Accident Analysis and Prevention 144 (2020) 105683

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

A crash severity analysis at highway-rail grade crossings: The random survival forest method

Amin Keramati[a], Pan Lu[b,*], Amirfarrokh Iranitalab[c], Danguang Pan[d], Ying Huang[e]

[a] Upper Great Plains Transportation Institute, Dept. 2880, North Dakota State University, Fargo, ND 58108-6050, USA
[b] Department of Transportation, Logistics, and Finance, Upper Great Plains Transportation Institute, North Dakota State University, Fargo, ND 58108-6050, USA
[c] Impact Research LLC, 10480 Little Patuxent Parkway, Suite 1050 (Corporate 40), Columbia, MD 21044, USA
[d] Department of Civil Engineering, University of Science and Technology Beijing, Beijing 100083, China
[e] Department of Civil and Environmental Engineering, North Dakota State University, Fargo, ND 58108-6050, USA

ARTICLE INFO

Keywords:
Accident prediction
Railroad grade crossing
Machine learning
Random survival forests

ABSTRACT

This paper proposes a machine learning approach, the random survival forest (RSF) for competing risks, to investigate highway-rail grade crossing (HRGC) crash severity during a 29-year analysis period. The benefits of the RSF approach are that it (1) is a special type of survival analysis able to accommodate the competing nature of multiple-event outcomes to the same event of interest (here the competing multiple events are crash severities), (2) is able to conduct an event-specific selection of risk factors, (3) has the capability to determine long-term cumulative effects of contributors with the cumulative incidence function (CIF), (4) provides high prediction performance, and (5) is effective in high-dimensional settings. The RSF approach is able to consider complexities in HRGC safety analysis, e.g., non-linear relationships between HRGCs crash severities and the contributing factors and heterogeneity in data. Variable importance (VIMP) technique is adopted in this research for selecting the most predictive contributors for each crash-severity level. Moreover, marginal effect analysis results real several HRGC countermeasures' effectiveness. Several insightful findings are discovered. For examples, adding stop signs to HRGCs that already have a combination of gate, standard flashing lights, and audible devices will reduce the likelihood of property damage only (PDO) crashes for up to seven years; but after the seventh year, the crossings are more likely to have PDO crashes. Adding audible devices to crossing with gates and standard flashing lights will reduce crash likelihood, PDO, injury, and fatal crashes by 49 %, 52 %, 46 %, and 50 %, respectively.

1. Introduction and background

Safety at highway-rail grade crossings (HRGCs) in the United States is an important social concern because crashes at these crossings may result in severe injuries and fatalities as a consequence of the huge differences in mass between motor vehicles and trains. It is important for decision makers, transportation agencies, and stakeholders to reduce the frequency and severity of these traffic collisions and to improve HRGCs' safety performance. According to the Federal Railroad Administration (FRA, 2018), from 1981 to 2018, crash frequency at HRGCs decreased by 76.6 %. This reduction may be associated with the upgrading of crossing controls from passive to active (Lenné et al., 2011; Meeker et al., 1997; Millegan et al., 2009). However, even though crash frequency decreased significantly, HRGC fatality and injury rates (per crash) have increased from 7.7 % to 12.2 %, and from 35 % to 37 %, respectively, during the same period (FRA, 2018). Consequently, only considering crash frequency as a risk measure results in neglecting crossings with low expected crash frequency, but high potential for more severe collisions.

Most previous studies of HRGC safety either focus on the analysis of crash frequency (Zheng et al., 2019, 2016; Lu et al., 2020; Zhou et al., 2020) or crash severity. Crash severity studies at HRGCs are often based on categorical outcome modeling using historical police reports and FRA HRGC crash datasets (Eluru et al., 2012; Fan et al., 2015; Ghomi et al., 2016; Haleem and Gan, 2015; Hao and Daniel, 2016, 2014, 2013; Hu et al., 2010; Kang and Khattak, 2017; Liu and Khattak, 2017; Ma et al., 2018; Savolainen et al., 2011; Zhao et al., 2018; Zhao and Khattak, 2015; Zheng et al., 2018). Because of the random, discrete, and non-negative nature of collision data, one of the most common approaches to transportation safety studies is the use of generalized

* Corresponding author.
E-mail addresses: Amin.Keramati@ndsu.edu (A. Keramati), pan.lu@ndsu.edu (P. Lu), Airanitalab@huskers.unl.edu (A. Iranitalab), pdg@ustb.edu.cn (D. Pan), Ying.Huang@ndsu.edu (Y. Huang).

https://doi.org/10.1016/j.aap.2020.105683
Received 5 December 2019; Received in revised form 21 May 2020; Accepted 6 July 2020
Available online 10 July 2020
0001-4575/

linear models (GLMs) with crash frequency or severity as the response variables. For example, Heydari and Fu (2015) used a Poisson Weibull model for crash frequency. Statistical models (Lee et al., 2004; Lu and Tolliver, 2016; Oh et al., 2006; Ye et al., 2018), including zero inflated, hurdle, and generalized event count models, were applied to address data issues such as the excess number of zero accidents and over/under dispersions. Data mining methods, such as the hierarchical tree-based regression technique, are also adopted by Yan et al. (2010), Zhou et al. (2020), and Lu et al. (2020) to predict train-vehicle crash frequencies at public passive HRGCs in the United States. Heydari et al. (2018) proposed and applied a method to compare different geographic areas in terms of a pre-specified safety performance, which can measure crash frequency of a given type. Recently, Keramati et al. (2020) investigated highway-rail grade crossing geometric effects.

Most of the modeling techniques used in previous crash severity studies focused on discrete choice approaches because of the discrete nature of crash severity levels. Hao and Daniel (2013) and Abdel-Aty and Keller (2005) applied an ordered Probit model to consider U.S. crash severity levels are naturally ordered. Hao and Daniel (2014) followed up on the previous work and adopted an ordered probit model by considering various control measures at HRGCs. Eluru et al. (2012) investigated crash severity by considering passive and active controls with a latent ordered response model using 10 years of crash data at HRGCs in the United States. Their results indicated that crashes at crossings with gates were less likely to result in injuries compared with crossings with only crossbucks, while crashes at crossings with flashing lights are more likely to result in injuries and fatalities. Hu et al. (2010) applied a generalized logit model to assess factors affecting crash severity at Taiwan's railroad grade crossings. Recently, Zhao et al. (2018) used binary logit models and generalized linear mixed models to study the association of potential factors with pedestrian injury severity levels using 10 years of crash data at HRGCs in the United States.

The characteristics of HRGCs at the time of a crash (e.g., crossing controls and highway traffic volume) may change over time, including before and after the crash occurrences (Liu and Khattak, 2017). Therefore, estimation of crash severity likelihoods by accounting for the time effects of contributing factors may improve the accuracy. The above studies do not account for these time effects, probably because of the complexity this consideration may add to the modeling process. In this study, to account for these time effects, the main interest is the time until the occurrence of a crash with a specific severity level. The analysis may be complicated because of (1) the need to identify a connection between a set of contributors to the time of occurrence of each crash severity level in a prediction model, and (2) crossings' crash histories can be recorded for a limited time period (e.g., 29 years in this study), and only the time of a crash (with any severity) during the study period can be recorded (otherwise crossings are recorded as event-free without any information from after the study time-period, leading to right-censored data). Consequently, specific algorithms are needed to take these characteristics into account.

Survival analysis has been commonly used to analyze the failure time, including biological death, engineering failure (e.g., mechanical failure), and other factors. (Fan et al., 2010). In survival analysis, time-to-event data are modeled, and are usually open to censoring because of the study period termination. In addition, the main goal of survival analysis is to explore the dependence of the survival time (failure time) on the covariates vector (Fan et al., 2010). Therefore, survival analysis is one approach that is able to consider the mentioned characteristics of the HRGC time-to-crash occurrence data. In this study, to quantify the crash severity likelihood over the timespan, the competing risk model (CRM) is used as the specific type of survival analysis. For crash analysis, CRM's main goal is estimating crash occurrence likelihoods in the presence of more than one crash events, including property damage only (PDO), injury, and fatal crashes. Moreover, those multiple events are competing with each other. In other words, a crossing categorized in one of these events will not be categorized in any of the other events. Correspondingly, in this study, the target of CRM utilization is quantifying the likelihood of collisions occurring at a crossing during a 29-year time span, while the crossing possibly experiences more than one event. These events are assigned the above mentioned three crash severity levels.

