On Scene Injury Severity Prediction (OSISP) machine learning algorithms for motor vehicle crash occupants in US

(article starts on next page)

Contents lists available at ScienceDirect

Journal of Transport & Health

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

On Scene Injury Severity Prediction (OSISP) machine learning algorithms for motor vehicle crash occupants in US

Stefan Candefjord [a,b,c,*], Azam Sheikh Muhammad [d], Pramod Bangalore [d], Ruben Buendia [a,b,c,e]

[a] Department of Electrical Engineering, Chalmers University of Technology, 412 96, Gothenburg, Sweden
[b] SAFER Vehicle and Traffic Safety Centre at Chalmers, Sweden
[c] MedTech West, Sahlgrenska University Hospital, Röda Stråket 10 B, 413 45, Gothenburg, Sweden
[d] Department of Computer Science and Engineering, Chalmers University of Technology, 412 96, Gothenburg, Sweden
[e] Department of Information Technology, University of Borås, 501 90, Borås, Sweden

ABSTRACT

A significant proportion of motor vehicle crash fatalities are potentially preventable with improved acute care. By increasing the accuracy of triage more victims could be transported directly to the best suited care facility and be provided optimal care. We hypothesize that On Scene Injury Severity Prediction (OSISP) algorithms, developed utilizing machine learning methods, have potential to improve triage by complementing the field triage protocol. In this study, the accuracy of OSISP algorithms based on the "National Automotive Sampling System - Crashworthiness Data System" (NASS-CDS) of crashes involving adult occupants for calendar years 2010–2015 was evaluated. Severe injury was the dependent variable, defined as Injury Severity Score (ISS) > 15. The dataset contained 37873 subjects, whereof 21589 included injury data and were further analyzed. Selection of model predictors was based on potential for injury severity prediction and perceived feasibility of assessment by first responders. We excluded vehicle telemetry data due to the limited availability of these systems in the contemporary vehicle fleet, and because this data is not yet being utilized in prehospital care. The machine learning algorithms Logistic Regression, Ridge Regression, Bernoulli Naïve Bayes, Stochastic Gradient Descent and Artificial Neural Networks were evaluated. Best performance with small margin was achieved with Logistic Regression, achieving area under the receiver operator characteristic curve (AUC) of 0.86 (95% confidence interval 0.82–0.90), as estimated by 10-fold stratified cross-validation. Ejection, Entrapment, Belt use, Airbag deployment and Crash type were good predictors. Using only a subset of the 5–7 best predictors approached the prediction accuracy achieved when using the full set (14 predictors). A simplified benefit analysis indicated that nationwide implementation of OSISP in the US could bring improved care for 3100 severely injured patients, and reduce unnecessary use of trauma center resources for 94000 non-severely injured patients, every year.

* Corresponding author. Electrical Engineering, Chalmers University of Technology, 412 96, Gothenburg, Sweden.
E-mail address: stefan.candefjord@chalmers.se (S. Candefjord).

1. Introduction

Motor vehicle crashes (MVC) in US produce around 30000 fatalities and 4 million injured people every year, adding up to a societal economic burden of totally $240 billion or $800 per citizen (Blincoe et al., 2015). A significant proportion of the fatalities are potentially preventable (Ray et al., 2016; Berwick et al., 2016). Military trauma care has achieved remarkably high survival rates, 98%, for patients reaching a treatment facility (Berwick et al., 2016). If similar outcomes can be achieved in civilian care, up to 20% of trauma fatalities can be saved (Berwick et al., 2016). For a local sample in Miami–Dade ($n = 98$) it was concluded that over a third of MVC deaths were potentially preventable (Ray et al., 2016). Patients with severe injury have a higher probability of surviving if they are directly transported to a trauma center, to provide specialized care with minimal delay (Hu et al., 2017; Haas et al., 2010; MacKenzie et al., 2006; Candefjord et al., 2020). A key to provide adequate care to a larger proportion of patients is to attain a high triage accuracy (Ray et al., 2016; Sasser et al., 2012), so that the rate of appropriate decisions on where to transport the patient can be increased.

In this study, we evaluate if methods employing machine learning and variables that can be assessed on the scene of accident has potential to amend field triage. First, we provide a literature review including a description of the field triage process and the challenges of attaining a high triage accuracy. It also identifies previous studies that form the foundation for the present study. At the end of the introduction, we provide the aim of the current study.

1.1. Literature review and study motivation

The triage protocol is the most important decision support for identifying patients with severe injury, while using health care resources efficiently by recognizing patients not likely to be in need of specialized care. The current US guidelines for field triage of injured patients are based on four steps: 1) Vital signs, i.e. Glasgow Coma Scale, systolic blood pressure, and respiratory rate; 2) Anatomy of injury, e.g. penetrating injuries and flail chest; 3) Mechanisms of injury, e.g. falls from >20 feet and high-risk auto crash; and 4) Special considerations, e.g. older adults (aged > 55 years) (Sasser et al., 2012). The steps are assessed in sequential order, i.e. if any of the step 1 criteria indicate severe injury the decision scheme is completed with the recommendation to transport the patient to a trauma center, followed by steps 2–4 if no previous step signals severe injury. If any criteria in steps 1–3 is fulfilled the protocol recommends transport to a trauma center, and if step 4 is fulfilled transport to a trauma center should be considered. By employing four consecutive steps based on different risk factors the rate of undertriage can be decreased.

Even though sophisticated guidelines for field triage have been developed, the rate of undertriage of US trauma patients in general, and MVC occupants in particular, is high. The American College of Surgeon's Committee on Trauma (ACS-COT) (American College of Surgeons Committee on Trauma (ACS-COT), 2014) states that *"an acceptable undertriage rate could be as high as 5%"*, when undertriaged patients are defined as having an Injury Severity Score (ISS) of 16 or more who were taken to a non-trauma center. Furthermore, ACS-COT states that *"most agree that an acceptable percentage of overtriage is in the range of 25% to 35%"*. Xiang and colleagues (Xiang et al., 2014) showed that more than one third (34%) of patients with major trauma in the US emergency departments were undertriaged in 2010. Stitzel et al. (2016) showed, in a US population weighted sample of MVC based on the National Automotive Sampling System - Crashworthiness Data System (NASS-CDS) for years 2000–2011 ($n = 9, 763, 984$), that the rate of undertriage was 20% with an overtriage of 54%. Note that using machine learning terminology, low undertriage corresponds to high sensitivity/true positive rate (low number of false negatives), and low overtriage corresponds to high specificity (low false positive rate).

The criteria for high-risk auto crash in the field triage guidelines (Sasser et al., 2012) are: i) compartment intrusion measures; ii) passenger ejected from vehicle (partial or complete); iii) death in same passenger compartment; iv) vehicle telemetry data consistent with high risk of injury. Criteria iv is based on the development of Advanced Automatic Crash Notification (AACN) algorithms, which use vehicle sensor data such as ΔV (total change of velocity during crash), principal direction of impact force, belt status and airbag deployment to predict the probability of any MVC occupant being severely injured. AACN shows promise to improve triage for MVC (Augenstein et al., 2003; Champion et al., 2005; Kononen et al., 2011; Stitzel et al., 2016). The use of AACN in field triage protocols was acknowledged by the "Expert Panel of the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention" in 2008 (National Center for Injury Prevention and Control, 2008). The panel recommended that for an estimated risk ≥20% of having a severe injury (defined as ISS > 15), the AACN provider should inform the Public Safety Answering Point (PSAP) that the occupant is at risk for a severe injury.

