WILEY | Hindawi

Research Article

Collective Risk Ranking of Highway Segments on the Basis of Severity-Weighted Crash Rates

Dawei Li [ID] **and Mustafa F. M. Al-Mahamda** [ID]

School of Transportation, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Mustafa F. M. Al-Mahamda; 223185614@seu.edu.cn

This study is intended to focus on the major factors affecting traffic crash rates and severity levels, in addition to identifying crash-prone locations (i.e., black spots) based on the two indicators. The available crash data for different road segments used for the analysis were obtained from the Washington state database provided by the Highway Safety Information System (HSIS) for the years 2006 to 2011. A Random Forest (RF) classifier was used to predict the outcome level of crash severity, while crash rates were predicted by applying RF regressor. Certain features were selected for each model besides the abstraction of new features to check if there are unobserved correlations affecting the independent variables, such as accounting for the number and weight of crashes within $1\,km^2$ area by implementing the Getis-Ord G_i* index. Moreover, to calculate the collective risk (CR) score, crash rates were adjusted to incorporate crash severity weights (cost per severity type) and regression-to-the-mean (RTM) bias via Empirical Bayes (EB) method. Finally, segments were ranked according to their CR score.

1. Introduction

In recent years, crash rate and crash severity were the two major indicators used to assess roadway's traffic safety, as well as identifying crash-prone locations. Reducing crash rates and severity requires different strategic approaches and policies that aim at reducing the exposure for traffic accidents. Both reliable historical crash data and well-developed prediction models are essential in the process of estimating the impact of human and environmental-based features on the frequency and outcome of a crash incident. A weighted risk index or a combination of crash frequency and severity should prove an advantage over having only crash rate as an index of traffic hazards, without any regard to severity level or vice versa (e.g., having an urban road segment with high crash rate but lower levels of crash severity, and the opposite on a rural highway segment). Predictions should also account for fluctuations and trends in crash rates. Furthermore, unobserved features and correlations (e.g., spatial correlations between adjacent segments that can amplify a hazardous impact) should be considered. Observed or predicted indicators can be utilized to rank road facilities according to their overall risk score, allowing traffic safety agencies to distinguish which facilities have the priority in future crash countermeasure policies.

This article hereby introduces a collective risk (CR) score that combines both crash rate and severity into a single index, and in the same time, gives an account of spatial correlations between adjacent road segments using the Getis-Ord G_i* statistic and provides control over bias in predictions of hazardous segments using the Empirical Bayes (EB) method. So as to illustrate the process mentioned above, Section 2 provides a brief review of related literature, while Section 3 presents the implemented methodology to process the given data, Section 4 lists the results with discussion, and finally, Section 5 concludes the findings of this article.

2. Review of Literature

With regard to crash data, many researchers focused on different factors affecting the estimation of crash rates and

crash severity. Aside from the observed features, there are other points that should be taken into consideration when analysing crash data, these factors include (a) under-reporting and source of data, as police reports count about 50% of actual property-damage-only (PDO) crashes; (b) ordinal nature of severity data and the correlation among neighbouring severity levels; (c) fixed parameters that restrict the effects of explanatory variables to be the same across individual observations, ignoring unobserved heterogeneity such as in risk-taking behavior; (d) fixed severity thresholds that lead to biased estimations of features (e.g., airbags implementation), affecting the likelihood of certain severity levels due to a constrained shift in thresholds; (e) endogeneity bias that occurs when explanatory variables (e.g., the amount and frequency of warning signs) are affected by the dependent variable (e.g., the crash rate); (f) within-crash and crash type correlations (e.g., number of crash-involved personnel, collision angle, cause, and responsibility, etc.) [1–3].

In addition to the combined effect of crash frequency and severity, spatial and temporal correlations between crash observations were also a major focus. Ma and Kockelman [4] performed a study on speed limit's effect on traffic safety by clustering homogenous highway segments from the state of Washington into crash counts by severity. Random-effects linear model and ordered probit regression were used for estimating crash rates and severity, respectively. On the other hand, Soltani and Askari [5] used spatial autocorrelation methods (Moran's I and Getis-Ord G_i* index) for clustering hotspots in Iran. Clusters were compared in three attributes; time of the crash, severity index, and location. In comparison, Bao et al. [6] utilized the RF technique to identify factors affecting crash severity levels with the purpose of studying unobserved travel patterns and spatial correlations related to crash rate distribution for large-scale taxi GPS data from the city of New York.

Efforts were also made to introduce a combined index used for ranking potential hotspots, providing a reference to impose safety countermeasures. Da Costa et al. [7] incorporated crash severity in a collective risk ranking of road segments in Australia. The severity-weighted frequencies per crash type (based on crash severity costs in Australia) were adjusted for a potential RTM bias via the EB method. Similarly, Afghari et al. [8] employed the EB method to predict hotspots in Australia besides a joint risk model of both crash rate and severity, stating that the latter has an advantage over traditional count models.