There are several approaches to competing risks, such as cause-specific Cox regression (Cox, 1972), the proportional hazard model of Fine and Gray (1999), and the Cox likelihood-based boosting (Binder and Schumacher, 2008). This paper proposed a novel study approach for crash severity at HRGCs based on survival analysis. In particular, the use of a machine learning approach, random survival forest (RSF), to account for competing risks for crash severity is selected to solve CRM (Ishwaran et al., 2008). According to Ishwaran et al. (2014), for the low-dimensional quadratic and interaction models, RSF has better performance compared with other approaches using competing risk models. They also emphasized that RSF outperforms the Cox-based boost in the high-dimensional quadratic and interaction model simulations. Another advantage of the RSF approach is that it is able to deal with heterogeneous data (Pölsterl et al., 2016). RSF can discover non-linear impacts and interactions adaptively and is a fully nonparametric approach (Ishwaran et al., 2010). Finally, because of averaging over trees and randomizing during the tree growing step, RSF is able to approximate complex survival functions with a low level of prediction error (Ishwaran et al., 2010).

In summary, the RSF approach is selected in this study to investigate crash severities as competing risks at HRGCs because (1) it is able to consider complexities in HRGC safety analysis, e.g., non-linear relationships between HRGCs crash severities and the contributing factors, and heterogeneity in data, (2) it has shown better performance (more accurate crash event predictions) compared with the other approaches to considering competing risks (Ishwaran et al., 2014), and (3) it is not only able to estimate long-term time effects on crash severities but also able to take crash severity dependency or competing nature into consideration. This method has been verified and applied extensively in medical research. However, to the knowledge of the authors, this is its first application to transportation safety. The authors seek to investigate this model's interpretive capabilities in crash severity analysis through the application of HRGC data.

2. Modeling methodology

2.1. Competing risk models

The competing risk model (CRM) is a specific type of survival analysis with the objective of estimating incidence probability in the presence of more than one failure cause. Correspondingly, in this study, the CRM's purpose is to estimate the likelihood of crash event occurring at a crossing during a time period, while it possibly experiences more than one crash event. These events are assigned the following crash severity levels in this study: PDO crash, injury crash, and fatal crash. The cause-specific hazard function describes the instantaneous failure rate at time t with severity level j for subjects that are not failing from severity level j by time t. The cause-specific hazard function for crash severity j is defined in Eq. (1) (Putter et al., 2007) shown below:

$$\alpha_j(t) = \lim_{\triangle t \to 0} \frac{P(t \le T_j < t + \triangle t | T_j \ge t)}{\triangle t} \tag{1}$$

In the above equation, T_j denotes the time to failure with severity level j. Note, the probability that a crash occurs with specific severity level in a specific time period depends on the cause-specific hazards of the other severity levels. The dependency assumption is met in this analysis because of their competing nature.

2.2. Competing risks forest algorithm

Random survival forest (RSF) is a new approach to competing risks, and which is an extension of Breiman's random forests (Breiman, 2001). The required steps to apply the competing risks forest algorithm can be summarized as follows (Ishwaran et al., 2014):

1) Draw B bootstrap samples from the learning data.
2) Grow a competing risk tree of each bootstrap sample by using splitting rules.
3) Quantify cumulative incidence function (CIF) for each tree b *after* the tree is grown to full size.
4) Obtain the ensemble.

In the next sub-sections, definitions and mathematical formulations needed to apply the above RSF approach steps are explained in detail (Ishwaran et al., 2014).

2.2.1. Bootstrap sampling (Step 1)

RSF is an ensemble method that grows random and independent survival trees and produces an aggregate outcome based on their outcomes. Similar to regular random forests, independent bootstrap sampling is used to grow each tree for the learning data by using random feature selection at each node. RSF is able to handle multicollinearity with bootstrap sampling. The process of random node splitting decreases the competition between highly correlated variables and in return decreases the correlation between the trees in the forest. RSF trees are generally grown deeply with many terminal nodes. In competing risk forests, trees are grown similarly to regular random forests. However, the splitting rules (Section 2.2.2) can be different. The trees are grown with splitting rules, then the estimated values, which are calculated within the terminal nodes, are selected to make the ensemble (Section 2.2.4). In this study, to grow a competing risk forest, a single competing risk tree is grown in each bootstrap sample. The advantages of this approach are (1) different forests are fitted based on both event-specific and combined event-specific splitting rules across the j crash severity levels, which helps to compare and select the better one according to their prediction performance; (2) this approach is efficient for large data settings, according to Ishwaran et al. (2014).

2.2.2. Splitting rules (Step 2)

In growing a competing risk tree for each bootstrap sample, candidate covariates are randomly selected at each node of the tree. The next step is splitting using the candidate covariate that maximizes competing risk splitting rule. In this study, the composite splitting rule is used to grow competing risk trees, which can be estimated based on combining the cause-specific splitting rules of the log-rank test or that of the approximation of Gray's test (Fine and Gray, 1999). The reason for selecting the composite splitting rules in this study is because the study's aim is to predict the CIF of all causes simultaneously (Ishwaran et al., 2014).

Let $(T_i, \delta_i)_{1\le i\le n}$ denote the survival times and event indicator, and $t_1 < t_2 < \cdots < t_m$ indicate the distinct event times. If the split at the current node can be formed like $x \le c$ and $x > c$ for a continuous or categorical covariate x, this split can form two child nodes including two new sets of competing risk data. If $\alpha_{jl}(t)$ and $\alpha_{jr}(t)$ are the cause-j (in this study, severity-j) specific hazard rates for the left and right child nodes, respectively, the log-rank test is a test of the null hypothesis H_0: $\alpha_{jl}(t) = \alpha_{jr}(t)$ for all $t \le \tau$. To split at the value c of variable x, the splitting rule can be formulated in Eq. (2).

$$L_j^{LR}(x, c) = \frac{1}{\hat{\sigma}_j^{LR}(x, c)}\sum_{k=1}^{m} W_j(t_k)\left(d_{j,l}(t_k) - \frac{d_j(t_k)Y_1(t_k)}{Y(t_k)}\right) \tag{2}$$

In the above equation, $\hat{\sigma}_j^{LR}(x, c)$ indicates the amount of standard deviation of value c for variable x. $Y_l(t)$ and $Y_r(t)$ denote the number of HRGC records at risk at time t in the left and right child nodes, respectively. The total number of HRGC records at risk is $Y(t) = Y_l(t) + Y_r(t)$. Similarly, $d_{j,l}(t)$ and $d_{j,r}(t)$ represent the number of type j events at time t for the left and right child, respectively, and $d_j(t) = d_{j,l}(t) + d_{j,r}(t)$ denotes the total number of j events at time t. As can be seen in Eq. (2), the test is influenced by the weighted difference of the cause-specific Nelson-Aalen estimates for the two child nodes. The time-dependent weight $W_j(t)$ defines the j severity level event involvement in splitting node. $W_j(t) = 1$ indicates the standard test that has optimal detecting power for event type j based on proportional cause-specific hazards and $W_j(t) = 0$ indicate event type j is not considered in splitting a node. Maximizing $|L_j^{LR}(x, c)|$ over x and c, the optimized split can be found.