Candefjord, Buendia and colleagues (Candefjord et al., 2015; Buendia et al., 2015) showed that so called On Scene Injury Severity Prediction (OSISP) algorithms, which are based on only crash characteristics that are feasible to assess on the scene of crash by first responders, achieved high accuracy for prediction of severe injury for MVC occupants in passenger cars (Buendia et al., 2015) and trucks (Candefjord et al., 2015). The OSISP concept has many similarities with AACN. The fundamental difference is that OSISP is designed to be implemented in a handheld device such as a tablet and information about the crash should be interpreted and input to the device by first responders (Olaetxea Azkarate-Askatsua, 2017), whereas AACN is integrated into the vehicle and retrieves data about the crash from the so called Event Data Recorder (Kononen et al., 2011). The AACN prediction can be executed directly following a crash, whereas the OSISP prediction is available after on scene assessment. AACN can utilize precise measurements of ΔV and principal direction of force (Kononen et al., 2011). OSISP, on the other hand, can employ some variables that can be assessed on scene but are typically not detected by vehicles, such as the sex and estimated age of the patient and whether the patient was ejected from or entrapped in the vehicle. The AACN and OSISP concepts are complementary, AACN can perform injury severity prediction at dispatch to aid planning the rescue and care operation and call adequate personnel and resources to the scene, while OSISP can be used on scene and incorporate data from on scene assessments, e.g. observations of the vehicle and patient.

Table 1

Definitions, descriptions and statistics for the model variables. For each variable level the proportion (percentage) of the MVC occupants in the full dataset having that characteristic is shown (denoted PF = proportion full), together with the proportion that sustained severe injury (denoted PSI = proportion severe injury). MD = missing data.

Variable	Definition	Level (PF/PSI [%])	Description
ISS	Injury Severity Score	ISS < 15 (94.3/ 0.0)	Whether the occupant was severely injured or not (dependent variable)
		ISS > 15 (5.7/ 100.0)	
Sex	Sex of occupant	Male (50.4/6.7)	Male or female occupant
		Female (48.6/4.9)	
		MD (1.0/0.0)	
Elderly	Victim over or under 55 years old	≤55 (77.3/5.5)	Age 55 years old is used as threshold, following the triage protocol (Sasser et al., 2012)
		>55 (20.7/7.1)	
		MD (2.0/0.2)	
PSL	Posted speed limit (km/ h)	$15 \leq x < 25$ (0.4/ 4.1)	Maximum speed allowed where the crash occurred. The posted speed limits in mph were converted into km/h and grouped in intervals, to facilitate comparisons with international studies (Buendia et al., 2015; Candefjord et al., 2015).
		$25 \leq x < 35$ (0.3/ 4.3)	
		$35 \leq x < 45$ (10.6/ 3.1)	
		$45 \leq x < 55$ (12.8/ 4.4)	
		$55 \leq x < 65$ (30.8/ 5.2)	
		$65 \leq x < 75$ (16.5/ 5.7)	
		$75 \leq x < 85$ (4.6/ 6.0)	
		$85 \leq x < 95$ (10.9/ 8.1)	
		$95 \leq x < 105$ (3.1/ 11.7)	
		$105 \leq x < 115$ (7.8/8.1)	
		$115 \leq x < 125$ (0.9/15.5)	
		MD (1.4/2.3)	
Belt Use	Belt use	Belted (83.4/3.8)	Whether the occupant wore seat belt
		Unbelted (7.2/ 26.0)	
		Unknown (3.0/ 2.5)	
		MD (6.4/9.9)	
Vehicle	Type of vehicle	Car (60.8/6.3)	Whether the occupant was in a car or a truck
		Truck (39.2/4.9)	
Airbag	Airbag deployment	Not deployed (41.7/3.3)	Whether any airbag (if present) was deployed
		Deployed (38.0/ 8.1)	
		Not present (7.1/ 7.6)	
		MD (13.2/5.8)	
Location	Location of crash	Urban (67.0/6.0)	Whether crash occurred in urban or rural environment. Cities and metropolitan areas are categorized as urban, other areas as rural.
		Rural (33.0/5.3)	
Position	Seat position	Front (93.2/5.7)	Seat position of the injured occupant. Divided into front or back seat.
		Back (6.4/7.3)	
		MD (0.4/2.5)	
WN	Weekend night	No (90.1/5.2)	Whether crash occurred during a weekend night, defined as either Saturday or Sunday at 0:00–7:59 a.m.
		Yes (9.1/10.9)	
		MD (0.8/5.5)	
Crash	Type of crash	Single (24.9/12.0)	Single vehicle crashes into stationary object or departs from the road
		Rear End (20.3/ 2.5)	Vehicle impacts another vehicle from behind
		Head On (4.4/ 15.3)	Head on collision between two vehicles
		Other (48.2/3.1)	Other type of crash
		MD (2.1/4.8)	
Eject	Ejection	No (97.7/4.8)	Whether occupant was ejected from the vehicle. The category 'Yes' includes both complete and partial ejection.
		Yes (1.7/58.4)	
		MD (0.6/12.0)	

Table 1 (*continued*)

Variable	Definition	Level (PF/PSI [%])	Description
Entrap	Entrapment	No (89.8/4.5)	Whether occupant was entrapped in the vehicle
		Yes (2.0/46.3)	
		MD (8.2/9.7)	
Rollover	Rollover	No (90.5/4.8)	Whether the vehicle rolled over, defined as $\geq 1/2$ turns
		Yes (8.6/14.2)	
		MD (0.9/26.1)	
Light	Light conditions	Daylight (61.2/4.1)	Light conditions at the scene of crash
		Dark (9.2/11.0)	
		Dark/Lighted (25.4/7.8)	
		Dawn (1.5/9.6)	
		Dusk (2.2/4.1)	
		MD (0.4/4.3)	

An advantage with OSISP compared to AACN is that applicability is not limited by the penetration rate of AACN systems in the contemporary vehicle fleet. This rate is limited by at least three important factors. First, The National Highway Traffic Safety Administration (NHTSA) predicted that only approximately 20% of the model year 2016 vehicle fleet are equipped with AACN (Lee et al., 2017). Second, functional AACN systems commonly require an active subscription based service, which may further reduce the proportion of vehicles for which AACN can be used (there exists no public data on rate of active subscriptions) (Lee et al., 2017). Third, to our knowledge there is no established standard for how AACN results are derived or presented to the PSAP, and few, if any, PSAPs/prehospital care systems have implemented routines for utilizing AACN data. OSISP can be used independently of vehicle telemetry data, and has potential to be used for most MVC. For most effective rescue and care and the widest coverage of MVC patients, AACN and OSISP could therefore be used in combination.

