COMPARISON OF INJURY SEVERITY PREDICTION USING SELECTED VEHICLES FROM REAL -WORLD CRASH DATA

Susumu Ejima,
Tsukasa Goto
SUBARU Corporation
Japan

Peng Zhang,
Kristen Cunningham,
Stewart Wang
University of Michigan, International Center for Automotive Medicine
USA

Paper Number 23-0034

ABSTRACT

Advances in automotive telemetry technology have the potential to predict occupant severity from vehicle conditions at the time of an accident, and appropriate triage, as well as transport to a trauma center, can greatly improve subsequent treatment. The National Automotive Sampling System Crashworthiness Data System (NASS-CDS: 1999-2015) was used to filter for new case selection criteria based on vehicle type and matched to Subaru vehicle categories. We have proposed four types of injury severity prediction algorithms that were matched with the categories of Subaru vehicles. Specifically, 1) ISP model that categorized the principal direction of force (PDOF) into four impact directions (front, left, rear, and right) , 2) ISP-R model that considers the effect of the right-front passenger in addition to the four impact directions, 3) ISP-f1R model that represents PDOF as a continuous function using periodic basis splines, called functional data analysis, and 4) ISP-f2R model in which the knot position was modified in 3). In this study, five-fold cross-validation was performed within the training data (NASS-CDS 1999-2015) to evaluate the performance of these four models. In addition, external validation was performed using the National Automotive Sampling System Crash Investigation Sampling System (NASS-CISS: 2017-2019). The results of the cross-validation showed that the area under the receiver operating characteristic curve (AUC) was used to evaluate the model performance, which was 0.854 for the ISP model and 0.862 for the ISP-R model, indicating that the ISP-R, which considered the influence of the right-front passenger, was more accurate. The AUC values were 0.847 for the ISP-f1R model and 0.856 for the ISP-f2R model using a continuous function for the direction of impact, indicating that the ISP-R model had the highest AUC value among the models. On the other hand, the validation results with NASS-CISS were 0.817 for the ISP model and 0.828 for the ISP-R model, and 0.831 for the ISP-f1R model and 0.828 for the ISP-f2R model, indicating that all models had AUC values above 0.8. The important factors for the occupant injury prediction algorithm were delta-V, belt use, age, and crash direction, and the presence of a right-front occupant was a significant injury risk modifier, especially in side impact crashes.

INTRODUCTION

Advanced Automatic Collision Notification (AACN), which determines the severity of possible injuries to occupants in traffic accidents, was expected to not only shorten the time to initiate patient care, but also save patients' lives by selecting the hospital to which they will be transported based on appropriate triage [1]. According to a report [2] by the Centers for Disease Control and Prevention (CDC) Field Triage Task Force, the triage process consists of four stages (physiological, anatomic, mechanical, and special considerations). In the third, "mechanics," the effectiveness of predicting severity based on data transmitted from telematics has been demonstrated.

Research on injury severity prediction algorithms for automobile occupants [3-8] has been conducted in the United States since around 2000. As mentioned above, the background of this research was the appropriate triage decision during emergency medical care, which started with the determination of the injury level of the occupants based on the damage to the exterior and the interior of the vehicle during an accident. The existence of an underlying traffic accident database was essential to the development of the injury severity prediction algorithm, and the contribution of the extensive traffic accident database conducted in the United States [9-10] was extremely significant.