According to the second splitting rules, Gray's test, splitting the trees can be obtained by selecting variables based on their direct impact on the cumulative incidence. Therefore, the splitting rules modeled after Gray's test (Gray, 1988), in which testing null hypothesis is H_0: $F_{jl}(t) = F_{jr}(t)$ for all $t \le \tau$ where, $F_{jl}(t)$ and $F_{jr}(t)$ are the results of the CIF for the left and right child nodes, respectively. The splitting rule based on the statistical score that uses the approximation of Gray's test is denoted by $L_j^G(x, c)$ and is given by substituting Y_j^* as modified risk set for Y and correspondingly Y_{jl}^* for Y_l in Eq. (2). The modified risk set can be estimated in Eq. (3).

$$Y_j^* = \sum_{i=1}^{n} I(T_i \ge t \cup (T_i < t \cap \delta_i \ne j \cap C_i^0 > t)) \tag{3}$$

Note that in above equation, Let T_i^o be the crash/event occurrence time associated with the i^{th} record of crash dataset (e.g., $i = 1,\ldots,n$), and let δ_i^o be the crash severity level (event type), $\delta_i^o \in \{1,\ldots, J\}$, where $J \ge 1$. Consequently, let C_i^o indicate the censoring time of individual record i while the actual crash occurrence time T_i^o is unobserved and one only observes $T_i = \min(T_i^o, C_i^o)$ and the crash (event) indicator $\delta_i = \delta_i^o I(T_i^o \le C_i^o)$.

According to Ishwaran et al. (2014), if the aim is to predict the CIF of all causes simultaneously, combining the cause-specific splitting rules across the event type (crash severity levels) can be useful, and test rules can be estimated in Eqs. (4) and (5).

$$L^{LR}(x, c) = \frac{\sum_{j=1}^{J} (\hat{\sigma}_j^{LR}(x, c))^2 L_j^{LR}(x, c)}{\sqrt{\sum_{j=1}^{J} (\hat{\sigma}_j^{LR}(x, c))^2}} \tag{4}$$

$$L^{G}(x, c) = \frac{\sum_{j=1}^{J} (\hat{\sigma}_j^{G}(x, c))^2 L_j^{G}(x, c)}{\sqrt{\sum_{j=1}^{J} (\hat{\sigma}_j^{G}(x, c))^2}} \tag{5}$$

Note, estimated $L^{LR}(x, c)$ and $L^G(x, c)$ in above equations can be determined by estimates of $L_j^{LR}(x, c)$ and $L_j^G(x, c)$ by Eq. (2).

2.2.3. Estimating cause-specific CIF for each tree (Step 3)

In an independent bootstrap sample, let $C_{i,b}$ denote the number of times that record i occurs in bootstrap sample b. To estimate the CIF related to the b^{th} tree, a case's covariate x is taken and dropped from the tree. In addition $h_b(x)$ contains the indices for cases from the learning data whose covariates share the terminal node with x. Considering $N_{j,b}(t x) = \sum_{i\in h_b(x)} C_{i,b} I\{T_i \le t, \delta_i = j\}$ as specific event counts and $Y_b(t x) = \sum_{i\in h_b(x)} C_{i,b} I\{T_i \ge t\}$ as the number of records at risks, the x's CIF can be estimated based on Eq. (6) (Ishwaran et al., 2014).

$$\widehat{F}_{j,b}(t x) = \int_0^t \widehat{S}_b(u{-}x) Y_b(ux)^{-1} N_{j,b}(du|x) \tag{6}$$

In Eq. (6), $\widehat{S}_b(t x) = \prod_{u\le t}(1 - \sum_j N_{j,b}(dux)/Y_b(u|x))$ is x's Kaplan-Meier estimate of event-free survival.

2.2.4. Ensemble method (Step 4)

The final step is aggregating cause-specific CIF for each tree ($\widehat{F}_{j,b}$), which are estimated in terminal nodes ($n_0 > 0$) of RSF trees to form the ensemble. The ensemble method used in this study is the out-of-bag (OOB) ensemble method. Based on the standard bootstrap theory, each bootstrap sample leaves out about 37 % of the data which constitutes the OOB portion of the data in that bootstrap sample. Therefore, OOB data are used to construct the OOB ensemble. The $O_i\{1,...,B\}$ is defined as the index set of trees where $c_{i,b}=0$ indicating O_i records trees, which case i is out-of-bag. Eq. (7) indicates the OOB ensemble estimates of the cause-*j* CIF.

$$\overline{F}_j^{oob}(tx) = \frac{1}{|O_i|}\sum_{b\in O_i}\widehat{F}_{j,b}(t|x_i) \tag{7}$$

2.3. Prediction performance

Brier score (BS) is a strictly proper scoring rule to assess prediction performance by defining the prediction error (Brier, 1950; Gneiting and Raftery, 2007; Winkler and Murphy, 1968). BS can be estimated by the squared difference of actual and predicted outcome. In this study, the time-dependent BS (Gerds and Schumacher, 2006; Graf et al., 1999) is considered to assess the ensemble CIF performance of each fitted RSF model as explained in Sub-section 2.2.3. The expected BS for crash severity *j* can be estimated by Eq. (8).

$$\hat{BS}_j(t) = E_i\{N_{j,i}(t) - \widehat{F}_j(t|x_i)\}^2 \tag{8}$$

In the above equation, $N_{j,i}(t)$ is equal to 1 if crossing record *i* experiences crash severity *j* before time *t*, and it is equal to 0 if record *i* has not experienced any of crash severities (event-free) until time *t*. correspondingly, $\widehat{F}_j(t|x_i)$ is estimate of cumulative incidence function of cause *j* for record *i* before time *t*.

2.3.1. Bootstrap cross-validation

To compare the prediction performance of different fitted RSF models, the bootstrap cross-validation approach is used. The term "bootstrap cross-validation" was first defined in Fu et al. (2005), whereby models are trained and tested in each bootstrap sample and cross-validated with all the data. Bootstrap cross-validation method was compared with many other cross-validation algorithms and recommended by Mogensen et al. (2012). The bootstrap cross-validation approach adopted by this research splits the original data D_N into a number of bootstrap training samples D_b (100 in this study) and corresponding test samples $D_N\backslash D_b$ (b = 1,..., B) without replacement from the original data. Decision tree models $\widehat{F}_{j,b}$ are then trained with each bootstrap training data D_b and prediction errors are calculated and tested with corresponding test sample. In the last step, the bootstrap cross-validation estimate of the prediction error for each crash severity (cause *j*) (*BootCvErr*) can be calculated by averaging over the test datasets by Eq. (9). For detailed information regarding to the cross-validation method, please refer to Fu et al. (2005) and Mogensen et al. (2012).

$$BootCvErr(t, j, \widehat{F}) = \frac{1}{B}\sum_{b=1}^{B}\frac{1}{M_b}\sum_{i\in D_N\backslash D_b}\widehat{W}_i(t)E_i\{N_{j,i}(t) - \widehat{F}_{j,b}(t|x_i)\}^2 \tag{9}$$

In Eq. (9), $\widehat{W}_i$ denotes the inverse probability of censoring weights (IPCW, Gerds and Schumacher, 2006). In addition, M_b indicates the size of the bootstrap samples for resampling without replacement.

3. Data

In this study, 29-year HRGC information is used, which includes all reported crashes/incidents and their current and historical crossing inventory information, geometric features of the crossings during the analysis period in North Dakota (ND). These data were extracted from three main data sources: (1) North Dakota roadway network, railway network, roadway intersections, and HRGCs from the ND Department of Transportation (ND_Hub, 2020); (2) highway-rail grade crossing accident/incident from Federal Railway Administration (FRA_Accident, 2020); and 3) highway-rail grade crossing inventory from the FRA database (FRA_Inventory, 2020). A total of 3310 crossings for ND public HRGCs from 1990 to 2018, including three crash severity levels (PDO, injury, and fatal) for each crash, were considered as the final study data. Table 1 summarizes all original variables used in the study analysis within 29 years. Note for each crossing, the corresponding categorical variable values are not always consistent within the 29-year analysis period, especially for crossing control these values need to be updated and treated accordingly in the model analysis. The values listed in Table 1 for all categorical variables and continuous variables indicate their changing ranges within 29 years. More descriptive statistics about the study covariates are provided in Table 1.

Table 1
Descriptive Statistics of Key Variables.