1.2. Aim

Current OSISP algorithms have been developed from the Swedish Traffic Accident Data Acquisition (STRADA) database using the method Logistic Regression (Candefjord et al., 2015; Buendia et al., 2015). The aim of this study is to develop an OSISP algorithm for MVC occupants in US based on NASS-CDS data, using several machine learning algorithms and compare their performance. A high-performing OSISP algorithm may be used to refine US field triage protocols, to improve the care for severely injured patients while decreasing unnecessary use of trauma center resources.

2. Materials and methods

2.1. Data selection

The scope of this study was adult MVC occupants registered in the NASS-CDS database for calendar years 2010–2015 (Radja, 2016). NASS-CDS includes investigations of around 5000 crashes per year involving passenger cars, light trucks, vans, and utility vehicles. The rationale for selecting calendar years 2010–2015 was that in 2010 injury scores were updated according to the AIS 2005 standard, and 2015 was the last full year available at the time the study commenced.

From 2010–2015, 45075 MVC occupants sustaining injury were identified in NASS-CDS (four data sets were linked, i.e. *Accident, Event, GV* and *OA*, described in Radja (2016)). Out of these, 43712 were car or light truck occupants, whereof 37873 occupants were ≥18 years old. Furthermore, 16284 occupants had no ISS data, leaving 21589 cases that were further analyzed.

2.2. Model variables

To classify occupants as being severely injured or not the ISS was used (Baker and O'Neill, 1976). ISS builds on classification of severity of each injury according to the Abbreviated Injury Scale (AIS) (Association for the Advancement of Automotive Medicine (AAAM), 2005). The threshold used to define severe injury was ISS >15, which is commonly recommended (Sasser et al., 2012).

The dependent variable was whether the occupant sustained severe injury or not. The predictor variables were chosen based on experience about potential for injury severity prediction, gained from several literature sources (Buendia et al., 2015; Candefjord et al., 2015; Stitzel et al., 2016; Kononen et al., 2011; Augenstein et al., 2003). Furthermore, a requirement for all selected predictors were that they were deemed to be feasible to assess at the scene of crash by first responders. We excluded vehicle telemetry data due to the limited availability in the contemporary vehicle fleet. All variables included in the model are detailed in Table 1.

2.3. Data representation and handling missing data

All of the variables in this study were categorical (Table 1). In order to achieve best performance for a machine learning task, the right choice of data representation technique is vital, especially for categorical data. Some machine learning algorithms, such as

Support Vector Machines and Multi-layer Perceptron (Deep Learning), explicitly require all the input variables to be numerical (Hastie et al., 2009). There are plenty of techniques to transform categorical values to numerical data enabling one to use any algorithm for all types of data - numeric or categorical. Two widely used techniques are *label encoding* and *one hot encoding*, described in Brink et al. (2017, pp. 36–43). We found that one hot encoding consistently yielded improved performance over label encoding, and decided to use it for this study.

Except the variables *Vehicle* and *Location*, all other variables exhibited missing values to varying extent (Table 1). We applied four different methods to handle missing data (Brink et al., 2017, pp. 36–43), creating one separate instance of the dataset per method. The methods were: remove cases with missing values (discarding 5394 cases that contained missing data), turn missing values into a new category level, imputing missing values using mode (most frequent value), and imputing missing values using conditional probability. While the first three methods are self explanatory, it is important to describe some details of the imputation using conditional probability. Assuming independence of the variables, the inherent information about probabilistic relationships between the predictor and target variables can be determined from the available complete data set (without missing values). Subsequently, these conditional probabilities can be utilized to impute the data for variables with missing values. Imputation via conditional probability will ensure that the missing value for a variable is filled with a value having the highest probability with respect to the target case, i.e., filling in a missing value with the most likely one determined by the similar cases but without missing value for that particular variable. Unlike the imputation using mode, imputation with conditional probabilities has the advantage of respecting the distribution of the variable and therefore should not introduce unexpected patterns in our data.

2.4. Machine learning algorithms

In this study we view machine learning as a broad concept, including traditional mathematical/statistical models that can be used for binary classification. Logistic Regression is a method that is commonly used in studies of MVC, see e.g. (Harrell, 2001; Schiff et al., 2008; Augenstein et al., 2003; Kononen et al., 2011; Candefjord et al., 2015; Buendia et al., 2015). We performed a literature study of similar problem domains and identified some natural competitors to Logistic Regression. We also diversified our search and tried several linear and non-linear machine learning methods. In an initial round of tests, we included Decision Trees, Random Forest, Linear Discriminant Analysis and Support Vector Machines (Hastie et al., 2009). The initial tests demonstrated subpar performance of these algorithms, therefore they were excluded from this study. We finally concluded with a set of four different algorithms that were deemed to have high potential: Ridge Regression, Stochastic Gradient Descent, Bernoulli Naïve Bayes, and Artificial Neural Networks. Every model was compared to the established Logistic Regression algorithm with respect to the performance metrics defined in Section 2.5.

Python (version 3.5) was used as base for all data analysis, with the data science libraries Pandas (version 0.22.0) and Scikit-learn (version 0.19.0) (Pedregosa et al., 2011), using the default settings for classifiers. For Artificial Neural Networks, TensorFlow (version 1.0.0) (Abadi et al., 2016) was used. Output results were adapted so that all classifiers could be easily compared.

2.4.1. Logistic regression

Studies of MVC have commonly focused on risk factors and the potential of individual predictor variables for injury severity prediction (Schiff et al., 2008; Kononen et al., 2011; Candefjord et al., 2015; Buendia et al., 2015). This objective is well met by Logistic Regression, which describes the probabilistic relationships between the dependent variable and the individual predictor variables. Let P denote the probability of severe injury. Y is the dependent variable. We choose $Y = 0$ for non-severe injury and $Y = 1$ for severe injury. The model can be defined using the form in Equation (1),

$$\ln\left(\frac{P(Y=1)}{1 - P(Y=1)}\right) = \beta_0 + \sum_{k=1}^{n} \beta_k X_k \tag{1}$$

where the left hand side of Equation (1) is referred to as the *logit transformation*. The fraction $\frac{P(Y=1)}{1-P(Y=1)}$ is called the odds ratio (OR) for the event $Y = 1$. The logit transformation enables estimating the logged odds of the event $Y = 1$. The OR for the kth predictor variable X_k is given by the coefficient e^{β_k} and represents the *constant effect* of the predictor X_k on the likelihood that $Y = 1$ will occur. This is exactly what we aim to measure in MVC analysis, i.e. quantifying the isolated effect of each X on Y with a single metric. Another advantage is that Logistic Regression does not require a linear relationship between the dependent variable and the predictors, since it applies a non-linear log transformation to the predicted OR (it does not mean that the model is non-linear, the transformation is applied on the output).

Results of Logistic Regression modeling were expressed as adjusted OR, with corresponding 95% confidence intervals (CI) and levels of statistical significance (p-values). The overall low proportion of severe injury in this study ($<6\%$) suggested that OR is a reasonable approximation of the relative risk. To report the required results, we implemented a wrapper class that encapsulated the default Logistic Regression from Scikit-learn and added the functionality to compute the aforementioned quantities.