Table 1 lists additional articles related to crash risk analysis.

3. Dataset and Methodology

The dataset from the HSIS database (years 2006 to 2011), which covers 38 counties of Washington state, contains accident data for each year, including crash scene characteristics and conditions. Different data logs were merged using a Visual Basic for Applications (VBA) automated MS Excel worksheet. The data were then split into two groups, the first group for building the models (years 2006 to 2009) and the other was used as a testing sample (years 2010 and 2011). Crash data (including time, weather, and severity level) referenced a single point on the roadway and only account for the most severe observation, while crash scene characteristics were collected from the homogenous segment of the roadway where the crash occurred [18]. Data went through a preprocessing stage which involved new categorization based on ordinal ranking for each feature, in addition to the elimination of redundant features and observations, and finally adding new features (such as hotspots).

3.1. Hotspots. To verify whether a crash scene or a road segment is a part of a larger cluster of hazardous locations, the Getis-Ord statistic G_i* was calculated. Additional data for milepost coordinates were obtained from the Washington State Department of Transportation (WSDOT) geospatial database [19]. Using crash scene milepost and the related coordinates of that milepost, and by implementing a VBA automated Excel worksheet, all crash observations from the years 2006 to 2009 were incorporated in the G_i* statistic calculation process. The collected crash data categorized severity into 8 classes. However, to conform with the crash cost data for the state of Washington [20], crash severity categories were transformed into 4 severity classes based on the KABCO ranking as shown in Table 2.

In order to incorporate crash counts into the calculation process, all crash observations were weighted according to their cost value ($C = \text{Cost}/\text{Cost}_{\text{PDO}}$). The study area (about 580 km × 380 km) was divided into 1 km^2 cells and each crash observation was assigned to its cell centroid. The G family of statistics can be used as measures of spatial correlations in many aspects, yet G_i and G_i* (i.e., local statistics) can detect pockets of spatial links within the global study area. In this article, the standardized local statistic G_i* was chosen over the other global and local statistics for the sake of both simplicity and efficiency in finding local clusters since it can detect correlations within the same local area (e.g., in the case of adjacent segments or an intersection that are located within the same grid unit i).

$$G_i^* = \frac{\sum_{j=1}^{n} w(d_{ij}) x_j - \left(\sum_{j=1}^{n} x_j / n\right) \sum_{j=1}^{n} w(d_{ij})}{\sqrt{\left(\sum_{j=1}^{n} x_j^2 - \sum_{j=1}^{n} x_j\right)/n} - \sqrt{\left(n \sum_{j=1}^{n} w(d_{ij})^2 - \left[\sum_{j=1}^{n} w(d_{ij})\right]^2\right)/(n-1)}}, \tag{1}$$

TABLE 1: Highlights on Crash Risk Analysis (in chronological order).

Author	Domain	Methodology
Chiou and fu [9]	Joint models	Multinomial-generalized Poisson joint model with error components.
Sugiyanto et al. [10]	Collective risk	Upper Control Limit (UCL) for hotspots with equivalent accident number.
Ouni and Belloumi [11]	Risk factors	Multinomial logit model to identify factors affecting severity. Ripley's K-function and kernel density estimation (KDE) for vulnerable road users (VRUs) clusters.
Xu et al. [12]	Risk factors crash severity	A sequential logistic regression model with random parameters for multiple collision angles.
Feng et al. [13]	Risk factors	Real-time experiment on the psychological and behavioral changes of drivers in longitudinal underpass segments.
Wen et al. [14]	Spatial correlation crash rate	A Poisson-based model with spatial conditional autoregressive priors (CAR) for crashes within adjacent road segments.
Zeng et al. [15]	Spatial correlation crash severity	Bayesian generalized ordered logit model with CAR prior spatial term, fixed parameters, and flexible thresholds.
Ulak et al. [16]	Spatial correlation crash rate	Getis-Ord G_i* and local Moran's I distance-based spatial weights with crash prediction accuracy index (CPAI).
Wang et al. [17]	Risk factors	A questionnaire-based study on risk perception among cyclists, and the major influencing factors in the prediction of risky behaviors.

TABLE 2: Crash severity levels and costs in Washington state, 2010.