The CDC report [1] recommends that a patient with a predicted probability of serious injury of more than 20% was considered to be at high risk of serious injury and should be taken to an appropriate trauma center and treated accordingly. The severity used in injury prediction algorithms was mainly the Maximum Abbreviated Injury Scale (MAIS), which is the maximum value of the Abbreviated Injury Scale (AIS) [11], and the Injury Severity Score (ISS) [12], an index used to evaluate multiple trauma. Augenstein et al. [3] constructed the URGENCY algorithm using MAIS as the severity and several effective variables, which has been applied to BMW vehicles and verified using actual accidents [13]. Kononen et al.[5] developed a vehicle-level injury severity prediction model for each of the four impact directions using ISS, which was applied to the GM OnStar service. Subsequently, Wang et al. [14] improved the model and proposed a method to replace the crash direction with a continuous function using Principle Direction of Force (PDOF), which has been verified in actual accidents as an algorithm for each passenger seating position. Various injury prediction algorithms [15-17] have also been studied in Japan, and an injury prediction algorithm [18] has been integrated into the AACN called D-Call Net [19], which was now in practical use. The severity of injury was calculated based on the injuries (fatal, serious, minor, and non-injury) listed in the Japanese traffic accident database (Institute for Traffic Accident Research and Data Analysis: ITARDA) [20], and the injury probability of the occupant was calculated based on pseudo-delta-V, crash direction, belt use, age, and other factors.
In this study, the ISS, which has a high correlation with the fatality rate, was used as the severity of injury as an algorithm for predicting injuries to Subaru vehicles in the United States. Specifically, we focused on the vehicle category to which Subaru vehicles belong and conducted logistic regression analysis based on NASS-CDS [9] data to develop four ISPs models. The algorithms were based on NASS-CDS (1999-2015), and these algorithms were validated using the cross-validation with NASS-CDS and the latest dataset, NASS-CISS (2017-2019) [10], to clarify the characteristics of injury prediction by different models. We believe that the AACN provides a baseline for vehicle safety performance and post-accident safety (including emergency medical care) in actual accidents. By continuously improving this injury severity prediction algorithm, we can contribute toward "zero" traffic fatalities, one of the pillars of a safe transportation system.

METHODS

Data Source

Although there were various vehicle damage situations included in NASS-CDS, from a practical point of view, the variables used in the algorithm were based on the information transmitted by vehicle telematics, i.e., information obtained from the EDR (Event Data Recorder). As for the information about the occupants, the variables were age and gender, which may be obtained by the telematics service provider operators during an emergency call. The reason for setting these restrictions on variables was to reflect the current state of information available from current vehicle telematics systems.
For NASS-CDS (1999-2015) and model years 2000 or later were filtered using new case selection criteria based on vehicle type to match Subaru vehicle categories. Since limiting the sample to Subaru-make vehicles would result in a smaller sample size, we expanded our scope to investigate the vehicle belong to Subaru vehicle categories. The vehicle category was defined by limiting the study cohort to only the vehicle types recently produced by Subaru. We used the NHTSA standard definitions of passenger cars and sport utility vehicles (SUVs) as defined in the NASS-CDS coding [21]. The cohort was selected using the following specific vehicle codes: 02, 04, 05, and 06 are passenger cars and 14, and 15 are SUVs (Table 1).
An additional inclusion criterion for this study was that the principal direction of force (PDOF) was known. The PDOF was the direction of force input during a collision and was calculated from the longitudinal and lateral delta-V measured from the EDR. Then, the PDOF was used to classify four impact directions as shown in Figure 1(a): Frontal (11,12 or 1 o'clock), Left-side (2-4 o'clock), Rear (5-7 o'clock), and Right-side (8-10 o'clock). Furthermore, instead of classifying the PDOF into four impact directions, we used a statistical analysis called functional data analysis [14] where cyclic basis splines were used to model the PDOF curve as shown in the Figure 1(b). Vehicles that sustained a non-horizontal event, such as a rollover, were excluded. The number of collisions was classified into single or multiple.
The occupant and injury parameters were based on the following criteria. Belt status was available as “all front-occupants belted” vs. “at least one front-occupant unbelted”, since the injury severity prediction was created at the vehicle level. We did not consider the belt status of rear-seat passengers because current Event Data Recorders (EDRs) might not sense presence or rear-seat occupants. The age threshold was set at 55 years [14], and the presence of elderly passengers was defined as whether there was one passenger aged 55 years or older. The presence of female passengers was defined as if there was one female passenger. In addition, the presence or absence of a right-front passenger was taken into account in this analysis, so the front-right

passenger was also categorized as either "present" or "absent”.
ISS, defined as sum of the squares of the AIS of the three most significantly injured body regions, were available. The models were designed to predict ISS at the vehicle level. We chose highest ISS among the any of occupants. The total number of identified NASS cases that matched these conditions was 7,351. Table 2 lists the variables used in the analysis.