2.4.2. Ridge regression

The performance of Logistic Regression depends on two key assumptions. First, there are no outliers (misclassified instances) in the dataset. Second, there does not exist any high correlations among the predictor variables (multicollinearity). A closer analysis of the dataset in the present study revealed that there exist cases with the same set of predictor variable values having different dependent

variable outcome. This is expected in an MVC dataset, because similar conditions does not necessarily end with similar injury outcome. However, these outliers can unduly influence the results of the analysis and lead to incorrect inferences.

Ridge Regression uses a regularization approach that constrains/regularizes or shrinks the coefficients of the ordinary least squares regression model. It discourages learning a complex model by allowing misclassification of extreme outliers and thereby improves generalizability. Ridge Regression also solves the multicollinearity problem through a shrinkage parameter controlling penalty on model coefficients. The learner identifies coefficients that are close to zero, and does not aim to fit every training data point. The coefficient estimates produced by this method are also known as the L_2 norm. For detailed mathematical formulation and practical implementation of Ridge Regression please refer to Hastie et al. (2009, pp. 59–65), Müller and Guido (2016, pp. 51–57) and Tattar (2017, pp. 312–318).

2.4.3. Stochastic gradient descent

Maximum likelihood estimation is used by several machine learning methods, including Logistic Regression, to estimate the model coefficients (β_k) (Equation (1)). A minimization algorithm such as Gradient Descent (GD) optimization is usually employed for this purpose. An alternative approach is to replace GD with its counterpart called Stochastic Gradient Descent (SGD). In GD optimization the cost gradient is computed from the complete training set, which can become time consuming for large datasets. In SGD, the gradient is computed for one or a batch of training data points at a time until it converges. The term "stochastic" comes from the fact that the gradient based on randomly selected training samples is a "stochastic approximation" of the "true" cost gradient.

Besides being faster than GD, SGD is superior for datasets with redundant samples (all variable levels equal), which applied to the dataset in the present study. This observation formed the rationale to include a variant of Logistic Regression with SGD based training. We used a regularization approach called "Elasticnet", which is a convex combination of the L_1 and L_2 norms, which is available in Scikit-learn and well explained in the documentation.

2.4.4. Bernoulli Naïve Bayes

Usually, Naïve Bayes based algorithms are designed for text classification (McCallum and Nigam, 1998). However, Bernoulli Naïve Bayes is a variant where each feature is assumed to be a binary variable (Manning et al., 2008), and therefore appears as a strong candidate classifier to be considered for evaluation on the MVC dataset. The default implementation of Bernoulli Naïve Bayes provided in Scikit-learn was used.

2.4.5. Artificial Neural Networks

Artificial Neural Networks (ANN) is a mathematical formulation of a biological brain. Like the brain, an ANN model consists of several neurons interconnected over different layers (Haykin, 2009). This interconnected structure models and stores the information about the complex relationships between the predictor and target variables. The experience learned from the training data is stored as weights and biases for each individual neuron. The ANN model emulates a non-linear function between the input predictor variables and the output target variable as

$$y = F(U; w),\qquad(2)$$

where F is the non-linear function estimation, U is the set of input predictor variables and y is the target variable. The weights w for the ANN model are decided based on an optimization algorithm, which minimizes the error between the estimated and the actual target variable during the training process.

Consider a training set $(U(i), d(i))_{i=1}^{N}$ with N sample points. $F(U(i); w)$ is emulated by the ANN model, $d(i)$ is the desired value of the output corresponding to the inputs $U(i)$, and the matrix w is a weight matrix. The network training is achieved by minimizing the loss function $\mathcal{E}$ defined as

$$\mathcal{E} := \frac{1}{N} \sum_{i=1}^{N} [d(i) - F(U(i); w)]^2 \qquad(3)$$

In this study the ANN models used had one hidden layer with 20 neurons, and the training was achieved by the Levenberg-Marquardt training algorithm (Marquardt, 1963; Levenberg, 1944).

2.5. Performance estimates

The area under the receiver operator characteristic (ROC) curve (AUC) was used to measure the performance of the model. The ROC shows the power in terms of sensitivity and specificity for prediction of severe injury for different cutoff values of $P(Y = 1)$ (probability of sustaining severe injury). The cutoff value determines the trade-off between sensitivity and specificity; increasing the sensitivity (identifying more MVC occupants with severe injury) is at the cost of decreasing the specificity (more false positives). Use of the OSISP algorithm in the field will require finding a suitable value for this cutoff. This is best determined by the management for health care trauma systems and is outside the scope of this study.

A 10-fold stratified cross validation (SCV) procedure was performed to estimate the performance of the OSISP algorithm for unseen data, in terms of ROC and AUC. The dataset was divided into ten randomized folds with approximately equal number of MVC occupants and similar distributions of severe/non-severe injury. One fold at a time was left out, a classification model was derived on data from the remaining nine folds. The model was then validated by classifying the observations in the left out fold. This procedure

was repeated for all folds, i.e. performed ten times. Mean and 95% CI ($\pm$2 standard deviations from mean, assuming normal distribution) for the classification performance were calculated from the ten scores.

2.6. Feature ranking and variable subset performance

An advantage of machine learning is that it allows to determine which features are important and which can be considered redundant or unneeded. The outcome is a subset of relevant features. This process is often referred to as *feature selection*. An important goal in feature selection is *feature ranking*. When developing an OSISP model, it is valuable to know the ranking of the predictor variables (features) with respect to their importance in determining injury severity.

There are numerous approaches available for feature ranking. However, selection of a suitable method requires deep knowledge of the problem domain. In this study, our goal was to acquire ranking of the features for each machine learning algorithm. This requires the algorithm to inherently assign scores to the features. The coefficients in Logistic Regression (and its two variants), can serve for this purpose.

We followed the widely used *recursive feature elimination* (Guyon et al., 2002) approach to carry out the feature ranking. It starts with the full set of features and recursively removes the least significant feature(s), building a new model for each step. The process continues recursively on smaller subsets of features, until a desired number is reached. In our case, a regression model was trained on the full set of predictor variables, and their importance was determined through the regression coefficients. Then, by recursively eliminating the most redundant feature, the recursion stopped when only one feature was left (i.e. all were ranked). The algorithm's classification performance in terms of AUC was evaluated as a function of increasing number of features, in the order of their ranking from best to worst. It demonstrated which and how many features that were needed to approach the performance for the full feature set.

3. Results

3.1. Classification performance for all methods

Table 2 shows the classification performance in terms of AUC evaluated by 10-fold SCV for the top performing classifiers and imputation methods. The highest accuracy obtained was AUC = 0.86 (95% CI 0.82–0.90) by Logistic Regression. Consistently under all four approaches of handling missing data, both Logistic Regression and Ridge Regression achieved high accuracy. Logistic Regression, with marginally better 95% CI values, was the top performing classifier. SGD performed almost on par with the best classifiers, whereas Bernoulli Naïve Bayes and ANN showed slightly lower accuracy. The four methods of handling missing data produced relatively similar results, with Conditional Probabilities and New Category yielding best performance. In Section 3.2, we show more detailed results for the top performing method Logistic Regression.