Variable Names	Categorical Variable Values	Min Frequency/ Value	Max Frequency/ Value
Crash Severity			
	No Crash	3163	3192
	PDO	2	18
	Injury	0	11
	Fatal Crash	0	6
Type of Train Service			
	Freight	2718	2807
	Intercity Passenger	387	476
ENS Sign Displayed			
	Yes	2055	2177
	No	1017	1139
Crossing Illumination Status			
	Illuminated	69	81
	Not Illuminated	3113	3125
Pavement Markings			
	No Marking	3111	3124
	Stop Lines	49	67
	RR Xing Symbols	16	25
Train Detection System			
	None	2398	2402
	Constant Warning Time (CWT)	375	378
	Motion Detection (MD)	42	43
	PTC	1	1
	DC	374	376
Commercial Power			
	Available	2107	2107
	Not Available	1087	1087
Roadway Paved Condition			
	Paved	563	563
	Not Paved	2631	2631
Crossing Control Types			
	Gate Only	4	22
	Standard Flashing Light	5	62
	Gate + Standard Flashing Light	11	42
	Cross Bucks + Stop Sign	44	78
	Gates + Standard Flashing Light + Audible Device + Stop Signs	2	14
	Gates + Standard Flashing Light + Audible Device	27	184
	Cross Bucks Only	2451	2676
Total Day Time Through Trains		0	35
Total Night Time Through Trains		0	33
Maximum Train Speed		5	79
Total Switching Trains		0	12
Maximum Train Speed		5	79
Highway Speed Limit		1	70
Percent of Trucks		1	22.67
Distance to the Nearest Intersections		0.78	2502
Crossing Angles		7.9	90
Number of Traffic Lanes		1	4

4. Results

In this section, the process and results of the application of the RSF approach to the 3310 HRGC crossings are presented. This includes: (1) selecting the RSF model with the least BS for all crash severity levels, (2) comparing the contributors' prediction power by estimating the variable importance (VIMP), (3) estimating the predicted cumulative crash severity probability of the most important contributors during the 29-year period by estimating its CIF, and (4) summarizing countermeasure effectiveness in cumulative crash and severity probability. Four RSF models (forests) are fitted. Three forests are fitted using log-rank splitting (Eq. 4) for each crash severity level using weights $(W_{PDO}(t) = 1, W_{Injury}(t) = 0, W_{Fatal}(t) = 0)$, $(W_{PDO}(t) = 0, W_{Injury}(t) = 1, W_{Fatal}(t) = 0)$, and $(W_{PDO}(t) = 0, W_{Injury}(t) = 0, W_{Fatal}(t) = 1)$ for forests denoted by RSF_{100}^{lg}, RSF_{010}^{lg}, and RSF_{001}^{lg}, respectively. Additionally, one forest is fitted using Gray's modified splitting rules (Eq. 5), RSF_{111}^{G} used $(W_{PDO}(t) = 1, W_{Injury}(t) = 1, W_{Fatal}(t) = 1)$.

In machine learning, there is a group of parameters that cannot be directly learned from the regular training process names hyperparameters. These parameters will define some properties of the machine learning models such as model complexity, the time required to train and test a model, and model prediction performance. In this research, the authors rely on experimental result of the out-of-bag (OOB) prediction errors, Brier score, and runtime to tune hyperparameters of the RSF. The parameters the authors tuned in this study are: the node size which controls the structure of the each individual tree, the number of mtry of variables considered as splitting candidate variables which controls level of randomness, the number of trees in the forest which controls the structure and the size of the forest, and the splitting rule which is splitting criteria in the nodes. A set of combinations of the parameters are used in the experiment of the study, they are mtry ranging from 1 to 7 with 1 as incremental change, node size ranging from 1 to 10 with 1 as incremental change, and number of trees ranging from 1 to 1000 with 50 as incremental change. The final selected parameter combination is mtry of 5, node size of 10, and number of trees of 1000. Finally, for each forest, 1000 trees are grown with 6 terminal nodes ($n_0 = 6$), 10 splits, 15 node size and 5 mtry.

4.1. Comparing four forests' BS

The cross-validation bootstrap approach is used to compare the time-dependent BS of four fitted forests. A total of 100 bootstrap samples of training and test data are considered ($B = 100$). The bootstrap without replacement is used to find a training set with the size of about 63.2 % of original data (2092 records), and corresponding test sets of about 1218 crossings (3310 – 2092 = 1218).

Fig. 1 illustrates the cumulative prediction error (BS) curve of each fitted forest for each crash severity level. The superscripts of G and lg in the forest notations in Fig. 1 indicate the Gray's modified splitting rule and log-rank splitting rule that the forest was built upon respectively. The subscripts of 111, 100, 010, and 001 indicate weight combination scenarios for three severity levels. In general, the time dependent BS results in Fig. 1 indicate less than 6.2 %, 4 %, and 2 % errors for PDO, injury, and fatal crash probability, respectively, during the 29-year prediction period. One can see from Fig. 1 part *a* that RSF_{111}^{G} and RSF_{100}^{lg} forests have better performance for predicting the PDO crashes compared with the other two forests because their cumulative prediction error is less. Fig. 1 part *b* indicates that the RSF_{111}^{G} and RSF_{010}^{lg} forests perform slightly better at predicting injury crash compared with the other forests. For predicting fatal crashes, Fig. 1 part *c* indicates all forests have almost identical performance up to the 20th year, and after 20th year, one can see RSF_{111}^{G} performs slightly better than the other three models. Since, in general, the forest using the composite Gray splitting rule (RSF_{111}^{G}) showed slightly the better performance among all four RSF models regarding to different splitting rules, it is selected as the best forest (model) for further analysis. In the next sections, this forest is used to assess the covariates' prediction power and for predicting the long-term cumulative probability change based on different combinations of active and passive crossing controls.

4.2. Variable importance

In RSF, variable selection typically is based on filtering variables on the basis of VIMP. Technically, in RSF, VIMP measures the increase or reduction in prediction error for the forest ensemble while a variable is randomly "noised-up" (Breiman, 2001). The larger positive VIMP indicates that the prediction accuracy of the forest is considerably reduced by noising the corresponding variable up. This indicates that the variable with larger VIMP is more predictive and contributes significantly to the prediction precision of a model. Because, in this section the target is estimating the VIMP for each covariate for each crash severity level, the event-specific VIMP is estimated. To calculate the event-specific VIMP, the prediction error is estimated based on BS error of Eq. (8), then data are noised-up by random node assignment, and the prediction error is recomputed. The VIMP for each covariate for each crash severity (event) *j* can be estimated by the differences in these two values.

Fig. 2 details the event-specific VIMP ranking based on the average VIMP of all three crash severities for HRGC covariates in the study, arranged from largest (train detection) at the top, to smallest (crossing angle) at the bottom. Contributors with VIMP close to zero denote that the variables have no contribution in predictive accuracy, and contributors with negative VIMP show the predictive accuracy can be improved when the variable is mis-specified; In other words, zero or negative VIMPs indicate non-predictive variables or indicate noise ((Ishwaran et al., 2008, 2010). Thus, the variables with negative and near zero values of VIMP can be ignored. The VIMP in Fig. 2 is calculated based on the forest using the composite Gray splitting rule (RSF_{111}^{G}). Train detection device, total switching train traffic, total daytime train traffic, maximum train speed, total nighttime train traffic, motorists' speed, crossing controls, highway speed limit, annual average daily traffic (AADT), and functional classification of road are identified as the 10 most predictive contributors based on their VIMP. One can see from Fig. 2 that the crossing control is among the 10 highest calculated VIMP (seventh) related to average of all severity levels. Consequently, to show the RSF model application in predicting the long-term cumulative probability, change in cumulative probability of each crash severity level is calculated under the effect of different combinations of active and passive crossing controls by estimating CIF. This is shown in the next section.