Table 1 Select body type from NASS-CDS

NASS-CDS Code	Body type	SUBARU Category
02	Car	2-door sedan, hardtop, coupe
04		4-door sedan, hardtop
05		5-door/4-door hatchback
06		Station Wagon (excluding van and truck-based)
14	Utility	Compact Utility (Utility Vehicle Categories "Small" and "Midsize")
15		Large utility

(a) Four impact directions ***(b) PDOF as continues curves***

Figure 1 Two different methodology using PDOF

Injury Prediction Model
In this study, four different algorithms were proposed. First, an algorithm that categorizes PDOF into four impact directions (front, right, rear, and left) according to Kononen et al.[5] and predicts whether someone in the vehicle has a high probability of being seriously injured using vehicle and occupant factor was developed. In addition, an algorithm that takes into account the influence of the right-front passenger was also developed. Next, following Wang et al.[14] statistical analysis, called functional data analysis, was employed to represent the PDOF effect as a continuous function, and periodic basis splines were used to model the PDOF curves. The key factor here was the number of basis splines (degrees of freedom). The greater the number of splines, the finer the PDOF curve can be obtained, which depends on the number of data. In defining the PDOF curve, two models with different knot positions were proposed. The knot position can "shift" the basis curve to focus on a specific region. Two methods were used here: the traditional method of choosing knots where there were many data points, i.e., according to the quantile of the data, and the forced method of setting knots because 0, 90, 180, and 270 degrees were known to be transition points. Note that the effect of the right-front passenger was taken into account in both models.

The four different algorithms were developed here using logistic regression models to predict the probability of sustaining ISS 15+ injuries.

1. ISP: Model without consideration of presence of right-front passenger.
2. ISP-R: Model with consideration of presence of right-front passenger.
3. ISP-f1R: Functional data analysis using PDOF and quantile-based knot selection using PDOF.
4. ISP-f2R: Functional data analysis using PDOF and knot selection based on impact direction.

Restricting the vehicle category from Subaru might also limit the usefulness of certain predictors. For example, the majority of the Subaru were passenger cars and small to midsize SUVs. Vehicle type was less heterogeneous and thus might or might not be strongly associated with injury risk. The stepwise procedure was used to select a subset of more relevant features to construct the final algorithm. The concept of using a statistical approach was to predict a 20% probability of ISS 15 or greater and the importance of each variable defined within a predictive model. Logistic regression models were fitted under the NASS sampling schema to investigate ISS 15 or greater injury with different configurations of the variables which was defined Table 2. In this study, the algorithm using the NASS-CDS, was developed as the basis to predict injury. The logistic regression equations are as follows:

$$P(ISS\ 15+) = \frac{1}{\left(1 + exp(-z)\right)} \qquad Equation\ (1)$$

$$z = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \qquad Equation\ (2)$$

where *P (ISS 15+)* is the probability that the ISS is 15 or greater (considered to be a serious injury level), β_0 was the intercept, β_i were the coefficients of the predictors x_i, and n is the number of predictors. These coefficients were shown in Table 2. It is necessary to have a cutoff value for *P (ISS 15+)*. The CDC Expert panel's recommendations for development of an injury risk algorithm specified a probability cutoff value of 0.2 using ISS. In another word, probability of 0.2 was defined as a serious injury at the crashes. In this study, this cutoff value was used with *P (ISS 15+)*.

Selection of variables for inclusion in the predictive models was done using forward and backward stepwise regression [22] to determine a final model. Starting from the null model, in each step, the criteria to add or remove one variable minimize Akaike information criteria (AIC) [23]. AIC was an information criterion that addresses the trade-off between the goodness of fit of the model and the complexity of the model. AIC was used in the stepwise variable selection in developing the predictive model. AIC includes a penalty that increases when the number of estimated parameters increases. Therefore, this penalty discourages over-fitting. The procedure stopped when AIC cannot be improved, and the final model was outputted. The performance of regression models was assessed with AIC and area under the curve (AUC) characteristic [24] which represents the model sensitivity as a function of specificity. The data analysis was conducted using R 4.1.1[25] with packages survey, pbs, and MASS.

Table 2 Predictor variable used in models that are able to be measured from vehicle telemetry or call center from the vehicles belong to Subaru category.