3.2. Detailed results for proposed OSISP algorithm

The ROC curve for the top performing classifier Logistic Regression (Table 2), implemented with Conditional Probabilities imputation and evaluated by 10-fold SCV, is shown in Fig. 1. Examples of undertriage and overtriage rates, based on the recommendations by ACS-COT and adding overtriage at 1% undertriage, is shown in Table 4. The Logistic Regression model is detailed in Table 3, presenting the levels of statistical significance (*p*-values), OR and 95% CI for each variable.

The classification performance as a function of number of predictor variables, derived using the feature ranking procedure, is shown in Fig. 2. In order of importance, Ejection, Entrapment, Belt use, Airbag deployment and Crash type were the five strongest predictors, and together yielded an AUC approaching the full feature set (Fig. 2). The classification accuracy improvement leveled off after adding around five to seven of the variables with highest prediction power, using more variables generated relatively small improvements.

Table 2

AUC values for the different classifiers and imputation methods. CP = Conditional Probabilities.

Model	Imputation methods AUC [95% CI]			
	CP	New Category	Removed	Mode
Logistic Regression	0.86 [0.82, 0.90]	0.86 [0.82, 0.90]	0.85 [0.82, 0.89]	0.84 [0.79, 0.89]
Ridge Regression	0.86 [0.81, 0.90]	0.86 [0.82, 0.89]	0.85 [0.82, 0.89]	0.84 [0.81, 0.87]
Stochastic Gradient Descent	0.85 [0.82, 0.88]	0.85 [0.82, 0.88]	0.84 [0.80, 0.88]	0.83 [0.78, 0.88]
Bernoulli Naïve Bayes	0.84 [0.80, 0.87]	0.83 [0.80, 0.86]	0.82 [0.80, 0.84]	0.82 [0.79, 0.84]
Artificial Neural Networks	0.78 [0.55, 1.02]	0.83 [0.80, 0.87]	0.78 [0.59, 0.96]	0.82 [0.76, 0.87]
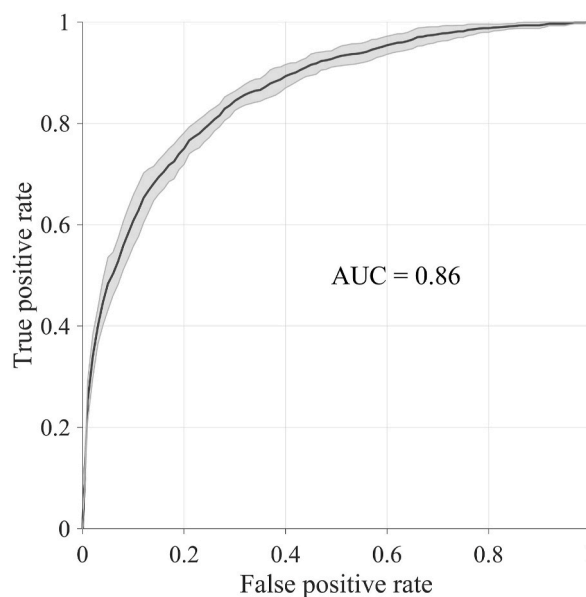
Fig. 1. The ROC curve and AUC value for the Logistic Regression classifier. The shaded area shows one standard deviation from the mean.

Table 3
Detailed results for the Logistic Regression classifier. The variable level used as reference is shown within parenthesis. The asterisks following variable levels indicate level of statistical significance: *, ** and *** denote $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$, respectively.

Variable	β	p-value	OR (e^{β}) [95% CI]
Sex (Male): Female	−0.049	0.48	0.95 [0.83, 1.09]
Elderly ($\leq$55): >55 ***	0.80	<10^{-4}	2.22 [1.90, 2.60]
PSL ($15 \leq x < 25$)			
$25 \leq x < 35$	−0.065	0.94	0.94 [0.16, 5.35]
$35 \leq x < 45$	−0.073	0.91	0.93 [0.26, 3.28]
$45 \leq x < 55$	0.29	0.65	1.34 [0.38, 4.67]
$55 \leq x < 65$	0.50	0.43	1.65 [0.48, 5.72]
$65 \leq x < 75$	0.64	0.32	1.89 [0.54, 6.57]
$75 \leq x < 85$	0.59	0.36	1.81 [0.51, 6.46]
$85 \leq x < 95$	0.80	0.21	2.24 [0.64, 7.79]
$95 \leq x < 105$	1.26	0.051	3.53 [0.99, 12.57]
$105 \leq x < 115$	0.79	0.21	2.21 [0.63, 7.77]
$115 \leq x < 125$	1.22	0.074	3.40 [0.89, 13.00]
Belt Use (Yes)			
Unbelted***	1.33	<10^{-4}	3.79 [3.19, 4.50]
Unknown**	−0.71	0.0092	0.49 [0.29, 0.84]
Vehicle (Truck): Car***	0.40	<10^{-4}	1.49 [1.29, 1.72]
Airbag (Not deployed)			
Deployed***	1.27	<10^{-4}	3.58 [3.06, 4.17]
Not available***	1.21	3.90×10^{-4}	3.56 [1.72, 6.55]
Location (Urban): Rural***	−0.30	1.50×10^{-4}	0.74 [0.63, 0.86]
Seat Position (Back): Front	0.58	0.10	1.78 [0.89, 3.58]
WN (No): Yes	0.10	0.32	1.11 [0.91, 1.35]
Crash (Rear End)			
Single***	1.16	<10^{-4}	3.20 [2.53, 4.04]
Head On***	1.41	<10^{-4}	4.09 [3.05, 5.49]
Other*	0.30	0.015	1.35 [1.06, 1.71]
Eject (No): Yes***	2.40	<10^{-4}	10.98 [8.32, 14.50]
Entrap (No): Yes***	2.54	<10^{-4}	12.62 [10.02, 15.92]
Rollover (No): Yes	0.18	0.078	1.20 [0.98, 1.47]
Light (Daylight)			
Dark**	0.27	0.0096	1.32 [1.07, 1.62]
Dark/Lighted***	0.37	<10^{-4}	1.45 [1.23, 1.71]
Dawn*	0.48	0.034	1.61 [1.04, 2.50]
Dusk	−0.25	0.35	0.78 [0.46, 1.31]
Intercept (β_0)***	−6.32	<10^{-4}	–

Table 4
Examples of undertriage (sensitivity or true positive rate) and overtriage ([1 - specificity] or false positive rate) for the Logistic Regression classifier.

| Undertriage (%)	Overtriage (%)
1	83
5	59
13	35
20	25

4. Discussion

4.1. Significance of findings

The main finding is that an OSISP algorithm is capable of predicting severe injury in a US population of MVC occupants with an AUC of 0.86 (95% CI 0.82–0.90, Table 2), based only on variables deemed to be feasible to assess on the scene of crash by first responders. Only a subset of the 5–7 strongest predictors is needed to attain good performance (Fig. 2). The study shows that similar classification performance is achieved by several machine learning methods, and that the traditional method Logistic Regression appears to be a good choice for developing injury severity prediction algorithms (Table 2).