4.3. Cumulative probability and countermeasure effectiveness prediction

In this section, crossing control is selected as an example for conducting the cumulative probability marginal effect analysis because it is among the 10 contributors with highest average event-specific VIMP. As previously stated, the RSF model using the composite Gray splitting rule (RSF_{111}^{G}) is used to predict cumulative probability because, as shown in Section 4.1, it showed better performance compared with the other models. In this section, the safety outcomes of the following combination of active and passive controls are assessed: (1) gate (only), (2) standard flashing light (only), (3) crossbucks and stop signs, (4) gate and standard flashing light, (5) gate and standard flashing light and audible device, (6) gate and standard flashing light and audible device and stop sign.

The marginal effect analysis is conducted through non-randomized controlled group comparison. CIFs for each crossing, each crash severity level, and each year are calculated through Eq. (7), then the average yearly CIF for each severity level is calculated with the Eq. (10) among all the CIFs for the same type of crossing control, and lastly, the

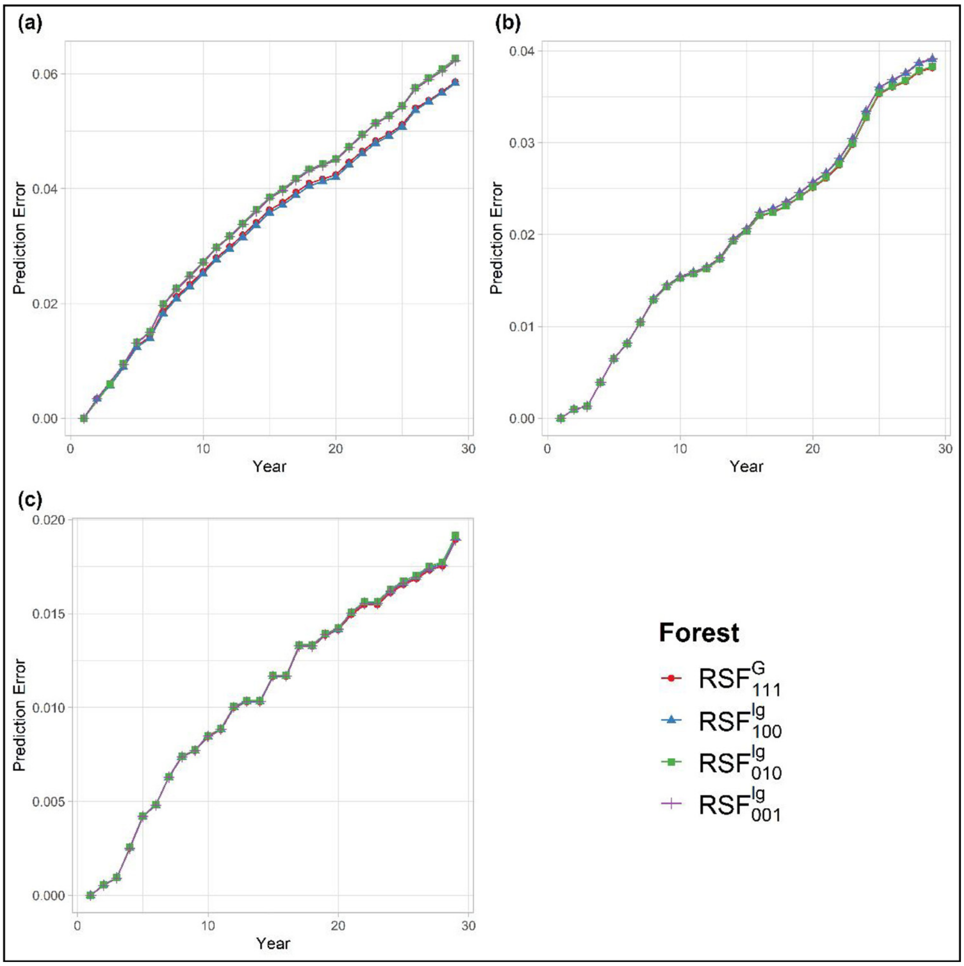

Fig. 1. Cumulative prediction error of each crash severity level over the 29-year study period.

marginal countermeasure difference is calculated with the following Eq. (11).

$$\bar{F}_j(tx_o) = \frac{\sum_{i=1}^{n} \bar{F}_j^{oob}(t|x_{oi})}{n} \tag{10}$$

Where, $\bar{F}_j(tx_o)$ denotes the average CIF for every control device o. t indicates the time and j indicated the severity level. X is variable of control device and o is index for six above levels of control device. n is the number of observations contains the same control device o for the year t and the severity level j.

$$D_{j,p-q}(t) = \bar{F}_j(tx_p) - \bar{F}_j(tx_q) \tag{11}$$

Where, $D_{j,p-q}(t)$ denotes the marginal effect of changing control device from q to p at year t for severity level j. if p is number 5 and q is number 4 in above mentioned control device, the D5-4 indicate the marginal effects at time t and severity j for adding control device of "audible device" to a crossing already equipped with "gate and standard flashing lights".

One of the advantages of RSF in this study is the direct estimation of predicted CIF for all the original data records during the 29-year period through the ensemble step by using Eq. (7). Therefore, to compare the change in predicted cumulative probability by comparing the different crossing control groups, the CIF for each crash severity level is stratified by the crossing control covariate values. In other words, the countermeasure effectiveness is calculated through non-randomized controlled group CIF comparison. Fig. 3 indicates the 29-year prediction of cumulative crash severity by comparing four pairs of crossing controls in parts *a*, *b*, *c*, and *d*.

4.3.1. Discussion

From Fig. 3 part *a*, one can see that crossings with gates only as an active crossing control will reduce all three crash severity likelihoods compared with the crossings with crossbucks and stop sign as a passive crossing control for almost 29 years. These results are consistent with previous studies that found that crossings with gates are associated with lower chances of injury severity compared with other crossings, specifically those with passive controls like stop signs and crossbucks, because gates are better able to prevent dangerous driving behaviors (Liu et al., 2015). However, the difference between PDO crash likelihood at crossings with gates only and crossings with crossbucks and stop signs begins to decrease after the middle of study time span (year 15). Although crossings with gates are less likely to have injury crashes compared with the crossings with crossbucks and stop signs during the first 20 years, crossings with both crossing control types show a similar performance after year 20. Fig. 3 part *b* indicates that by adding a stop sign to the combination of gate, standard flashing light, and audible device, reduces crossings' cumulative PDO and fatal probabilities up to year seven; but after the seven-year point, the crossings are more likely to have PDO and fatal crashes. In addition, over the 29-year period, injury crash likelihood is moderately higher (12 % increase on average) for crossings after adding the stop sign. These trends may be due to the fact that some traffic control devices, like stop signs, which are usually installed at regular highway intersections, may cause some drivers to be confused about the distance they should check for traffic when stop signs are used at HRGCs (Jeng, 2005). Drivers stopping in front of a