Variables	Categories
Delta V (mph)	Change in velocity (logarithm of Delta V)
Direction of impact	Four directions (Front/Left/Right/Rear) or PDOF (0 to 350 degree in 10-degree increment)
Vehicle Type	Car, Utility (MY 2000 and new car)
Number of Events	Single, Multiple
Belt Use	Belted, Unbelted
Right-front Passenger	Presence of right-front passenger
Age	At least one greater than 55 years
Gender	At least one female

Validation

Model validation was important to evaluate the expected prediction accuracy in the field. Five-fold cross validation [26] was performed within the training data (NASS-CDS 1999-2015) to evaluate model performance. The cross-validation delivered a merged version of all five validation data sets. By merging them, one summary was presented. Furthermore, we also externally validated the developed algorithms using the Crash Investigation Sampling System (CISS) cases. NASS-CISS is the improved NASS sampling scheme which better reflects the current population, newer vehicles, and enhanced injury coding. The number of cases by Subaru vehicle type identified from NASS-CISS was 1,163. NASS-CISS has a higher weighting than NASS-CDS, which means it represents more cases. Performance measures such as AUC were used to assess the accuracy of the developed predictive model.

RESULT

Statics Analysis

A statistical analysis was applied to the databases used (NASS-CDS, NASS-CISS) to quantify the contribution of each variable related to injury. Table 3 shows the comparison of each predictor between ISS 15+ vs. ISS 15- by each database. The numbers indicate the mean value of delta-V, and percentages for other categorical variables. In addition, t-test and Chi-squared test were used respectively for delta-V and other categorical variables for statistical comparison. The proportion of frontal impact was high as the direction of impact. The ISS 15+ in particular has a higher percentage of side impact from the left side. In terms of vehicle type, both categories have high percentages of Car. As for multiple collisions, ISS 15+ tended to have a higher rate, and the rate of seat belt use was higher for ISS less than 15. The proportion of right-front passenger was higher in ISS 15+, and the proportion of elderly occupants was higher in ISS 15+. Gender was not significantly different between the two groups. On the other hand, for NASS-CISS, a similar trend to that of NASS-CDS was observed, however, for each direction of impact (ISS 15+), the percentage of rear-end impact was higher than that of NASS-CDS, while the percentage of right-side impact was lower. Furthermore, the percentage of multiple collisions (ISS less than 15) was smaller than that of the NASS-CDS.

Table 3 Means and standard deviations of key variables stratified by ISS 15+ vs. ISS 15-. in NASS-CDS and NASS-CISS.

Variables		NASS-CDS(1999-2015) n=7,351			NASS-CISS(2016-2017) n=1,163		
		ISS 15+	ISS 15-	p-value	ISS 15+	ISS 15-	p-value
Delta V (mph)		29.0	20.4	<0.001	23.7	20.7	<0.001
Direction of impact	Front	55.3%	72.1%	<0.001	59.9%	75.9%	0.013
	Left	23.0%	7.1%		13.1%	5.5%	
	Right	17.8%	9.6%		8.6%	7.6%	
	Rear	3.9%	11.3%		18.4%	11.0%	
Vehicle type	Car	84.5%	79.7%	0.009	84.2%	81.3%	0.607
	Utility	15.5%	20.3%		15.8%	18.7%	
Number of Multiple Events		65.8%	54.1%	0.002	61.9%	29.1%	0.020
Seatbelt usage		63.3%	84.1%	<0.001	57.0%	83.9%	0.123
Presence of right-front passenger		38.5%	21.6%	<0.001	35.2%	18.5%	0.032
Presence of passengers older than 55 years		39.3%	17.5%	<0.001	27.4%	24.5%	0.069
Presence of female passengers		64.8%	60.6%	0.115	55.3%	58.5%	0.449