To put these findings into perspective, the proposed OSISP algorithm outperforms triage accuracy reported for field triage protocols, and performs on par with the best AACN algorithms. Rehn et al. (2012) reported an undertriage of 19.1% at an overtriage of 71.6% in their study on 1812 patients admitted to a primary trauma center, whereof 768 had major trauma (New Injury Severity Score over 15), after introducing a new two-tiered trauma team activation protocol. As previously mentioned, Stitzel et al. (2016) showed that undertriage was 20% with an overtriage of 54% for NASS-CDS data for years 2000–2011, and Xiang et al. (2014) demonstrated undertriage of 34% in US emergency departments. In contrast, the OSISP algorithm achieves undertriage of 5% at overtriage of 59% (Fig. 1, Table 4), indicating that the method could be used to improve performance of field triage for US MVC occupants.

We chose to highlight the results for the Logistic Regression classifier. It showed the highest performance; however, the 95% confidence intervals overlapped with the other classifiers (Table 2) so there is no evidence that Logistic Regression is likely to have the best performance on a prospective dataset. However, Logistic Regression has some important advantages for an OSISP model compared to most machine learning models. The model is explainable and the result is easier to interpret than models such as ANN that does not provide model coefficients or OR for the predictor variables. This could be an advantage when introducing the model in a prehospital setting, the prehospital staff can relate their experience to how the mathematical model works. Logistic Regression is also less complex than advanced mathematical models underlying many other classifiers, and can potentially be less prone to overfitting than more complex models, especially when regularization techniques are employed. Due to these advantages and the results from the current study we therefore suggest to use Logistic Regression as the basis for OSISP. However, it should still be benchmarked against other machine learning methods in future studies, since it may still be outperformed by other classifiers, especially on larger and more intricate datasets where more complex models may enhance performance. Furthermore, new methods may be needed to handle the problem with unobserved heterogeneity in the MVC data, which will limit the performance of injury severity prediction.

Compared to previous studies on OSISP using data from Sweden (Buendia et al., 2015; Candefjord et al., 2015), the algorithms
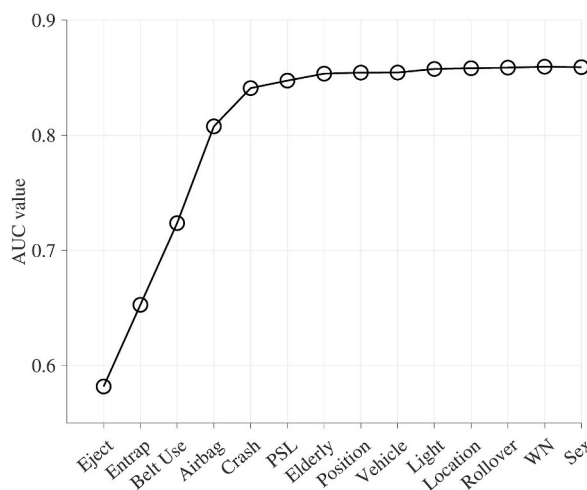
Fig. 2. The classification performance in terms of AUC for the Logistic Regression classifier evaluated as a function of number of predictor variables, input subsequently in the order of individual prediction power.

developed in the current study perform better. A plausible explanation is that the two strongest predictors in the current model, Eject and Entrap (Fig. 2), were not used in the earlier studies because this data is not available in the national Swedish MVC dataset (Buendia et al., 2015). The model developed in the present study includes both occupants in light trucks and cars, which should simplify field use compared to the earlier models that were separate for trucks (Candefjord et al., 2015) and passenger cars (Buendia et al., 2015).

The trends for the odds ratios for the different variables included in the study (Table 3) largely follow the trends reported in the literature (Candefjord et al., 2015; Buendia et al., 2015; Sasser et al., 2012). A surprising finding in the present study was that crashes in rural environment were less dangerous than urban crashes (odds ratio 0.74, $p < 0.05$). The finding that Logistic Regression slightly outperformed other machine learning methods is in agreement with the study by Kusano and Gabler (2014). They evaluated different injury risk classifiers for developing AACN, including Logistic Regression, Random Forest, AdaBoost, Naïve Bayes, Support Vector Machine, and classification k-nearest neighbors. They used a NASS-CDS dataset aggregating years 2002–2011 to include 16398 vehicles involved in non-rollover collisions. The best models used Logistic Regression and yielded AUC of 0.86–0.89, where the highest accuracy was obtained for models including age and sex as predictors but this only contributed to a small improvement (past AACN models have been criticized for relying on age and sex) (Kusano and Gabler, 2014).

Compared to AACN algorithms, Stitzel et al. (2016) demonstrated that their Occupant Transportation Decision Algorithm (OTDA) achieves < 50% overtriage and <5% undertriage in side impacts and 6–16% undertriage in other types of crashes. They showed that this is an improvement in terms of lowered undertriage compared to the algorithms URGENCY and On Star (Stitzel et al., 2016), developed and evaluated in previous works (Augenstein et al, 2002, 2003; Bahouth et al., 2004; Rauscher et al., 2009; Kononen et al., 2011). The OSISP algorithm performs on par with the OTDA algorithm, showing an undertriage of approximately 7% at 50% overtriage (Fig. 1, OSISP is a single model for all crash types). OSISP attains high accuracy without utilizing vehicle telemetry data. In the future, an effective way of assuring most effective rescue and improving field triage could be to use AACN in conjunction with OSISP. For supporting dispatch planning of rescue missions, an algorithm based on vehicle telemetry data alone could be used, such as the OTDA (Stitzel et al., 2016). For supporting transport destination decisions OSISP could be used at the scene of crash, as additional important information then can be recognized, such as occupant being entrapped or ejected. Currently, vehicle telemetry data is not available for most crashes. OSISP can be used for all crashes and is straightforward to implement in a handheld device to be used at the scene of crash (Olaetxea Azkarate-Askatsua, 2017), and thus has high potential for improving field triage in the near future. In the long term, models like OTDA and OSISP could also be combined into a set of algorithms utilizing the most predictive data momentarily available in a continuum from awareness of accident to the MVC patient is delivered to the appropriate hospital, to provide a dynamic injury severity prediction to support decisions from dispatch to field triage.

We can perform a simplified benefit analysis following the work by Stitzel et al. (2016, p. 1217 and Table 5) on a population-weighted sample of NASS-CDS data. If we choose a threshold for the OSISP algorithm such that undertriage equals 5%, as deemed acceptable by ACS-COT (American College of Surgeons Committee on Trauma (ACS-COT), 2014), the cost is an overtriage of 59% (Table 4). Assuming that the classification accuracy would be similar for the whole US population, we expect an improvement of undertriage from 20% to 5%, with some increase in overtriage (59% versus 54%). This would translate to approximately 4600 more patients with severe injury being correctly triaged and receiving more appropriate care every year, if OSISP is implemented nationwide. If lowering overtriage with help of OSISP would be prioritized by the emergency medical services we could set a threshold producing undertriage of e.g. 10%, i.e. halving the rate of undertriage compared to current outcomes, which would yield an overtriage of 42%. This translates to more appropriate care for approximately 3100 severely injured patients, while reducing unnecessary use of trauma center resources for 94000 patients, every year.