Category	K/A	B	C	O
Severity level	Fatal/Incapacitating injury	Non-incapacitating injury	Possible injury	No Injury/PDO
Cost (USD)	2,900,000	155,000	60,000	10,000

where $w(d_{ij})$ is the distance weight function between the centroid of interest i and the adjacent centroid j, and x_j is the weighted value of crashes in j (i.e., $\sum C_j$), while n is the total number of cells (or centroids) including centroid i [21]. The distance was assigned to be a 1.5 km radius around centroid i to reach centroids in corner cells (Queen neighborhoods), while the distance function $w = \text{INT}[1/(\text{INT}(0.67 \times d_{ij}) + 1)]$ produces only two outputs, 1 for $d < 1.5$ (including $d_{ii} = 0$) and 0 for $d \geq 1.5$. The G_i^* value is already a z-statistic and can be directly utilized to identify clusters within the grid according to the p-value of the cell. Figure 1 shows the clusters within the study area as cells with $p < 0.01$ were considered significant crash-prone areas [22]. A hotspot class was given for each crash observation based on their clusters. Hotspot classes were incorporated in the modeling process as an explanatory variable.

3.2. Crash Severity. Out of 183,400 training data points (years 2006 to 2009) and 85,160 testing observations (years 2010 and 2011), only 157,321 and 43,623 data points were selected, respectively. Preprocessing included removing data points with missing data, as well as the elimination of features with low variance, Spearman Correlation ranking, and recursive feature elimination and cross-validated selection (RFECV). Figure 2 provides a description of the top-ranked features with regard to their importance to crash severity prediction.

Two of the most popular machine learning algorithms supported by Python were used to fit the observed crash severity levels and the related crash features.

3.2.1. Random Forests. Random forest is an ensemble of K tree-structured estimators (h_k) that split according to the values of a random vector Θ_k sampled independently and with the same distribution for all trees in the forest. Classification accuracy results from growing the trees and collecting their votes for the most popular class y at input x. A margin function $mg(xy)$ measures the extent to which the average number of votes at x, y for the right class exceeds the average vote for any other class. The larger the margin, the more confidence in the classification.

$$mg(x, y) = avg_k[I(h_k(x) = y)] - \max_{j \neq y} avg_k[I(h_k(x) = j)], \quad (2)$$

$$PE^* = P_{x,y}(mg(x, y) < 0), \quad (3)$$

where $I(\cdot)$ is the indicator function and PE^* is the generalization error [23]. Nonetheless, given the convenience of the Scikit-learn ensemble package [24], a random forest of 1,000 estimators was used to classify the severity levels.

3.2.2. Ordered Logit. Discrete choice models, such as the multinomial logit and the mixed logit, were used extensively in the transportation domain (e.g., route choice and travel behavior) [25, 26]. However, as mentioned in Section 2, the nature of severity data requires an ordinal classifier. Based on the application of the generalized linear models (GLM), the ordered logit model (OLM) scales the response values into J ordered classes, and each response z_j is confined within θ_j thresholds ($\theta_{j-1} < z_j < \theta_j$). To estimate the optimal slope
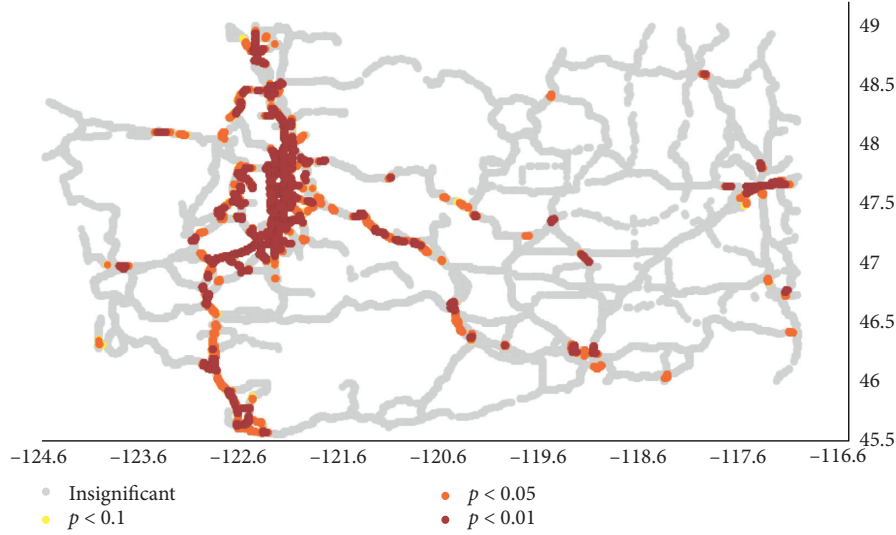
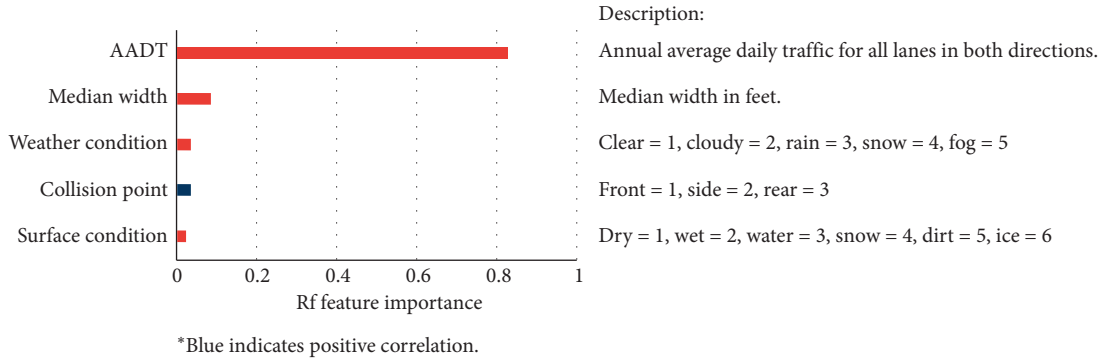FIGURE 1: Clusters within the study area (Washington state).