Logistic regression analysis was performed using the subject vehicles defined in the previous section to create ISPs and determine their contribution to each predictor variable. Table 4 shows the results of the coefficients of each variable in the logistic regression models (ISP, ISP-R, ISP-f1R, and ISP-f2R) using the stepwise variable selection method. p-values less than or equal to 0.05 were shown in bold. The most significant variable in the analysis was delta-V, which was common to all models. For each impact direction, ISP and ISP-R showed a relatively higher risk than the other impact directions from "left-side" impact. On the other hand, ISP-f1R and ISP-f2R, which approximate the PDOF effects with continuous functions, were represented by periodic curves (C1, C2, C3, C4) and have four degrees of freedom. The coefficients in the table approximate the effect of PDOF by combining the four curves and indicate the optimal degree of fitting. In addition, belt use was found to be a significant variable in all models. Age was found to be the most important variable among the factors related to the occupants. For ISP-R, in which the effect of the right-front passenger was taken into account, the variable for right-side impact (far side) was found to be significant. Although ISP-f1R and ISP-f2R were not significant, the selection of variables by AIC was an important factor to improve the accuracy of the model. On the other hand, only the ISP model selected gender differences as a predictor.

Table 4 Estimation of coefficients for variables from the ISPs (ISP, ISP-R, ISP-f1R, ISP-f2R).

Parameters		**ISP**	**ISP-R**			**ISP-f1R**	**ISP-f2R**
Intercept		-15.900	-16.559			-13.842	-12.754
ln Delta-V (mph)		**4.049**	**4.125**			**4.229**	**4.171**
Direction of impact	Front	0.481	0.470	PDOF	C1	-5.642	-6.076
	Left	**2.245**	2.586		C2	1.969	1.347
	Right	1.617	1.382		C3	**-2.989**	**-3.238**
	Rear	0.000	0.000		C4	**-2.588**	**-3.908**
Belt use	Belted	**-1.350**	**-1.250**			**-1.283**	**-1.256**
	Unbelted	0.000	0.000			0.000	0.000
Vehicle type	Utility	**-0.467**	-0.459			-0.499	-0.472
	Car	0.000	0.000			0.000	0.000
Number of events	Multiple	0.382	0.391			0.351	0.356
	Single	0.000	0.000			0.000	0.000
Presence of older occupants	55 or greater	**1.516**	**1.477**			**1.517**	**1.534**
	Under 55	0.000	0.000			0.000	0.000
Presence of right-front passenger	Front	-	0.847	PDOF	C1	2.222	1.026
	Left	-	-0.326		C2	-1.138	-0.220
	Right	-	**1.262**		C3	1.322	0.519
	Rear	-	0.648		C4	0.495	1.123
Presence of females	At least one female	-0.257	-			-	-
	No female	0.000	-			-	-

Algorithms were compared by their sensitivity and specificity the NASS-CDS with the five-fold cross-validation and the NASS-CISS. The sensitivity and specificity of the models when the cutoff values were varied from 0.10, 0.15, 0.20, 0.25, and 0.30 were shown in Table 5-6. The area under the receiver operating characteristic curve (AUC) results were also shown as a measure of the model's performance.

When the PDOF was categorized into four impact directions, the ISP-R model showed higher sensitivity than the ISP model by taking into account the influence of the right-front passenger, and there was no significant difference between the two models in terms of specificity. The models with PDOF represented as continuous functions (ISP-f1R, ISP-f2R) showed lower sensitivity values than ISP-R. No significant differences in specificity were observed for all models. The AUC, which evaluates model performance, was highest for the ISP-R model, followed by ISP-f2R, ISP-f1R, and ISP. On the other hand, validation using NASS-CISS showed that the ISP and ISP-R models, which were modeled with four impact directions, showed different trends depending on the value of the cutoff. Compared to the ISP model, the sensitivity of ISP-R varied depending on the cutoff value. The model using a continuous function for PDOF showed a similar trend, with ISP-f2R showing a slightly higher sensitivity than ISP-f1R. There was no significant difference in specificity among all models. The AUC was higher for ISP-f1R, followed by ISP-f2R, ISP-R, and ISP.

Table 5 Comparison of sensitivity and specificity with the five-fold cross-validation with the ISP, ISP-R, ISP-f1R and ISP-f2R.

