Equitable Traffic Crash Prediction Framework To Support Safety Improvement Grants Allocation

By Zihang Wei, Zihao Li, Mihir Mandar Kulkarni, and Ximin Yue

On December 16, 2021, the National Highway Traffic Safety Administration (NHTSA) announced the release of nearly $260 million in highway safety grants as part of the Bipartisan Infrastructure Law.[1] These grants aim to enhance traffic safety across all 50 states. Due to limited funding, it is crucial to allocate these grants optimally for maximum social benefit.[2] However, current approaches to grant allocation lack clear criteria and primarily rely on previous crash data and available funding.[3]

Shutterstock.com/William Potter

HSIS First Place Safety Data Award

This is the first-place winning paper of the Federal Highway Administration's (FHWA) 2023 Excellence in Highway Safety Data Award, which is designed to encourage university students to use Highway Safety Information System (HSIS) data to investigate a topic that advances highway safety and to develop a paper to document the original research. The HSIS Highway Safety Data Awards Program is jointly administered by FHWA and ITE. This paper has been edited for length from the original version.

For future allocation, an accurate crash prediction model is essential as a reliable decision-support tool. Previous studies have explored various models, including safety performance functions, Bayesian multivariate poison lognormal models, loglinear regression models, and negative binomial models to predict crash occurrences and identify candidate sites for safety improvement.[4-8] Recently, with advancements in AI algorithms and increased number of crash datasets, machine learning and deep learning algorithms have been applied for accurate crash prediction and safety improvement planning.[9, 10] However, these models can lead to inequality issues, as disadvantaged groups may experience lower prediction performance compared to advantaged groups.[11, 12] Directly applying imbalanced models for grant allocation may exacerbate social inequality.[13] Few studies have focused on model equality in traffic safety and grant allocation. Therefore, this study proposes an equitable traffic crash prediction framework to support grant allocation, ensuring equal performance for all groups.[14]

In this study, we quantitatively evaluate the equality of a traffic crash prediction model based on North Carolina census tracts. The study utilizes crash data from the Highway Safety Information System (HSIS) and sociodemographic data from the U.S. Census Bureau. Ideally, besides improved overall prediction performance, the traffic crash prediction model should perform equally well across different census tract groups, such as high-income and low-income areas. Statistically, the distribution of model prediction errors for these sensitive groups should be similar, indicating unbiased performance. In order to addressing the inequality issue in safety grant allocation, an equitable traffic crash prediction framework is proposed in this study. The results show that by incorporating the Synthetic Minority Oversampling Technique (SMOTE) with the attentive interpretable (TabNet) model, the proposed framework improves both model equality and overall prediction performance.[15, 16] The SMOTE is applied to resolve the training data imbalance issue. It is important to note that this framework does not generate the safety grant allocation plan directly, but it serves as an equitable and effective decision-support tool for allocation.

Data and Methodology

Data Preparation

This section presents the study's framework and data preparation process. Figure 1 depicts the proposed equitable traffic crash prediction framework. The input data comprises two parts: (1) crash data from the highway safety information system (HSIS), which includes accident, occupant, roadway, and vehicle information; and (2) sociodemographic data from the U.S. Census Bureau. The variables are selected and processed to create census tract level data for the crash prediction model's input (yellow block in Figure 1). The goal of the crash prediction models is to forecast the number of severe crashes in the upcoming year based solely on previous