FIGURE 2: Severity feature importance.

vector parameters w_j and thresholds θ_j, both prediction errors and the sum of all penalties (for crossing each threshold), given by the all-thresholds loss function, should be minimized in all classes.

$$loss(y, z) = \sum_{j=1}^{J-1} f\big(s(j, y)(\theta_j - z)\big), \tag{4}$$

where $f(\cdot)$ is a margin penalty function, y is the predicted range, and $s(j, y)$ is positive if $j \geq y$ and negative otherwise. The data can be fitted using the posterior probability $P(y \leq j \mid x)$, and the cumulative probability of the ordinal class follows a logistic function [27].

$$P(y \leq j | x) = \frac{1}{1 + \exp\big(w^T x - \theta_j\big)}. \tag{5}$$

This procedure is implemented in the MORD library under the all-threshold variant classifier [28].

3.3. Crash Rate.
The 157,321 observations from the years 2006 to 2009 and the 43,623 observations from the years 2010 and 2011 (i.e., data used in Section 3.2.) went through

further preprocessing to be implemented in the crash rate prediction process. Crash observations were clustered for each homogenous road segment, with a length of no less than 0.16 km for convenience [29]. Since the HSIS dataset only provides the most severe case for a single crash observation, grouping data into segments for each year makes temporal variables (e.g., weather conditions) redundant. 19,490 segments from the years 2006 to 2009 and 7,321 from 2010 to 2011 were selected for training and testing the model. The features were filtered by standardized coefficients and Spearman Correlation rankings, as well as the embedded RFECV method (See Figure 3).

3.3.1. Random Forest Regressor.
Similar to RF classifiers, RF regressors are formed by growing trees depending on the random vector Θ. The output of the tree predictors h_k is numerical and the training set is independently drawn from the distribution of the random vector y, x. An accurate regression forest has a low correlation between residuals and low error trees. The mean squared generalization error for the forest $E_{x,y}(y - avg_k h(x, \Theta_k))^2$ is formed by taking the average random predictor over k trees [23]. Scikit-learn's regressor was used with 1,000 estimators.
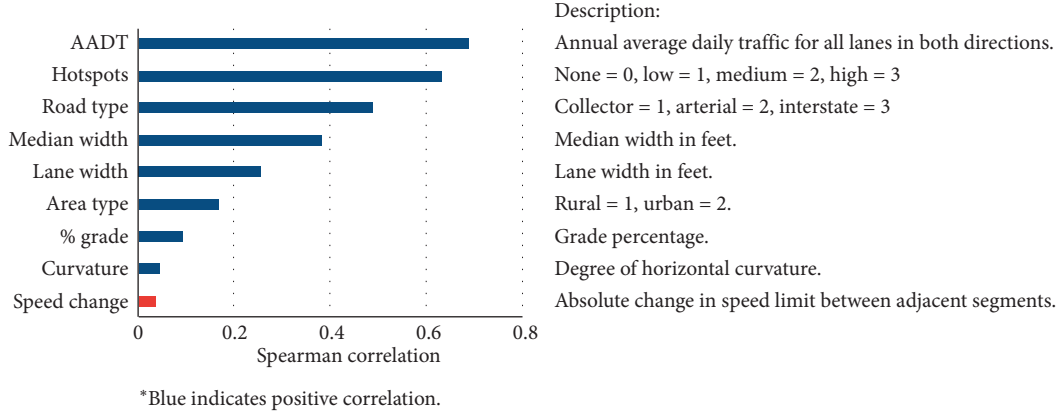
FIGURE 3: Crash rate feature importance.