NASS-CDS (Five-fold cross-validation)

cutoffs	**ISP**		**ISP-R**		**ISP-f1R**		**ISP-f2R**	
	sensitivity	**specificity**	**sensitivity**	**specificity**	**sensitivity**	**specificity**	**sensitivity**	**specificity**
0.10	0.580	0.903	0.586	0.900	0.581	0.906	0.579	0.899
0.15	0.493	0.946	0.518	0.943	0.461	0.945	0.474	0.946
0.20	0.430	0.965	0.451	0.966	0.389	0.968	0.406	0.967
0.25	0.370	0.976	0.401	0.978	0.349	0.979	0.360	0.977
0.30	0.324	0.986	0.353	0.983	0.316	0.984	0.320	0.983
AUC	0.854		0.862		0.847		0.856	

Table 6 Comparison of sensitivity and specificity with the NASS-CISS with the ISP, ISP-R, ISP-f1R and ISP-f2R.

NASS-CISS

cutoffs	ISP		ISP-R		ISP-f1R		ISP-f2R	
	sensitivity	specificity	sensitivity	specificity	sensitivity	specificity	sensitivity	specificity
0.10	0.658	0.896	0.652	0.909	0.652	0.899	0.653	0.894
0.15	0.533	0.928	0.444	0.929	0.465	0.931	0.512	0.934
0.20	0.389	0.953	0.376	0.954	0.385	0.944	0.414	0.953
0.25	0.289	0.963	0.333	0.969	0.342	0.964	0.347	0.967
0.30	0.230	0.969	0.240	0.971	0.282	0.969	0.284	0.972
AUC	0.817		0.828		0.831		0.828	

DISCUSSION

In this study, four different algorithms were developed using the NASS-CDS (1999-2015) to predict injuries focused on vehicles belonging to the Subaru category. The results showed that all algorithms showed higher risk curves for front, right, rear, and left impacts than Kononen et al.[5]. The main reason was the effect of limiting the vehicle body types in the NASS-CDS coding to 2, 4 to 6 for passenger cars and 14 and 15 for utility vehicles among the Subaru vehicle types. Subaru vehicles were relatively small or medium-sized, and the difference in vehicle body weight may have resulted in the higher risk curves.

The contribution to each predictor was determined from the results of the logistics regression model using the stepwise variable selection method. Delta-V, belt use, and age were common to all models as factors with high contribution, and in particular, delta-V was found to be the factor with the highest contribution rate. In the models (ISP and ISP-R) for the four impact directions, the contribution from the left-side impact was large, which was similar to other literature [27-28]. For the models that represent PDOF as a continuous function (ISP-f1R, ISP-f2R), the maximum number of degrees of freedom was four (C1, C2, C3, C4), even when the number of data was over 7,000 cases. In the end, a primary effect with PDOF as a continuous function and a secondary effect due to the presence of the right-front passenger were modeled with four degrees of freedom. Wang et al. [15] used 10 degrees of freedom for the PDOF periodic function and five degrees of freedom for the effect of the right-front passenger. For the four impact directions (front, right, rear, and left) model, the degree of freedom was three, indicating that the advantages of the functional data analysis were not fully demonstrated compared to the impact direction model.

The effect of the secondary effect was determined for the case with a right-front passenger in each crash direction (ISP-R) and for the models with continuous PDOF functions (ISP-f1R and ISP-f2R). The results for each impact direction (ISP-R) showed that the secondary effect for the case with a right-front passenger was larger in the case of right-side impact. Similarly, for the models representing PDOF as a continuous function (ISP-f1R, ISP-f2R), the secondary effects were observed, although the effects were not significant due to the limited number of samples.

Newland et al.[29] noted that drivers with a front seat passenger present were more at risk than drivers without a passenger as a risk factor for drivers. Thus, passenger interaction, which represents contact with adjacent occupants in the same seating row, was found to be more severe when a right-front passenger was present. The effect of gender difference was only chosen for the ISP in the stepwise variable selection method.

According to the AUC results using the NASS-CDS dataset, ISP-R, which categorized PDOF into four impact directions, was higher than the models that represented PDOF as a continuous function (ISP-f1R and ISP-f2R). The reason for this may be that, as stated previously, the functional data analysis was limited to Subaru vehicle categories from NASS-CDS, which narrowed down the number of cases, and therefore could not demonstrate sufficient performance. Wang et al.[15] reported that the AUC value was improved by using a continuous function for PDOF. In order to improve the performance of the algorithm, it is necessary to collect data on a larger number of accidents, and data consistency is important. There was no significant difference in the specificity of the models, while ISP-R performed better than the other models in terms of sensitivity.

