceleration anomaly and deceleration (sudden breaking), as it considers these actions the most common anomaly, the detection is complemented with proportions evaluated by SHRP2's data [1]. In addition, with the validation of video recording data, the assessment of driving styles is improved, and the consideration of behaviors as anomalies is more factual.

A. *Evaluation of SCE's Validation*

Since the process of validation is majorly based on video data, if an SCE fails to be acknowledged, the video and parametric data is able to distinguish it. Additionally, it assists categorizing the intensity of the SCE, and to estimate the timing of pivotal occurrences [10]. Therefore, the anomaly categories of *SafeDrive* analysis could compare its scores more accurately. Additionally, the process of validation allows a set of designated events to be categorized depending of their severities. Table 4.1 gives a brief description of each.

B. *Eyes-off-Road Evaluation*

The three types of glance metrics examined by SHRP2 in various interval times consisted on glances on the road and off the road, which significantly determined the association of risky behaviors or secondary activities with crash or near-crash events. The glance distribution allows the evaluation of gradual risk in relation to these activities. Even though the Off3to1 has been considered to be the most accurate for predictions compared to the others, SHRP2 study looked to achieve an even better forecaster.

Table 6.1 Model on Linear Combinations of Glance Metrics Associated with Crash and Near Crash

Model	AIC	Model Likelihood
Off3to1 + max.off + m.uncertainty	317.07	1.00
Off3to1	320.39	0.19
Off3to1 * max.off	321.44	0.11
Off3to1 * max.off * m.uncertainty	321.59	0.10
Off3to1 * m.uncertainty	322.01	0.08
max.off	337.58	0.00
m.uncertainty	345.86	0.00

The predictive success of the model is employed to explain glance behavior and the risk of distracting activities (secondary tasks). In this case the activities were examined from the 5 seconds previous to the incident and 1 second after. Thus, to determine the contribution of glances and distracting activities, the percentage of eyes-off-path in the time window of 2 seconds that overlaps the precipitating event is used, as the glance behavior occurs at the same time interval [13].

Some of the activities that were assessed to be predicted were texting, talking or listening to the phone, and visual manual. In Table 4.3 the proportion of eyes-off-road behavior in the 2 seconds where the glance behavior overlaps with the distractive activity is shown. In this case, the prediction of the period Off1to1after (one second previous and after the precipitating event) proved to be more accurate than when modeling the distractions alone, given then lower distribution of it [13]. The lowest Akaike Information Criterion (AIC) of 339.43 that belongs to the Off1to1after is clearly the most predictive model. Therefore, it is accurate to estimate the effect and role of glances off-path with relation to the distractive tasks to the crashes and near-crashes. Nevertheless, the study then combined the successful predictive model with the each of the events, and the AIC appeared to be lower than most events alone or the model alone–achieving an even more accurate predictor for each.

Table 6.2 Contribution to Risk Estimation from Distracting Activities

Model	AICc	Model Likelihood
Off1to1after	**339.43**	1
Texting	341.66	0.33
Talking/Listening	348.53	0.01
Visual/Manual	349.59	0.01
Off1to1+ after Talking/listening	**333.03**	1
Off1to1 + Texting	**334.23**	0.55
Off1to1 + Visual/Manual	**347.92**	

Table 6.2 Displays the contribution to risk estimation based on the Off1to1 Model, where the activities and model are examined in the 2 seconds that overlap the precipitating event.

C. Conditional Logistic Regression Evaluation

The data selected for the logistic regression assessment was part of a case-crossover study, as it used selected crashed and near crashed to find an event matched from the baseline to be matched with a particular driver. Thus, a matched-base control study took place, as the matching of such situations were linked the same driver [14]. Moreover, the features put to trial under this method are distracting activities and interactions among matching variables. It also allows most of the coefficients to be deciphered, even as β_0 cannot be approximated through the conditional logistic regression [13]. Also, due to the unbiased odds ratios assigned to the predictors, these are understood as risk ratios–which are evaluated by a list of distractions.

1) Talking or listening on a phone
2) Hands free usage of phone (voice activation)
3) Non-visual or electronic actions (talking, singing, writing, reaching for objects)
4) External distraction from the vehicle (pedestrian, accident, animal)
5) Loss of attention towards the roadway (facing towards left or right, even if to the mirrors or windows)
6) Others (not specific or unknown distractions)
7) No distraction whatsoever.

Figure 6.2 displays the odds ratio of some distractive activities (variables) depending on the event type: Crash, Near-Crash (NC), and both combined (CNC).