3.3.2. Bayesian Ridge Regression. Bayesian inference techniques are considered popular machine learning tools in the transportation field (e.g., intelligent transportation systems) [30]. Ridge regression is a form of linear regression that penalizes the coefficients w with the ridge coefficient $\alpha \geq 0$, hence minimizing the residual sum of squares.

$$\min_{w} \|xw - y\|_2^2 + \alpha \|w\|_2^2, \tag{6}$$

The larger the value of α, the greater the amount of shrinkage. Similarly, the Bayesian Ridge (BR) includes regularization parameters during the estimation procedure, where w has a prior $p(w|\lambda)$ given by a spherical Gaussian $N(w|0, \lambda^{-1}I_p)$. The numerical output y of the probabilistic model $p(y|x, w, \alpha) = N(y|xw, \alpha)$ is assumed to follow a Gaussian distribution and both priors for α and λ are randomly chosen from gamma distributions. w, α, and λ are estimated during the fit of the model by maximizing the log marginal likelihood [24]. Scikit-learn's BayesianRidge was used with default initial parameters.

3.4. Empirical Bayes. Traffic crashes are random events, thus predicting the rates have some limitations due to the inherent characteristics of the data itself, regardless of the methods used to collect it. These limitations can introduce bias and affect crash data reliability, especially in the long term. Statistically, a low crash rate period will likely be followed by a high rate period, and vice versa. This tendency is known as regression-to-the-mean (RTM). Applying short-term predictions for longer periods can cause a bias called RTM bias (See Figure 4).

The EB method is used to estimate the expected average crash rate N_E of an individual site (road segment or intersection) for a given time period T (in years) on the condition of unchanged site features (e.g., geometric design). The estimate relies on predictive models N_P (modified for similar characteristics or control function) combined with observed crash rates N_O.
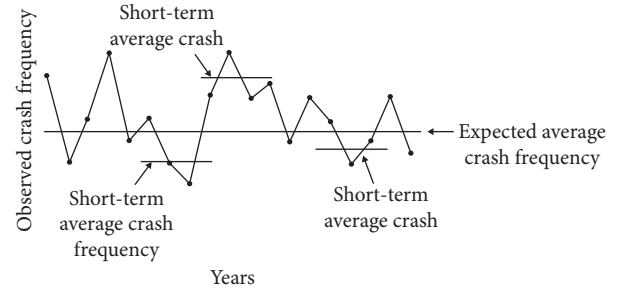
$$N_E = wN_P + (1 - w)N_O. \tag{7}$$

FIGURE 4: Rtm effect on crash rate prediction [29].

An overdispersion parameter (i.e., a variance of classification) is needed to provide an indication of the statistical reliability of N_P. The lower is the overdispersion parameter, the more statistically reliable is the model. This parameter is used in the EB method to provide the weights w and $1 - w$ to N_P and N_O, respectively [29]. As N_P increases, N_E becomes more dependable on the N_O value.

$$w = \frac{N_P}{N_P + V}. \tag{8}$$

Since the crash rate model already incorporated various variables (e.g., road features) in the fitting stage, in addition for providing more simplicity, N_P values for the study period (i.e., years 2006 to 2009) are taken directly from the model output without further modifications regarding these features. Aiming to give more importance to segments with low rates but high severity levels (see Figure 5), both N_P and N_O should be converted to severity-weighted rates S_P and S_O, respectively.

$$S_{it} = M_{it} \sum_{j=1}^{J} P_{itj} C_j, \tag{9}$$

where P_{itj} is the probability of crash severity j occurring on segment i in year t, C_j is the cost for severity level j, and M_{it} is the total number of crash predictions or observations in segment i and year t. When S_P is calculated, P is taken directly from the classifier probability predictions, while for S_O, P is simply taken as the ratio m_j/N_O.
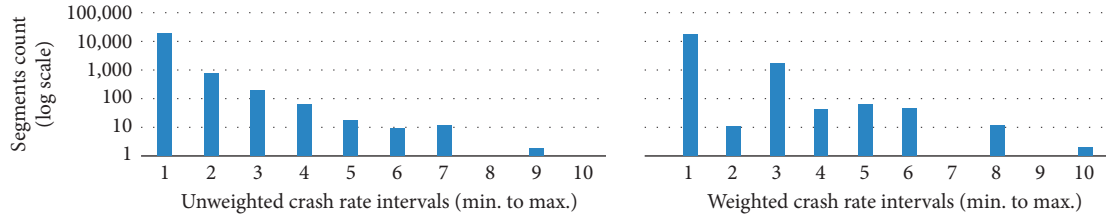
FIGURE 5: Distribution of severity-weighted and unweighted observed crash rates (2006 to 2009).

TABLE 3: F1-score for models and each severity class.