On the other hand, the AUC results using the NASS-CISS dataset showed that the model with PDOF as a continuous function produced better results. For sensitivity, ISP-f2R showed the highest results when the cutoff value was 0.2 (the value recommended by the CDC). It is important to set an appropriate cutoff value for the NASS-CISS because it includes relatively new vehicles and the safety performance of vehicles in a crash has been improved. Sensitivity and specificity were in a trade-off relationship, and it is important to balance sensitivity and specificity in setting cutoff values, and it is necessary to discuss appropriate triage decisions based on field data. In addition, NASS-CISS has larger weights than NASS-CDS and represents more cases. The median weight of the NASS-CDS was 75. In contrast, the median weight of NASS-CISS was 193. The larger weight for NASS-CISS was likely due to the greater number of crashes per year compared to when NASS-CDS was implemented. The new weighting of the NASS-CISS addresses this issue and thus has a larger weight. A larger sample was needed to evaluate the safety performance of newer vehicles of a given year, which also plays a role in improving the performance of the algorithm.

CONCLUSOIN

In this study, the Subaru vehicle category was selected as the vehicle body type from the NASS-CDS, and four different algorithms were used to predict the severity of injuries. All algorithms were able to predict the risk of serious injury (ISS 15 or higher), confirming that the injury prediction method was useful for on-site triage of occupants in the event of an accident. The area under the receiver operating characteristic curve (AUC) was used as a measure of model performance, and cross-validation using NASS-CDS showed that the ISP-R model had the highest result at 0.862. Additional validation using NASS-CISS resulted in AUC values above 0.8 for all models. The major predictors were delta-V, direction of impact, seatbelt use, and age, and the presence of a right-front passenger was found to be an important injury risk modifier, especially for side impact crashes. These models need to be continually improved with the collection of crash data.

The model that most predicted injury outcomes was at the occupant level, not at the vehicle level. However, in practically crash scenario, a vehicle's telematic system would make the initial contact and Emergency Medical Services (EMS) would respond to the vehicle as a unit. Therefore, this paper proposes an injury severity prediction model that aims to predict crash vehicle categories that may include serious injuries at the vehicle level. Many automakers are attempting to collect rear seat occupant information as well as front seat occupant information. Thus, by understanding the status of occupant information collection, the injury severity prediction model can be extended to the occupant level.

Reference

[1] MacKenzie, EJ., Rivara, FP., Jurkovich, GJ., et al. A national evaluation of the effect of trauma-center care on mortality. The New England Journal of Medicine, 354:366–378. 2006.

[2] Sasser, SM., Hunt, RC., Sullivent, EE., et al. Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage., MMWR Morbidity and mortality weekly report, 58:1–35. 2009.

[3] Augenstein, J., Digges, K., Ogata, S., Perdeck, E., and Stratton, J., Development and validation of the URGENCY algorithm to predict compelling Injuries. Proceeding of the 17th ESV Conference 2001, Paper No. 352

[4] Bahouth, GT., Digges, KH., Bedewi, NE,. et al. Development of URGENCY2.1 for the prediction of crash injury severity. Top Emergency Medicine. 26:157–165. 2004.

[5] Kononen, DW., Flannagan, CA., Wang, SC., Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accident Analysis and Prevention, 43:112–122. 2011

[6] Ayoung-Chee, P,, Mack, CD., Kaufman, R., Bulger, E., Predicting severe injury using vehicle telemetry data. Journal Trauma Acute Care Surgery. 74, 190–194; discussion 4–5. 2013.

[7] Kusano, KD., Gabler, HC., Comparison and Validation of Injury Risk Classifiers for Advanced Automated Crash Notification Systems. Traffic Injury Prevention, 15, S126–S133, 2014.

[8] McMurry, TL., Cormier, JM., Daniel, T., et al. An omni-directional model of injury risk in planar crashes with application for autonomous vehicles. Traffic Injury Prevention, DOI: 10.1080/15389588.2021.1955108, 2021.