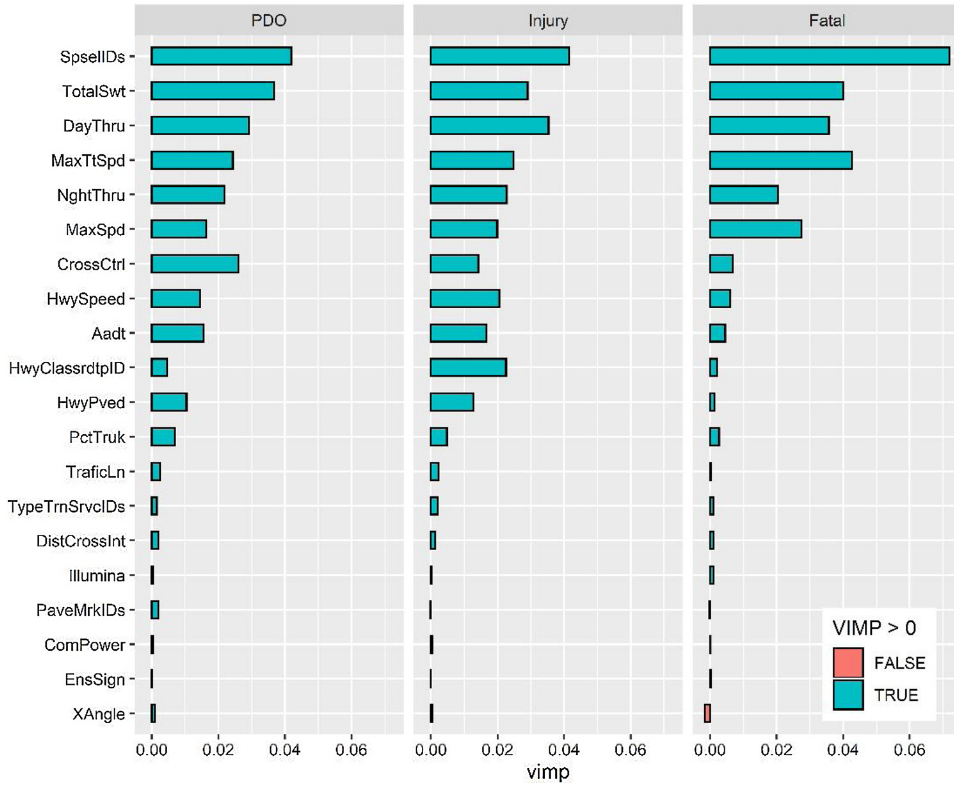

Fig. 2. Variable importance based on forest using composite Gray splitting rule.

highway-intersection stop sign would only check the traffic close to the intersection. However, drivers stopping in front of an HRGC stop sign should check a much longer distance for train traffic. Some researchers have argued against using passive controls like highway stop signs at grade crossings. For example, Burnham's (1995) investigations indicated that only 18 % of motorists could be alerted to the stop signs and 82 % were confused or semi-confused about the presence of stop signs at grade crossings.

Fig. 3 part *c* indicates that the probability of crashes (PDO, injury, and fatal) is considerably reduced for crossings with audible devices as additions to gates and standard flashing lights. They are reduced by 49 %, 52 %, 46 %, and 50 %, respectively, over 29 years. These results are also supported by previous investigations by the FRA (2018), which showed that crossings with gate and flashing lights without bells are more likely to be associated with driving around or through the gates, which indicates intentional trespassing behavior (Liu et al., 2015). Finally, Fig. 3 part *d* shows that crossings with gates and standard flashing lights are less likely to have PDO crashes compared with crossings with standard flashing lights only up to around year 23; then crossings with both control devices show similar performance. In addition, one can see from Fig. 3 part *d* that crossings with standard flashing lights only are less likely to have both injury and fatal crashes compared with crossings with gates and standard flashing lights, specifically after year five. This result is counterintuitive. It indicates that adding gates to crossings with only standard flashing lights will increase the long-term injury and fatal crash likelihoods. The potential rationale could be that people who are used to crossings with only standard flashing light crossings are more prone to gate violations. However, more specific experiments need to be conducted to better understand this phenomenon.

Table 2 summarizes the increase in average annual probability rates during the 29-year period. For example, although crossings with gate only show less likelihood of all crash severities during almost the entire 29 years compared with the crossings with crossbucks and stop signs (Fig. 3 part *a*), on average, the annual increase in likelihood of crash occurrence, injury crashes, and fatal crashes for crossings with gates only are about 1.33 %, 0.39 %, and 0.19 % compared with 1.31 %, 0.37 %, and 0.18 %, respectively, for crossings with the crossbucks and stop signs. Moreover, crossing control combinations are ranked based on annual crash likelihood increases, which showed that crossings with gates and standard flashing lights and audible devices are less likely to have both crash and crash severities in comparison with other crossing controls. Information delivered with Table 2 is the time effects on crash likelihood. Traditionally, HRGC agencies make safety improvement decisions based solely on the predicted crash/severity likelihood. With the results suggested in this study, time effect trend can be very different than the instantaneous likelihood trend and it should be considered in the safety improvement decision making. For example, crossings with gate only have less likelihood on injury crash however its annual increasing rate is much higher than the ones for crossings with crossbucks + stopsigns. If the improvement decision making is considering long-term effects then both instantaneous crash/severity likelihood and annual average crash/severity increasing rate should be considered.

5. Conclusions and discussions

In this study, the machine learning method of random survival forests (RSF) is used as a novel approach to applying the competing risk model to examine crash severity at public highway-rail grade crossings (HRGCs) in North Dakota from 1990 to 2018. RSF demonstrates its capability to identify cause-specific contributors and its ability to model crash severity likelihoods. RSF is able to produce easy-to-interpret results, which are expressed as the estimated crash severities cumulative probabilities. These results are produced by direct estimation of the cumulative incidence function (CIF). It also demonstrates its accuracy in prediction, as the time-dependent Brier Score results indicated less than 6.2 %, 4 %, and 2 % errors for PDO, injury, and fatal crash probability, respectively, over the 29-year prediction. RSF is also able to capture contributors' effects on prediction error of each competing

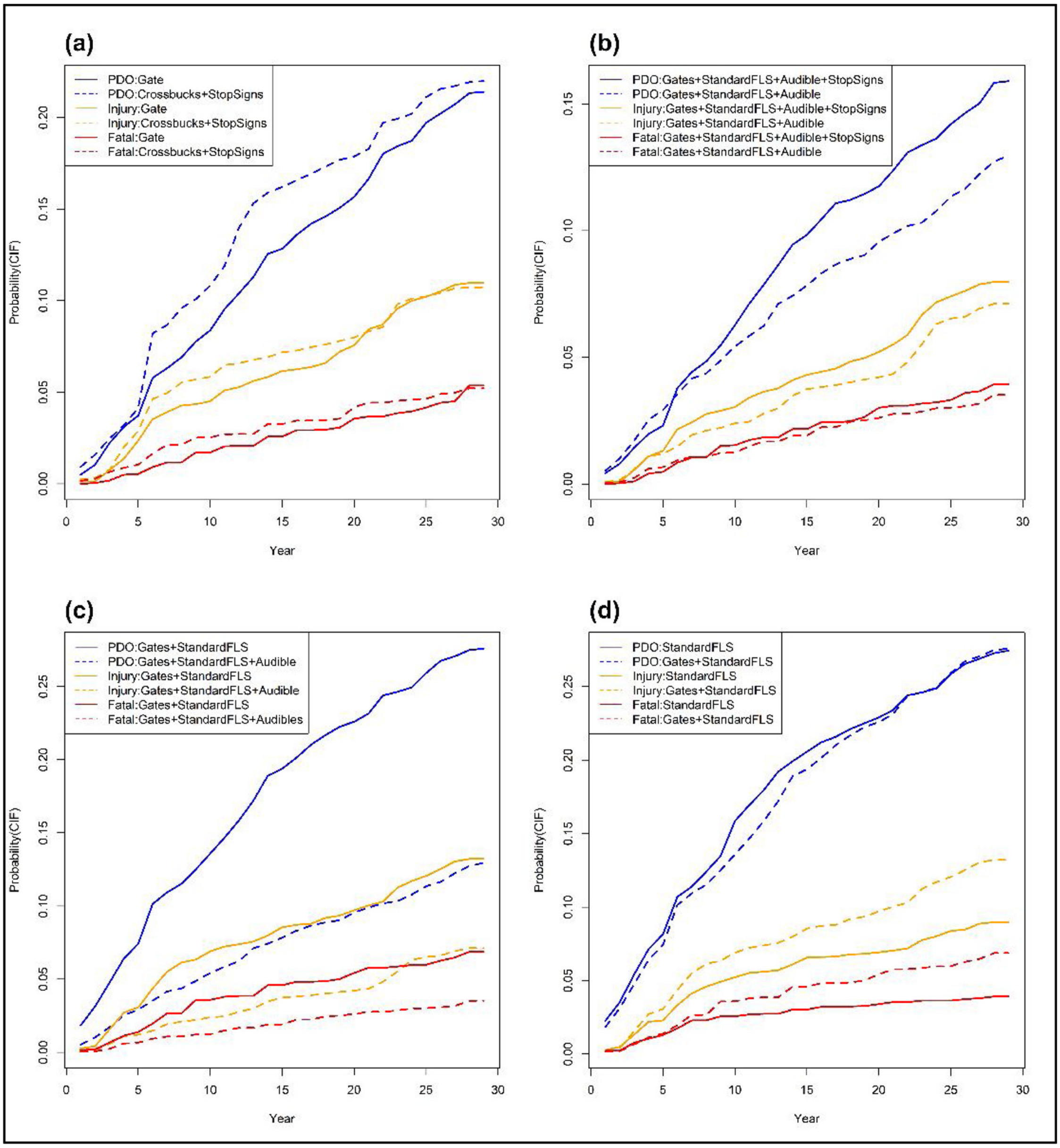

Fig. 3. Cumulative probabilities of crash severity for crossing control pairs.