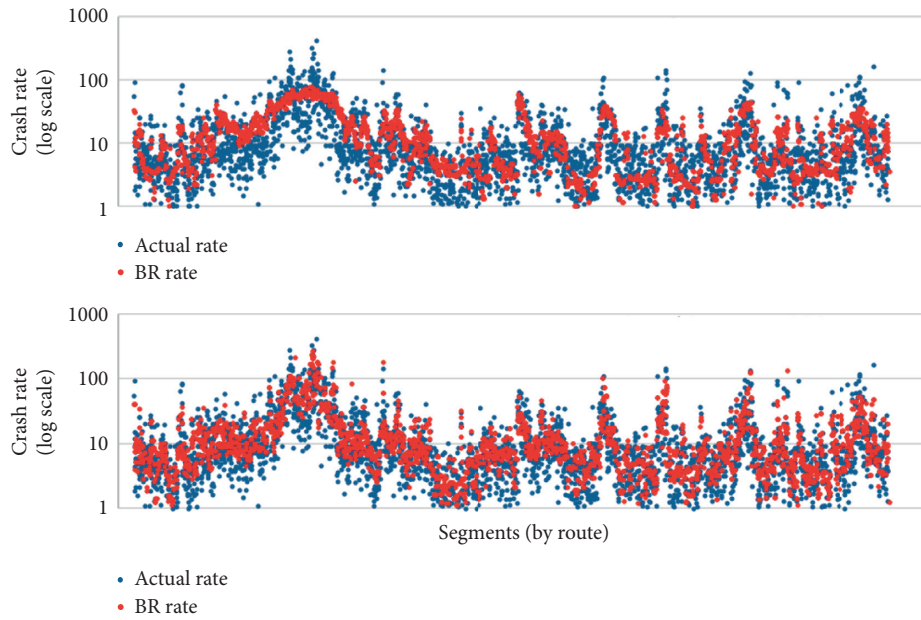| Dataset	With random features				Without random features			
	Random forest		Ordered logit		Random forest		Ordered logit	
	Class	Model	Class	Model	Class	Model	Class	Model
No Injury/PDO	0.72		0.79		0.77		0.80	
Possible injury	0.19		0.06		0.09		0.00	
Non-incapacitating injury	0.12	0.54	0.00	0.55	0.07	0.54	0.00	0.54
Fatal/Incapacitating injury	0.06		0.00		0.02		0.00	

FIGURE 6: Crash rate models performance (2011 predictions).

Finally, N_E is modified to estimate the collective risk (CR) by predicting the weighted crash rate S_F for future year t^* (i.e., the year 2010 or 2011) [7].

$$w = \frac{\sum_{t=1}^{T} S_{Pit}}{\sum_{t=1}^{T} S_{Pit} + \sum_{t=1}^{T} (S_{Pit} - S_{Oit})^2/T}, \tag{10}$$

$$CR_{it^*} = \frac{S_{Fit^*}}{\sum_{t=1}^{T} S_{Pit}} \left[w \sum_{t=1}^{T} S_{Pit} + (1-w) \sum_{t=1}^{T} S_{Oit} \right]. \tag{11}$$
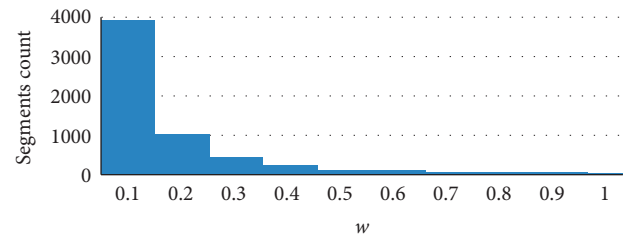

FIGURE 7: EB weight (w) distribution.

TABLE 4: Segment ranking according to their CR values.

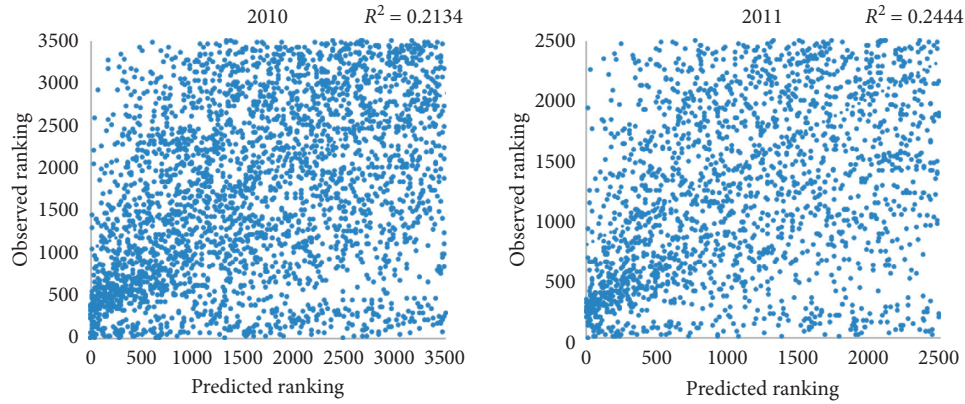| 2010			2011		
Segment	Predicted CR rank	Observed CR rank	Segment	Predicted CR rank	Observed CR rank
5–162.91	1	237	5–162.91	1	317
5–172.85	2	279	5–166.52	2	208
5–163.43	3	2	5–161.33	3	219
5–164.43	4	5	5–153.86	4	293
5–148.23	5	392	5–164.43	5	172

FIGURE 8: Correlation of predicted and observed CR ranking.