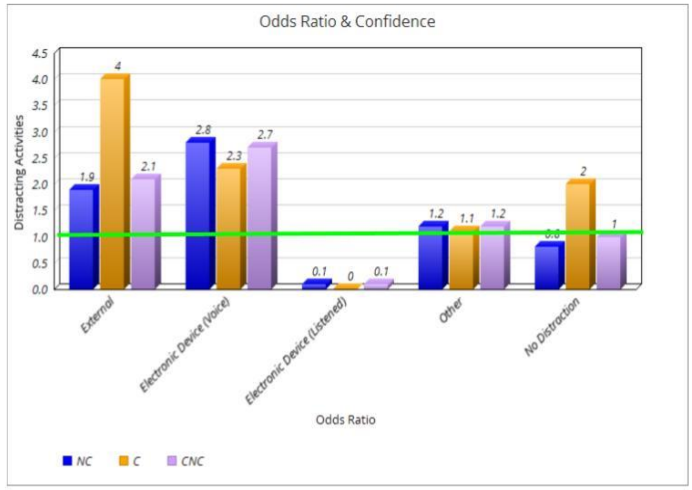

Chart 6.1 Odds ratio of Distractions

All things considered, there are around 50 different distracting activities that are examined to determine the *odds ratios*. The frequencies of some are higher than others, and some show higher levels of risk, while others are not as notably dangerous. For the odds ratios to be noteworthy, based on a 95% confidence interval, if they range fully above 1 (from y-axis in the graph), or fully below it.

Table 6.3 displays the confidence interval of each event to determine which are important. The bold numbers show that either these events are significant as they are go way above 1 (like the electronic devices distractions)–which means that SCEs are highly influenced by these—or below zero (or way below, like when there are no distractions), which demonstrate that such distractions are less likely to result in SCEs.

Table 6.3 Odds Ratio and Confidence Interval for the Distracting Activities

Distraction Type	NC	C	CNC
External Distraction	0.5 - 5	0.2 - 8	1 - 5
Electronic Devices (voice) Distraction	**1 - 6**	**0.5 -8**	**1.5 - 5**
Electronic Devices (listen) Distraction	**0 - 8**	**0.5 -8**	**0 - 0.5**
Not Looking Forward the Road	0.6 - 2	0.5 - 3	**1 - 2**
Other Types of Distraction	0.2 - 3	**0.2 - 8**	0.5 - 3.5
No Distraction	**0.5**	**0.2**	**0.5**
	Confidence Interval (Ranges from 0 to 8)		

The odds ratio analysis shows that many of these distracting activities are indeed main factors that lead to SCEs. They occur much more regularly than other activities (the ones with an odds ratio < 1). Additionally, all of these are analyzed through the video recording data, since their interpretations require to be judiciously reviewed. For this reason, SHRP2's accuracy on visual data is optimal for the detection, categorization, and distinction abnormal driving behavior that result in SCEs.

D. Stuctural Modeling Evaluation

The baseline selection has 20,000 samples available and around 5 million files of driven trips with over 3,000 identified crashes and near-crashes. From these, the many of the crash-related events consider trajectories with successive breaks and stops when driving in a freeway lane. This case involves a leftmost vehicle and a rightmost vehicle, whose trajectories collapse in the event of a crash.

Trajectories examined through video recording data allow creating a simulation of trajectories described by the model previously detailed in the trajectory-based reconstruction. To apply the model into a real-world scenario, the trajectories would adopt the physical model as follows:

$$v_k t, \quad t \le t0_k$$

$$y_k(t) = \frac{v_k t0_k + c(t - t0_k)2}{2}, \quad t0_k < t \le \frac{t0_k + v_k}{(-c_k)\,2}$$

$$\frac{v_k t0_k - v_k^2}{2c_k}, \quad t > \frac{t0_k + v_k}{(-c_k)\,2}$$

Here, $y_k(t)$ presents the position of a vehicle k at a time t, v_k is the initial rate, and a_k the increase of breaking with $t0_k$ being the initial time at which the acceleration began [6]. The adoption of this model can then be used for rear-end collision events where a rear-end collision k would have a reaction time of a driver k of $e_k = t0_k - t0_{k-1}$. Then for a driver k, its initial forward movement with a following headway among vehicles k and k-1 of $f_k = \frac{y_k\left(t_{0_{k-1}}\right) - y_{k-1}(t_{0_{k-1}})}{v_k}$ $f_k = (y_k(t0_{k-1}) - y_{k-1}(t0_{k-1})) / v_k$. With this it is possible to calculate braking rates, the initial velocities, and the time frame of the braking from the beginning.

VI. Conclusion

The focus of this paper was to identify a project that could benefit driving with safety in matters data needed for the application's further development in assessing abnormal driving behavior. Overall, *SHRP2*'s extensive database proves to have accurate data, which could complement the detection of the behavior and other factors that surround the anomaly. The evaluation result demonstrate that this work performs well in detecting various driving anomalies without using labeled training data. This work offers an ideal option to detect driving anomalies from large-scale vehicle data.