4.2. Influence of variable encoding and data imputation

This study followed a rigorous protocol to assess the effects of different variable encoding and data imputation methods. Adapting some predetermined variable encoding and imputation method would have been simpler; however, such an approach is only viable if it is supported by the literature from similar studies. To our knowledge, no previous studies predicting the effect of different data representation and imputation methods on the accuracy of machine learning for an MVC dataset exist. In this study, we adapted two methods for encoding data and four ways to handle missing data. Independent of the imputation method and the type of learning algorithm, dummy variable encoding consistently produced better results than label encoding. For handling missing data, the Conditional Probabilities and New Category imputation methods produced the highest AUC scores (Table 2). However, the influence of different imputation methods was relatively small, which indicates that the classification performance is stable.

4.3. Limitations of the study

The size of the dataset in this study ($n = 21589$ occupants) is of the same order of magnitude as several similar studies (Candefjord et al., 2015; Kusano and Gabler, 2014; Kononen et al., 2011). However, it is smaller than the study by (Stitzel et al., 2016) ($n = 115159$ occupants). Unfortunately, ISS data was not available for a large proportion of the compiled NASS-CDS cases (approximately 43%), which had to be excluded. OSISP models can potentially achieve higher accuracy for larger datasets in future studies.

This study did not employ weighting factors provided by NHTSA to account for the NASS-CDS sampling system (Radja, 2016). NASS-CDS samples events that are harmful (property damage and/or injury), and at least one vehicle needs to be towed away. The system samples a small proportion of all crashes based on a design where the country is divided into 1195 geographic areas, which are further divided into police jurisdictions. Within each jurisdiction, every week crashes are selected for investigation using a strategy that increases the probability that high severity crashes are included. The weighting factors can be used to derive estimates

representative of the entire country, and assign less weight to more severe crashes (Radja, 2016). Since OSISP is mainly aimed to be used by prehospital personnel to complement the triage protocol (Buendia et al., 2015), we reasoned that it is likely that the crashes those teams will experience are more severe than most crashes in NASS-CDS, and that adjusting the NASS-CDS sample may make the dataset less representative for an OSISP model. Furthermore, crashes with no ISS data were excluded, and the lack of that information may not occur completely at random. There is potential bias in our models due to that the data sample may not accurately represent the patient population seen by prehospital personnel prospectively.

It is clear from Table 2 that all the algorithms selected for this study consistently showed good performance, except ANN. When performing the 10-fold SCV using ANN, we observed that individual AUC scores of some of the folds were as low as 0.5, i.e. no better than random. Thus, the average of ten folds dropped to 0.78 in two out of the four results reported for ANN in Table 2. There is never a guarantee that ANN learning will not get stuck in a local optimum, which is the most plausible explanation of the behavior of ANN in this study. Another possible reason is that we have used only one hidden layer, there may be need for a more complex network architecture to overcome this problem. On the other hand, one important aspect is that ANN often requires huge datasets to outperform traditional machine learning methods.

4.4. Future work

Both of the variants of Logistic Regression considered in this study, namely SGD and Ridge Regression, make use of regularization to avoid overfitting. Logistic Regression showed marginally better results than its counterparts with regularization. This affirms that the performance scores for Logistic Regression are realistic and not an outcome of overfitting. Thus, selecting Logistic Regression as the OSISP algorithm is well justified in our scenario. However, the importance of Ridge Regression and SGD cannot be completely overlooked, since the differences in performance (AUC scores) are not statistically significant. An OSISP algorithm based on Ridge Regression or SGD is expected to deliver results comparable to those of Logistic Regression, or even better if the data available for model construction is noisy. Therefore, designing future models using new data consisting of more variables for the problem addressed in this study, all three versions of regression presented here should be considered for developing the OSISP model. ANN may have larger potential to improve its performance using larger and more diverse datasets, as compared to other machine learning methods. In the future, an OSISP algorithm could potentially also incorporate vehicle telemetry data and additional information from the scene of crash. Such rich information, with potentially complex patterns distinguishing severe injury, could lend itself well to the power of ANN. Therefore, we encourage the continued use of ANN for injury severity prediction.

To prove the potential benefits of OSISP for the US population, we recommend to perform a prospective clinical study in collaboration with emergency medical services. Ideally, such a study would benchmark the accuracy of the field triage protocol in use against the performance of the OSISP algorithm, with aim to validate the findings in the present retrospective study. The OSISP algorithm can be implemented in a smartphone or tablet for quick and easy recording of the included variables. A first version of a suggested design for a smartphone app has been presented in Olaetxea Azkarate-Askatsua (2017). The implementation should preferably be designed in close collaboration with medical responders.

OSISP has so far only been developed for MVC involving cars and trucks. Future studies could address models for other road users, such as motorcyclists, cyclists and pedestrians.

5. Conclusion

An OSISP algorithm for use in the US by first responders to predict the probability of MVC occupants being severely injured was developed, based on evaluations of several machine learning algorithms. The selected algorithm used Logistic Regression and showed high classification accuracy for differentiating severe and non-severe injury, and needed only a subset of the 5–7 strongest predictors to achieve good performance. Regression models appear to be well suited for MVC injury severity prediction. This study indicates that a simple to use OSISP tool for first responders could be utilized to improve field triage accuracy in the US, to improve care for thousands of severely injured patients every year, while reducing unnecessary use of trauma center resources for non-severely injured patients.

Declaration of competing interest

None.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Sanjay Ghemawat, Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, November 2–4, 2016. USENIX Association, Savannah, GA, USA, pp. 265–283.

American College of Surgeons Committee on Trauma (ACS-COT), 2014. Resources for Optimal Care of the Injured Patient: 2014. American College of Surgeons, Chicago, IL.

Association for the Advancement of Automotive Medicine (AAAM), 2005. Abbreviated Injury Scale (AIS) 2005. AAAM, Barrington, IL.

Augenstein, J., Perdeck, E., Stratton, J., Digges, K., Bahouth, G., 2003. Characteristics of crashes that increase the risk of serious injuries. Annu. Proc. Assoc. Adv. Automot. Med. 47, 561–576.

Augenstein, J., Perdeck, E., Stratton, J., Digges, K., Steps, J., Bahouth, G., 2002. Validation of the URGENCY algorithm for near-side crashes. Annu. Proc. Assoc. Adv. Automot. Med. 46, 305–314.

Bahouth, G.T., Digges, K.H., Bedewi, N.E., Kuznetsov, A., Augenstein, J.S., Perdeck, E., 2004. Development of URGENCY 2.1 for the prediction of crash injury severity. Top. Emerg. Med. 26, 157–165.

Baker, S.P., O'Neill, B., 1976. The injury severity score - an update. J. Trauma Inj. Infect. Crit. Care 16, 882–885.

Berwick, D., Embrey, E., Goldkind, S.F., Haider, A., Holcomb, J.B., James, B.C., Klein, J., Kupas, D.F., Laurencin, C., MacKenzie, E., Marcozzi, D., McCannon, C.J., McSwain Jr., N., Parrish, J., Redberg, R., Reinhardt, U.E., Robinson, J., Scalea, T., Schwab, C.W., Spinella, P.C., 2016. A National Trauma Care System: Integrating Military and Civilian Trauma Systems to Achieve Zero Preventable Deaths after Injury. The National Academies of Sciences, Engineering and Medicine. Technical Report.