4. Results and Discussion

Crash severity was found to be correlated to some unobserved random features such as weather conditions and collision points. These features were needed to be drawn from a suitable distribution to be used in the model's testing stage. Both weather and road surface conditions follow a Gaussian distribution, while the collision point follows a Beta distribution (with $\alpha = 0.2$ and $\beta = 0.4$). AADT was taken as an average from the years 2006 to 2009. Crash severity results are summarized in Table 3.

It can be seen that both methods could not produce accurate results. Furthermore, predictions were biased towards no Injury/PDO severity class (largest number of observations). However, the RF method generated more correct predictions in the higher severity classes; therefore, class probabilities shall be taken from the RF output. Incorporating the random features did not affect the overall performance of the model; hence, only AADT and median width can be used to predict severity class. Figure 6 illustrates the results of crash rate models.

RF regressor gave more accurate predictions for crash rates, while the BR model had underdispersed results. Thus, RF predictions are selected to be used as N_P and N_F values, respectively. EB weights for S_P indicate that predicted values have less importance to the estimated crash rate (see Figure 7).

Since not all segments had crash observations during both periods (i.e., years 2006–2009 and 2010–2011), only 3,649 and 2,647 segments from the years 2010 and 2011, respectively, were ranked according to their CR values. The top 5 ranked segments are listed in Table 4.

CR ranking indicates a local cluster on route 5 in the city of Seattle. Nevertheless, only 9 segments were included in the top 100 ranks of both actual and predicted rankings. The predictions of the year 2011 showed a slight improvement in accuracy for the top 1000 segments (see Figure 8). This can be understandable since the EB method is more effective in long-term estimates rather than the short-term. Severity weights also have a role in some bias towards crash clusters.

5. Conclusion

Machine learning algorithms introduce simple alternatives to the conventional regression models by providing both reasonable estimations and a user-friendly interface. In other words, it does not require comprehensive statistical knowledge. Random Forest accuracy for regression and classification outperformed traditional models (i.e., linear and logistic estimators). Yet, a multistage algorithm that combines RF and other models might produce better results and should be further investigated. Annual average daily traffic was a major factor in predicting both crash severity and frequency among some road features. On the other hand, random and temporal variables had a very small impact. The Empirical Bayes method demonstrated significant control over predicted crash rates, keeping it in a reasonable range. Although CR rank predictions were relatively poor in terms of reliability, having a collective risk estimate exhibits more potential to identify hotspots or localized clusters of safety hazards, thanks to the combination of frequency and severity of crashes. Attention should be paid to severity weights (i.e., considering an alternative to

the cost index C) and the screening method (e.g., using road segments instead of grid cells) to avoid allocating large weights on outliers, especially when reported data points are related to the most severe case and there is only one observation per segment.

Data Availability

The traffic crash datasets used to support the findings of this study were supplied by HSIS. Requests for access to these data for research and educational purposes should be made to HSIS staff at https://www.hsisinfo.org/datarequest.cfm. Other types of data are cited at relevant places within the text as references.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] D. Lord and F. Mannering, "The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives," *Transportation Research Part A: Policy and Practice*, vol. 44, no. 5, pp. 291–305, 2010.

[2] P. T. Savolainen, F. L. Mannering, D. Lord, and M. A. Quddus, "The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives," *Accident Analysis & Prevention*, vol. 43, no. 5, pp. 1666–1676, 2011.

[3] D. Li, Y. Song, and Q. Chen, "Bilevel programming for traffic signal coordinated control considering pedestrian crossing," *Journal of Advanced Transportation*, vol. 2020, Article ID 3987591, 18 pages, 2020.

[4] J. Ma and K. Kockelman, "Crash frequency and severity modeling using clustered data from Washington state," in *Proceedings of the IEEE Intelligent Transportation Systems Conference*, pp. 1621–1626, IEEE, Toronto, Canada, September 2006.

[5] A. Soltani and S. Askari, "Exploring spatial autocorrelation of traffic crashes based on severity," *Injury*, vol. 48, no. 3, pp. 637–647, 2017.

[6] J. Bao, P. Liu, X. Qin, and H. Zhou, "Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data," *Accident Analysis & Prevention*, vol. 120, pp. 281–294, 2018.

[7] S. Da Costa, X. Qu, and P. M. Parajuli, "A crash severity-based black spot identification model," *Journal of Transportation Safety & Security*, vol. 7, no. 3, pp. 268–277, 2015.

[8] A. P. Afghari, M. Haque, and S. Washington, "Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes," *Accident Analysis & Prevention*, vol. 144, Article ID 105615, 2020.