[9] National Highway Traffic Safety Administration. National Automotive Sampling System. Washington, DC: National Highway Traffic Safety Administration; Available at: https://www.nhtsa.gov/crash-data-systems/national-automotive-sampling-system. Accessed May 15, 2020.

[10] National Highway Traffic Safety Administration. National Automotive Sampling System. Washington, DC: National Highway Traffic Safety Administration; Available at: https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/CISS/ Accessed May 15, 2020.

[11] Association for the Advancement of Automotive Medicine. "Abbreviated Injury Scale (AIS)": Available at: https://www.aaam.org /abbreviated-injury-scale-ais/. Accessed May 1, 2020.

[12] Baker, SP., O'Neill, B., Haddon, WJr., Long, WB., The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma.*14:187–196. 1974.

[13] Rauscher, S., Messner, G., Baur, P. Augenstein, J., Digges, K., Perdeck, E., Bahouth, G., and Pieske, O., Enhanced automatic collision notification system -Improved rescue due to injury prediction- First filed experience. Proceeding of the 21th ESV Conference 2009, Paper Number. 09-0049

[14] Wang, SC., Kohoyda-Inglis, CJ. et al., Second Generation Aacn Injury Severity Prediction Algorithm: Development and Real-World Validation. Proceeding of the 25th ESV conference, Paper Number: 17-0133-O, 2017.

[15] Kuniyuki, H., A Study of Injury Prediction Method and Influential Factors in Frontal Collision Using Accident Data, JSAE Transaction Vol.43. No. 2. p.261-267. 2012
[16] Tominaga, S., Nishimoto, T., Motomura, T., Matsumoto, M, Lubbe, N., and Kiuchi, T., Injury prediction algorithm for AACN based Japanese road accident data, JSAE Transaction Vol.46 p.925-930. 201
[17] Kubota, K., Nishimoto, T., Tominaga, S., and Miyoshi, T., Development of Injury Prediction Algorithms Based on Seating Positions, Impact Directions and Injury Regions. JSAE Spring Annual Conference. Paper Number 20175518. 2017
[18] Nishimoto, T., Mukaigawa, K., Tominaga, S., Motomura, T., Lubbe, N., Kiuchi, T., Matsumoto, M, and Matsumoto, H., Serious Injury Prediction Algorithm based on large-scale data and under-triage control. Accident Analysis and Prevention, Vol 98, p.266-276. 2017
[19] D-Call Net https://www.helpnet.co.jp/. Accessed Nov 24, 2022.
[20] Institute for Traffic Accident Research and Data Analysis (ITARDA). https://www.itarda.or.jp/english/ Accessed Nov 24, 2022.
[21] National Highway Traffic Safety Administration, National Automotive Sampling System (NASS) Crashworthiness Data System Analytical User’s Manual 2007 File. National Center for Statistics and Analysis, NHTSA, U.S. Department of Transportation, Washington, DC.
[22] Hocking RR,: The Analysis and Selection of Variables in Linear Regression. Biometrics, 32, (1976).
[23] Akaike H,: A new look at the statistical model identification, IEEE Transactions on Automatic Control,19(6):716–723, doi:10.1109/TAC.1974.1100705, MR 0423716, (1974).
[24] Fawcett T,: An Introduction to ROC Analysis. Pattern Recognition Letters. 27 (8): 861–874. doi:10.1016/j.patrec.2005.10.010, (2006).
[25] R software, https://www.r-project.org/. Accessed May 2020.
[26] Stone M,: Cross-validatory choice and assessment of statistical predictions. J. Royal Stat. Soc.,36(2),111 147, (1974).
[27] Farmer, CM,. Braver, ER., Mitter, EL., Two-vehicle side impact crashes: The relationship of vehicle and crash characteristics to injury severity. Accident Analysis and Prevention, 29, Issue 3, Pages 399-406, 1997.
[28] Stigson, H., and Kullgren, A,. Effect of Side Impact Protection in Reducing Injuries. 22nd International Technical Conference on the Enhanced Safety of Vehicles (ESV), Washington, D.C., June, Paper Number 11-0319, 2011
[29] Newland, C., Belcher, T., Bostrom, O. et al., , Occupant-to-occupant interaction and impact injury risk in side impact crashes. *Stapp Car Crash J,* 52**,** 327-47. 2008.