Table 2
Annual average crash probability increasing percentages for crossing controls.

Rank	Crossing Control	%PDO	%Injury	%Fatal	%Crash
1	Gates + StandardFLS + Audible	0.44	0.25	0.12	0.82
2	Gates + StandardFLS + Audible + StopSigns	0.55	0.28	0.14	0.98
3	Crossbucks + StopSigns	0.75	0.37	0.18	1.31
4	Gates	0.75	0.39	0.19	1.33
5	StandardFLS	0.90	0.31	0.13	1.34
6	Gates + StandardFLS	0.92	0.46	0.24	1.62

event by estimating event-specific variable importance (VIMP).

A few important findings identified with the research:

1) Train detection device, total switching train traffic, total daytime train traffic, maximum train speed, total nighttime train traffic, motorists' speed, crossing controls, highway speed limit, annual average daily traffic (AADT), and functional classification of road are identified as the most predictive contributors based on their VIMP.
2) Although, crossings with gate-only reduce all severity likelihoods compared with the crossings with crossbucks and stop signs, their PDO and injury crash likelihood rates increase faster than for crossings with crossbucks and stop sign for each year.
3) CIF results indicate that by adding stop signs to the combination of gate and standard flashing lights and audible devices, crossings are lower in cumulative PDO and fatal probability for about seven years. However, after year seven, the crossings are more likely to have PDO and fatal crashes. This result indicates the improved ability of the model to predict long-term cumulative probabilities compared with models used in previous studies. Additionally, results showed that injury crash likelihood is 12 % more for crossings after adding the stop sign.
4) Crash occurrence risks and all crash severities increase substantially for crossings with gates and standard flashing lights with no audible devices compared with crossings with the same combination and with an audible device. Adding audible devices to crossings with gates and standard flashing lights causes crossings to have, on

average, 49 %, 52 %, 46 %, and 50 % less probability of crash, PDO, injury, and fatal crashes, respectively, compared with crossings with the same combinations of controls but with no audible device.

5) Crossings with gates and standard flashing lights are less likely to have PDO crash severities compared with the crossings with standard flashing lights for about 23 years, then crossings with both control devices show similar performance. In addition, crossings with standard flashing lights only are less likely to have both injury and fatal crashes compared with crossings with gates and standard flashing lights for almost the entire study time span (specifically after year five).

The RSF approach to competing risks demonstrates its capability to identify risk factors and their marginal probability for crash severity by estimating event-specific VIMP and event-specific CIF. The model provides accurate results to identify and summarize the contributors and risks. Future research may (1) compare other approaches to CRM prediction performance for HRGC safety outputs, and (2) use RSF approach for high-dimensional HRGC data with large numbers of contributors, as it is shown that RSF is less affected by the challenges of dimensionality and has better performance for predicting high-dimensional data compared with other approaches (Ishwaran et al., 2014; Pölsterl et al., 2016).

It is also worth to note that ideally to truly understand countermeasure's marginal effectiveness, controlled before-and-after experiment should be conducted and the effectiveness measurements should be collected and analyzed. However, due to the unavailability of such data, the research conducted in this study is based on empirical data in ND. The uncertainty associated with empirical pilot study should be further investigated before they can be used as guidance for HRGC safety improvement decision making.

Author statement

The authors' confirmed contributions to the paper are as follows:

Study Conception and Design: Dr. Pan Lu, Amin Keramati and Dr. Ying Huang.

Data Collection: Amin Keramati, and Dr. Pan Lu.

Mathematical Modeling Development: Amin Keramati, and Dr. Pan Lu.

Modeling Analysis and Interpretation of Results: Amin Keramati, Dr. Pan Lu, Dr. Amirfarrokh Iranitalab, Dr. Danguang Pan and Dr. Ying Huang.

Draft Manuscript Preparation: Amin Keramati, and Dr. Pan Lu.

Final manuscript writing and results review: Amin Keramati, Dr. Pan Lu, Dr. Amirfarrokh Iranitalab, and Dr. Danguang Pan.

Final manuscript approve: Amin Keramati, Dr. Pan Lu, Dr. Amirfarrokh Iranitalab, Dr. Danguang Pan and Dr. Ying Huang.

Final submission: Dr. Pan Lu, and Amin Karamati.

Revision and re-submissions: Dr. Pan Lu, Amin Karamati, Dr. Amirfarrokh Iranitalab, and Dr. Ying Huang.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgements

The authors express their gratitude to the following funding agencies for their support: North Dakota State University and the Mountain-Plains Consortium (MPC), a University Transportation Center funded by the U.S. Department of Transportation.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.aap.2020.105683.

References

Abdel-Aty, M., Keller, J., 2005. Exploring the overall and specific crash severity levels at signalized intersections. Accid. Anal. Prev. 37 (3), 417–425.

Binder, H., Schumacher, M., 2008. Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. BMC Bioinformatics 9 (1), 14.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. Mon. Weather Rev. 78 (1), 1–3.

Burnham, A., 1995. Stop sign effectiveness at railroad grade crossings (abuse without excuse). In: Third International Symposium on Railroad-Highway Grade Crossing Research and Safety. Southeastern Transportation Center.

Cox, D.R., 1972. Regression models and life-tables. J. R. Stat. Soc. Ser. B 34 (2), 187–202.

Eluru, N., Bagheri, M., Miranda-Moreno, L.F., Fu, L., 2012. A latent class modeling approach for identifying vehicle driver injury severity factors at highway-railway crossings. Accid. Anal. Prev. 47, 119–127.

Fan, J., Feng, Y., Wu, Y., 2010. High-dimensional variable selection for Cox's proportional hazards model. Borrowing Strength: Theory Powering Applications–a Festschrift for Lawrence D. Brown. Institute of Mathematical Statistics, pp. 70–86.

Fan, W., Kane, M.R., Haile, E., 2015. Analyzing severity of vehicle crashes at highway-rail grade crossings: multinomial logit modeling. J. Transp. Res. Forum 39–56.

Fine, J.P., Gray, R.J., 1999. A proportional hazards model for the subdistribution of a competing risk. J. Am. Stat. Assoc. 94 (446), 496–509.

FRA, 2018. All Highway-Rail Incidents at Public and Private Crossings, 1981-2018 [WWW Document]. (Accessed 27 September 19). https://oli.org/about-us/news/collisions-casulties.

Fu, W.J., Carroll, R.J., Wang, S., 2005. Estimating misclassification error with small samples via bootstrap cross-validation. Bioinformatics 21 (9), 1979–1986.

Gerds, T.A., Schumacher, M., 2006. Consistent estimation of the expected Brier score in general survival models with right-censored event times. Biom. J. 48 (6), 1029–1040.

Ghomi, H., Bagheri, M., Fu, L., Miranda-Moreno, L.F., 2016. Analyzing injury severity factors at highway railway grade crossing accidents involving vulnerable road users: A comparative study. Traffic Inj. Prev. 17 (8), 833–841.

Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. J. Am. Stat. Assoc. 102 (477), 359–378.

Graf, E., Schmoor, C., Sauerbrei, W., Schumacher, M., 1999. Assessment and comparison of prognostic classification schemes for survival data. Stat. Med. 18 (17–18), 2529–2545.

Gray, R.J., 1988. A class of $ K $-sample tests for comparing the cumulative incidence of a competing risk. Ann. Stat. 16 (3), 1141–1154.