References

[1] M. Zhang, C. Chen, T. Wo, T. Xie, M. Z. A. Bhuiyan, and X. Lin, "SafeDrive: Online Driving Anomaly Detection from Large-Scale Vehicle Data." IEEE Transactions On Industrial Informatics, 2017, https://doi.org/10.1109/TII.2017.2674661.

[2] C. Kenneth. "Analyzing Driver Behavior Using Data from the SHRP 2 Naturalistic Driving Study." *Analysis of In-Vehicle Field Study Data and Countermeasure Implications*, 2013, pp. 1-4.

[3] Antin, Jon, Suzie Lee, Jon Hankey, and Tom Dingus. "The SHRP2 Naturalistic Driving Study Addressing Driver Performance and Behavior in Traffic Safety." *Design of the In-Vehicle Driving Behavior and Crash Risk Study: In Support of the SHRP 2 Naturalistic Driving Study,* 2011. doi:10.17226/14494.

[4] H. Jim. "SHRP2: Accelerating Solutions for Highway Safety, Renewal, Reliability, and Capacity." Transportation Research Board of the National Academies, 2013. Web. www.sae.org/events/gim/presentations/2012/campbell.pdf

[5] Richard van der Horst, and Jeroen Hogema. "Time-To-Collision And Collision Avoidance Systems." *TNO Institute for Human Factors*.

[6] SHRP2 Staff. "Development of Analysis Methods Using Recent Data." *National Academy of Sciences*. 2012.

[7] Implementing the Results of the Second Strategic Highway Research Program: Saving Lives, Reducing Congestion, Improving Quality of Life." Transportation Research Board. N.p., 2009. Web. Mar. 2017.

[8] Eiriksdottir Hekla, "Quantitive analysis of rear-end crash causation mechanisms based on naturalistic crash data." Chamlers University of Technology. April 20, 2016. Accessed January 2017. http://publications.lib.chalmers.se/records/fulltext/235332/235332.pdf

[9] Campbell, Ken. "The Exploratory Advanced Research Program: Automated Video Feature Extraction Workshop Summary Report." Federal Highway Administration Research and Technology Coordinating, Developing, and Delivering Highway Transportation Innovations. October 10, 2012. Accessed December 2016. https://www.fhwa.dot.gov/advancedresearch/pubs/13037/001.cfm

[10] Trent, Victor, Marco Dozza, Jonas Bargman, Christian-Nils Boda, Johan Engstrom, Carol Flanngan, John D. Lee, and Gustav Markkula. "Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk." The Second Strategic Highway Research Program, 2015, 3-128. 2015. ISBN: 978-0-309-27423-4

[11] Federal Highway Administration. "AASHTO and FHWA Announce Round 7 Recipients of SHRP2 Implementation Assistance Program ." SHRP2 Solutions Tools For The Road Ahead, June 20, 2016. http://www.fhwa.dot.gov/goshrp2/Content/documents/fhwa_aashto_rd7_ann_160616.pdf.

[12] Hankey, Jonathan M., Miguel A. Perez, and Julie A. McClafferty. "Description of the SHRP2 Naturalistic Database, And The Crash, Near-Crash, And Baseline Data Sets." Prepared for The Strategic Highway Research Program 2 Transportation Research Board of The National Academies, April 2016, 1-37.

[13] Tao, Zongwei, and Elizabeth Hess. "Requirements and Feasibility of a System for Archiving and Disseminating Data from SHRP 2 Reliability and Related Studies, 17-45." The Second Strategic Highway Program: Report S2-L13-RW-1, 2011. doi:10.17226/22881

[14] YouTube. August 22, 2016. Accessed December 2016. https://www.youtube.com/watch?time_continue=7&%3Bv=J_NwK0VJGmM

[15]] K. D. Kusano, R. Chen, A. H. Tsoi, and H. C. Gabler, "Comparison of Event Data Recorder and Naturalistic Driving Data for the Study of Lane Departure Events," Proc. 24th Enhance. Saf. Veh. Conf., no. 15–0149, 2015.

[16] Jaynes, Mike. "Can We Move Beyond the CAN Bus for Vehicle Networks?" Cyber Operations, Analysis, and Research. July 11, 2016. Accessed December 2016. https://coar.risc.anl.gov/can-move-beyond-can-bus-vehicle-networks/.

[17] Tao, Zongwei, and Elizabeth Hess. "Requirements and Feasibility of a System for Archiving and Disseminating Data from SHRP 2 Reliability and Related Studies, 17-45." *The Second Strategic Highway Program: Report S2-L13-RW-1*, 2011. doi:10.17226/22881