Blincoe, L.J., Miller, T.R., Zaloshnja, E., Lawrence, B.A., 2015. The Economic and Societal Impact of Motor Vehicle Crashes (Report No. DOT HS 812 013). National Highway Traffic Safety Administration, Washington, DC. Technical Report.

Brink, H., Richards, J.W., Fetherolf, M., 2017. Real-World Machine Learning. Manning, Shelter Island.

Buendia, R., Candefjord, S., Fagerlind, H., Bálint, A., Sjöqvist, B.A., 2015. On scene injury severity prediction (OSISP) algorithm for car occupants. Accid. Anal. Prev. 81, 211–217. https://doi.org/10.1016/j.aap.2015.04.032.

Candefjord, S., Asker, L., Caragounis, E.C., 2020. Mortality of trauma patients treated at trauma centers compared to non-trauma centers in Sweden: a retrospective study. Eur. J. Trauma Emerg. Surg. https://doi.org/10.1007/s00068-020-01446-6.

Candefjord, S., Buendia, R., Fagerlind, H., Bálint, A., Wege, C., Sjöqvist, B.A., 2015. On-scene injury severity prediction (OSISP) algorithm for truck occupants. Traffic Inj. Prev. 16, S190–S196. https://doi.org/10.1080/15389588.2015.1057578.

Champion, H., Augenstein, J., Blatt, A., Cushing, B., Digges, K., Flanigan, M., Hunt, R., Lombardo, L., Siegel, J., 2005. New tools to reduce deaths and disabilities by improving emergency care: URGENCY software, occult injury warnings, and air medical services database. In: Proceedings of the 19th International Technical Conference on Enhanced Safety of Vehicles. NHTSA Sponsored), Washington, DC.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422.

Haas, B., Gomez, D., Zagorski, B., Stukel, T.A., Rubenfeld, G.D., Nathens, A.B., 2010. Survival of the fittest: the hidden cost of undertriage of major trauma. J. Am. Coll. Surg. 211, 804–811. https://doi.org/10.1016/j.jamcollsurg.2010.08.014.

Harrell, F.E., 2001. Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis and Logistic Regression. Springer Science and Business Media, Inc., New York.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed. Springer, New York.

Haykin, S.S., 2009. Neural Networks and Learning Machines, third ed. Prentice Hall, New York.

Hu, W., Dong, Q., Huang, B., 2017. Correlations between road crash mortality rate and distance to trauma centers. J. Transport Health 6, 50–59.

Kononen, D.W., Flannagan, C.A.C., Wang, S.C., 2011. Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accid. Anal. Prev. 43, 112–122.

Kusano, K., Gabler, H.C., 2014. Comparison and validation of injury risk classifiers for advanced automated crash notification systems. Traffic Inj. Prev. 15, S126–S133.

Lee, E., Wu, J., Kang, T., Craig, M., 2017. Estimate of mortality reduction with implementation of advanced automatic collision notification. Traffic Inj. Prev. 18 https://doi.org/10.1080/15389588.2017.1317090. S24–S30.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. 2, 164–168.

MacKenzie, E., Rivara, F., Jurkovich, G., Nathens, A., Frey, K., Egleston, B., Salkever, D., Scharfstein, D., 2006. A national evaluation of the effect of trauma-center care on mortality. N. Engl. J. Med. 354, 366–378.

Manning, C.D., Raghavan, P., Schütze, H., 2008. Introduction to Information Retrieval. Cambridge University Press, New York.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11, 431–441. https://doi.org/10.1137/0111030.

McCallum, A., Nigam, K., 1998. A comparison of event models for Naive Bayes text classification. In: AAAI-98 Workshop on Learning for Text Categorization, pp. 41–48.

Müller, A.C., Guido, S., 2016. Introduction to Machine Learning with Python: A Guide for Data Scientists, first ed. O'Reilly Media, Inc, Sebastopol, CA.

National Center for Injury Prevention and Control, 2008. Recommendations from the Expert Panel: Advanced Automatic Collision Notification and Triage of the Injured Patient. Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, Georgia. Technical Report.

Olaetxea Azkarate-Askatsua, I., 2017. An IT-Solution for Clinical Evaluation of an On-Scene Injury Severity Prediction Algorithm Intended for Road Crash Victims. Master's Thesis. Chalmers University of Technology, Gothenburg, Sweden.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Radja, G.A., 2016. National Automotive Sampling System – Crashworthiness Data System, 2015 Analytical User's Manual. National Highway Traffic Safety Administration (NHTSA). NHTSA Technical Report DOT HS 812 321.

Rauscher, S., Messner, G., Baur, P., Augenstein, J., Digges, K., Perdeck, E., Bahouth, G., Pieske, O., 2009. Enhanced automatic collision notification system - improved rescue care due to injury prediction - first field experience. In: Proceedings of the 21St (ESV) International Technical Conference on the Enhanced Safety of Vehicles, Held June 2009. National Highway Traffic Safety Administration (NHTSA), Stuttgart, Germany.

Ray, J.J., Meizoso, J.P., Satahoo, S.S., Davis, J.S., Haren, R.M.V., Dermer, H., Jill, G., Bahouth, G.T., Blackbourne, L.H., Schulman, C.I., 2016. Potentially preventable prehospital deaths from motor vehicle collisions. Traffic Inj. Prev. 17, 676–680.

Rehn, M., Lossius, H.M., Tjosevik, K.E., Vetrhus, M., Østebø, O., Eken, T., 2012. Efficacy of a two-tiered trauma team activation protocol in a Norwegian trauma centre. Br. J. Surg. 99, 199–208.

Sasser, S., Hunt, R., Faul, M., Sugerman, D., Pearson, W., Theresa, D., Wald, M., Jurkovich, G., Newgard, C., Lerner, E., Cooper, A., Wang, S., Henry, M., Salomone, J., Galli, R., 2012. Guidelines for field triage of injured patients - recommendations of the national expert panel on field triage. MMWR (Morb. Mortal. Wkly. Rep.) 61 (No. 1), 1–20, 2011.

Schiff, M.A., Tencer, A.F., Mack, C.D., 2008. Risk factors for pelvic fractures in lateral impact motor vehicle crashes. Accid. Anal. Prev. 40, 387–391.

Stitzel, J.D., Weaver, A.A., Talton, J.W., Barnard, R.T., Schoell, S.L., Doud, A.N., Martin, R.S., Meredith, J.W., 2016. An injury severity-, time sensitivity-, and predictability-based advanced automatic crash notification algorithm improves motor vehicle crash occupant triage. J. Am. Coll. Surg. 222, 1211–1219.e6.

Tattar, P.N., 2017. Statistical Application Development with R and Python: Power of Statistics Using R and Python, second ed. Packt Publishing, Birmingham.

Xiang, H., Wheeler, K.K., Groner, J.I., Shi, J., Haley, K.J., 2014. Undertriage of major trauma patients in the U.S. Emergency departments. Am. J. Emerg. Med. 32, 997–1004.