Haleem, K., Gan, A., 2015. Contributing factors of crash injury severity at public highway-railroad grade crossings in the US. J. Safety Res. 53, 23–29.

Hao, W., Daniel, J.R., 2013. Severity of injuries to motor vehicle drivers at highway–rail grade crossings in the United States. Transp. Res. Rec. 2384 (1), 102–108.

Hao, W., Daniel, J., 2014. Motor vehicle driver injury severity study under various traffic control at highway-rail grade crossings in the United States. J. Safety Res. 51, 41–48.

Hao, W., Daniel, J., 2016. Driver injury severity related to inclement weather at highway–rail grade crossings in the United States. Traffic Inj. Prev. 17 (1), 31–38.

Heydari, S., Fu, L., 2015. Developing safety performance functions for railway grade crossings: a case study of Canada. In: 2015 Joint Rail Conference. American Society of Mechanical Engineers. p. V001T06A017-V001T06A017.

Heydari, S., Fu, L., Thakali, L., Joseph, L., 2018. Benchmarking regions using a heteroskedastic grouped random parameters model with heterogeneity in mean and variance: applications to grade crossing safety analysis. Anal. Methods Accid. Res. 19, 33–48. https://doi.org/10.1016/j.amar.2018.06.003.

Hu, S.-R., Li, C.-S., Lee, C.-K., 2010. Investigation of key factors for accident severity at railroad grade crossings by using a logit model. Saf. Sci. 48 (2), 186–194.

Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random survival forests. Ann. Appl. Stat. 2 (3), 841–860.

Ishwaran, H., Kogalur, U.B., Gorodeski, E.Z., Minn, A.J., Lauer, M.S., 2010. High-dimensional variable selection for survival data. J. Am. Stat. Assoc. 105 (489), 205–217.

Ishwaran, H., Gerds, T.A., Kogalur, U.B., Moore, R.D., Gange, S.J., Lau, B.M., 2014. Random survival forests for competing risks. Biostatistics 15 (4), 757–773. https://doi.org/10.1093/biostatistics/kxu010.

Jeng, O.-J., 2005. Survey of Driver Perceptions of Railroad and Light Rail Warning devices/grade Crossings.

Kang, Y., Khattak, A., 2017. Cluster-based approach to analyzing crash injury severity at highway–rail grade crossings. Transp. Res. Rec. 2608 (1), 58–69.

Keramati, Amin, Lu, Pan, Tolliver, Denver, Wang, Xingju, 2020. Geometric effect analysis of highway-rail grade crossing safety performance. Accid. Anal. Prev. 138 (April), 105470. https://doi.org/10.1016/j.aap.2020.105470.

Lee, J., Nam, D., Moon, D., 2004. A zero-inflated accident frequency model of highway–rail grade crossing. In: Proceedings of the Transportation Research Board Annual Meeting. Washington, DC.

Lenné, M.G., Rudin-Brown, C.M., Navarro, J., Edquist, J., Trotter, M., Tomasevic, N., 2011. Driver behaviour at rail level crossings: responses to flashing lights, traffic signals and stop signs in simulated rural driving. Appl. Ergon. 42 (4), 548–554.

Liu, J., Khattak, A.J., 2017. Gate-violation behavior at highway-rail grade crossings and

the consequences: Using geo-Spatial modeling integrated with path analysis. Accid. Anal. Prev. 109 (May), 99–112. https://doi.org/10.1016/j.aap.2017.10.010.
Liu, J., Khattak, A.J., Richards, S.H., Nambisan, S., 2015. What are the differences in driver injury outcomes at highway-rail grade crossings? Untangling the role of pre-crash behaviors. Accid. Anal. Prev. 85, 157–169. https://doi.org/10.1016/j.aap.2015.09.004.
Lu, P., Tolliver, D., 2016. Accident prediction model for public highway-rail grade crossings. Accid. Anal. Prev. 90, 73–81. https://doi.org/10.1016/j.aap.2016.02.012.
Lu, Pan, Zheng, Zijian, Ren, Yihao, Zhou, Xiaoyi, Keramati, Amin, Tolliver, Denver, Huang, Ying, 2020. A gradient boosting crash prediction approach for highway-rail grade crossing crash analysis. J. Adv. Transp. 2020, 6751728.
Ma, C., Hao, W., Xiang, W., Yan, W., 2018. The impact of aggressive driving behavior on driver-injury severity at highway-rail grade crossings accidents. J. Adv. Transp. 2018.
Meeker, F., Fox, D., Weber, C., 1997. A comparison of driver behavior at railroad grade crossings with two different protection systems. Accid. Anal. Prev. 29 (1), 11–16.
Millegan, H., Yan, X., Richards, S., Han, L., 2009. Evaluation of effectiveness of stop-sign treatment at highway–railroad grade crossings. J. Transp. Saf. Secur 1 (1), 46–60.
Mogensen, U.B., Ishwaran, H., Gerds, T.A., 2012. Evaluating random forests for survival analysis using prediction error curves. J. Stat. Softw. 50 (11), 1.
Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction model for railway-highway interfaces. Accid. Anal. Prev. 38 (2), 346–356. https://doi.org/10.1016/j.aap.2005.10.004.
Pölsterl, S., Conjeti, S., Navab, N., Katouzian, A., 2016. Survival analysis for high-dimensional, heterogeneous medical data: Exploring feature extraction as an alternative to feature selection. Artif. Intell. Med. 72, 1–11.
Putter, H., Fiocco, M., Geskus, R.B., 2007. Tutorial in biostatistics: competing risks and multi-state models. Stat. Med. 26 (11), 2389–2430.
Savolainen, P.T., Mannering, F.L., Lord, D., Quddus, M.A., 2011. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accid. Anal. Prev. 43 (5), 1666–1676.
Winkler, R.L., Murphy, A.H., 1968. Good' probability assessors. J. Appl. Meteorol. Climatol. 7 (5), 751–758.
Yan, X., Richards, S., Su, X., 2010. Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings. Accid. Anal. Prev. 42 (1), 64–74. https://doi.org/10.1016/j.aap.2009.07.003.
Ye, Z., Xu, Y., Lord, D., 2018. Crash data modeling with a generalized estimator. Accid. Anal. Prev. 117, 340–345.
Zhao, S., Khattak, A., 2015. Motor vehicle drivers' injuries in train–motor vehicle crashes. Accid. Anal. Prev. 74, 162–168.
Zhao, S., Iranitalab, A., Khattak, A.J., 2018. A clustering approach to injury severity in pedestrian-train crashes at highway-rail grade crossings. J. Transp. Saf. Secur 9962, 1–18. https://doi.org/10.1080/19439962.2018.1428257.
Zheng, Z., Lu, P., Tolliver, D., 2016. Decision tree approach to accident prediction for highway–rail grade crossings: empirical analysis. Transp. Res. Rec. 2545 (1), 115–122.
Zheng, Z., Lu, P., Lantz, B., 2018. Commercial truck crash injury severity analysis using gradient boosting data mining model. J. Safety Res. 65, 115–124.
Zheng, Z., Lu, P., Pan, D., 2019. Predicting highway–rail grade crossing collision risk by neural network systems. J. Transp. Eng. Part A Syst. 145 (8), 4019033.
Zhou, Xiaoyi, Lu, Pan, Zheng, Zijian, Tolliver, Denver, Keramati, Amin, 2020. Accident prediction accuracy assessment for highway-rail grade crossing using random forest algorithm compared with decision tree. Reliab. Eng. Syst. Saf. 200 (106931).
FRA_Accident, (2020). Accident Data as reported by Railroads [WWW Document], n.d. https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx (Accessed 24 September 19).
FRA_Inventory (2020). Highway-Rail Crossing Inventory Data [WWW Document], n.d. URL https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/DownloadDbf.aspx (Accessed 24 September 19).
ND_Hub, (2020). North Dakota GIS Hub Data Portal [WWW Document], n.d. URL https://www.nd.gov/gis/apps/Download/ (Accessed12 June 19).