[9] Y.-C. Chiou and C. Fu, "Modeling crash frequency and severity using multinomial-generalized poisson model with error components," *Accident Analysis & Prevention*, vol. 50, pp. 73–82, 2013.

[10] G. Sugiyanto, A. Fadli, and M. Y. Santi, "Identification of black spot and equivalent accident number using upper control limit method," *ARPN Journal of Engineering and Applied Sciences*, vol. 12, no. 2, pp. 528–535, 2017.

[11] F. Ouni and M. Belloumi, "Spatio-temporal pattern of vulnerable road user's collisions hot spots and related risk factors for injury severity in tunisia," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 56, pp. 477–495, 2018.

[12] C. Xu, Y. Wang, P. Liu, W. Wang, and J. Bao, "Quantitative risk assessment of freeway crash casualty using high-resolution traffic data," *Reliability Engineering & System Safety*, vol. 169, pp. 299–311, 2018.

[13] Z. Feng, M. Yang, W. Zhang, Y. Du, and H. Bai, "Effect of longitudinal slope of urban underpass tunnels on drivers' heart rate and speed: a study based on a real vehicle experiment," *Tunnelling and Underground Space Technology*, vol. 81, pp. 525–533, 2018.

[14] H. Wen, X. Zhang, Q. Zeng, J. Lee, and Q. Yuan, "Investigating spatial autocorrelation and spillover effects in freeway crash-frequency data," *International Journal of Environmental Research and Public Health*, vol. 16, no. 2, Article ID 219, 2019.

[15] Q. Zeng, W. Gu, X. Zhang, H. Wen, J. Lee, and W. Hao, "Analyzing freeway crash severity using a bayesian spatial generalized ordered logit model with conditional autoregressive priors," *Accident Analysis & Prevention*, vol. 127, pp. 87–95, 2019.

[16] M. B. Ulak, E. E. Ozguven, O. A. Vanli, and M. W. Horner, "Exploring alternative spatial weights to detect crash hotspots," *Computers, Environment and Urban Systems*, vol. 78, Article ID 101398, 2019.

[17] C. Wang, W. Zhang, Z. Feng, K. Wang, and Y. Gao, "Exploring factors influencing the risky cycling behaviors of young cyclists aged 15-24 Years: a questionnaire-based study in China," *Risk Analysis*, vol. 40, no. 8, pp. 1554–1570, 2020.

[18] A. P. Nujjetty, Y. M. Mohamedshah, and F. M. Council, "Guidebook for the washington state data files," highway safety information system," 2014, https://www.hsisinfo.org/pdf/guidebook_WA.pdf.

[19] Washington State Department of Transportation (WSDOT), "Wsdot - state route milepost markers," geospatial open data portal," 2019, https://gisdata-wsdot.opendata.arcgis.com/datasets/wsdot-state-route-milepost-markers.

[20] T. Harmon, G. Bahar, and F. Gross, *Crash Costs for Highway Safety Analysis*, USDOT FHWA Office, Washington, DC, USA, 2018.

[21] J. K. Ord and A. Getis, "Local spatial autocorrelation statistics: distributional issues and an application," *Geographical Analysis*, vol. 27, no. 4, pp. 286–306, 1995.

[22] M. Lee and A. J. Khattak, "Case study of crash severity spatial pattern identification in hot spot analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, no. 9, pp. 684–695, 2019.

[23] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[24] Scikit-learn Developers, "Scikit-learn user guide: release 0.23.1," 2020, https://scikit-learn.org/stable//_downloads/scikit-learn-docs.pdf.

[25] D. Li, T. Miwa, T. Morikawa, and P. Liu, "Incorporating observed and unobserved heterogeneity in route choice analysis with sampled choice sets," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 31–46, 2016.

[26] D. Li, T. Miwa, C. Xu, and Z. Li, "Non-linear fixed and multi-level random effects of origin-destination specific attributes on route choice behaviour," *IET Intelligent Transport Systems*, vol. 13, no. 4, pp. 654–660, 2019.

[27] J. D. M. Rennie and N. Srebro, "Loss functions for preference levels: regression with discrete ordered labels," in *Proceedings of the IJCAI Multidisciplinary Workshop on Advances in Preference Handling*, Edinburgh, Scotland, August 2005.

[28] F. Pedregosa-Izquierdo, *Feature Extraction and Supervised Learning on FMRI: From Practice to Theory,"* Medical Imaging, Université Pierre et Marie Curie, Paris, France, 2015.

[29] American Association of State Highway Transportation Officials (AASHTO), *Highway Safety Manual*, American Association of State Highway Transportation Officials, Washington, DC, USA, 2010.

[30] D. Li, M. T. Miwa, and T. Morikawa, "Modeling time-of-day car use behavior: a bayesian network approach," *Transportation Research Part D: Transport and Environment*, vol. 47, pp. 54–66, 2016.