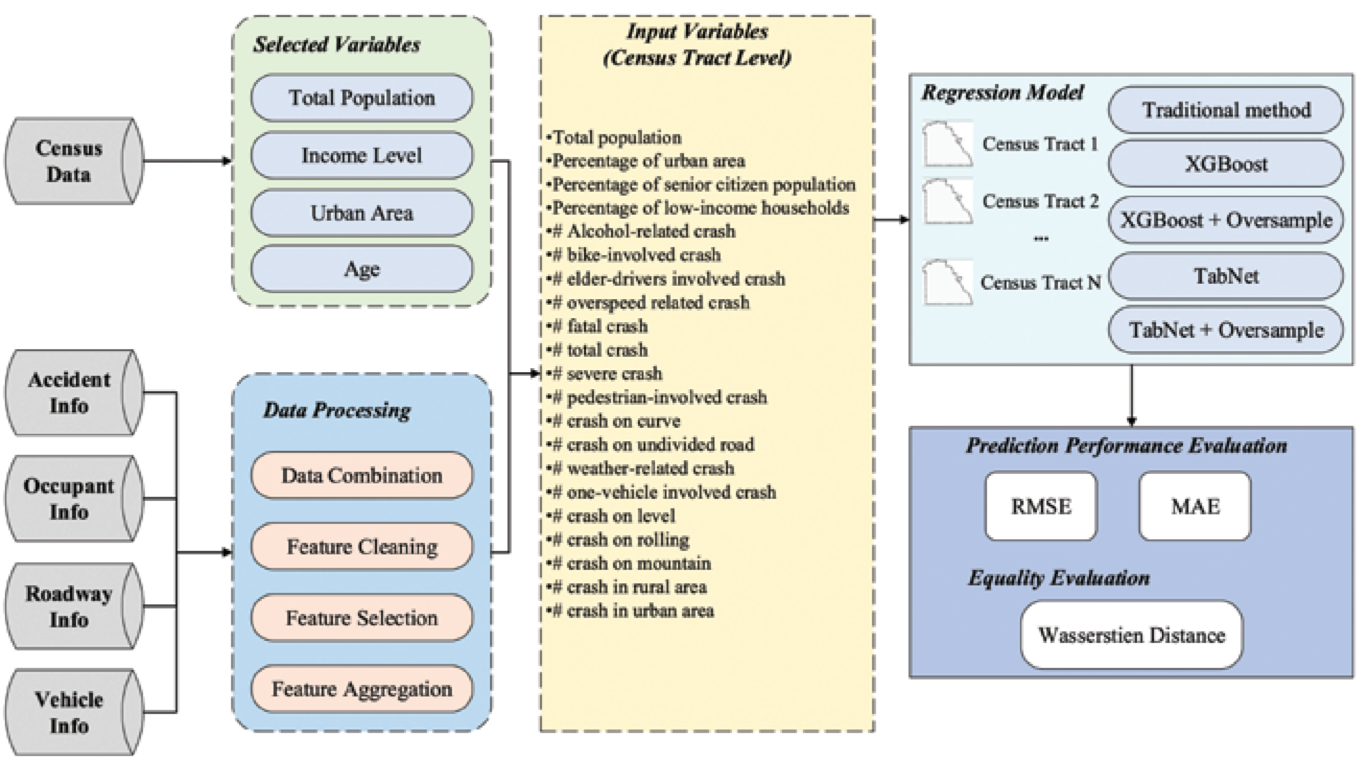
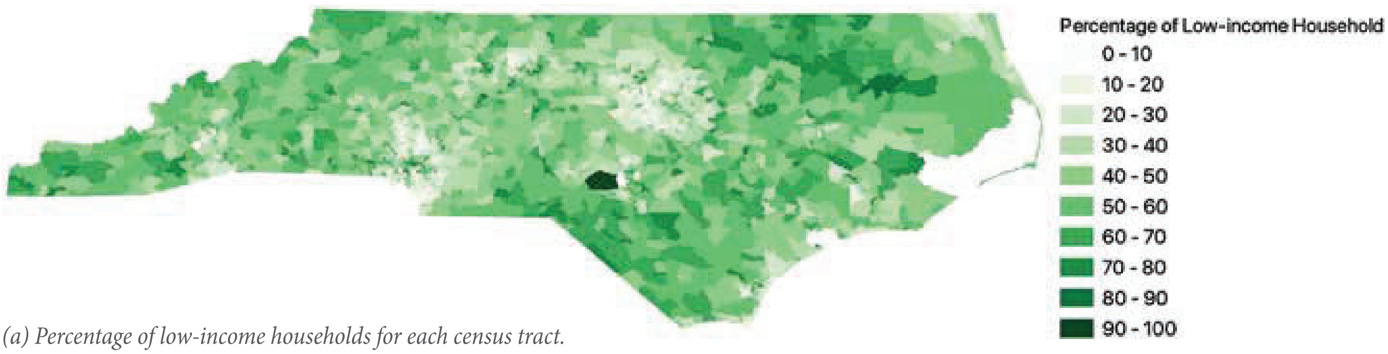


Figure 1. Data description and flow chart of proposed framework (notes: # indicates the number of).

year's input variables. Severe crashes are chosen for allocating safety improvement grants due to their higher social costs compared to minor crashes, impacting healthcare, productivity, and quality of life. Five models are tested: (1) current practice (using the previous year's severe crash count as the prediction for the next year), (2) XGBoost learning model, (3) XGBoost learning model with SMOTE, (4) TabNet learning model, and (5) TabNet learning model with SMOTE. SMOTE is applied to improve model equality by reducing data biases. Model evaluation includes assessing overall performance through root mean square error (RMSE) and mean absolute error (MAE), reflecting average prediction errors. Model equality is assessed by comparing three group pairs (high/low-income, urban/rural, and aging/non-aging) using Wasserstein Distance (WD) to determine if prediction error distributions differ between the groups. Detailed information about each step is discussed in the following sections.

HSIS crash data

Four years of crash data (2015-2018) are collected from the HSIS, including accident, occupant, roadway, and vehicle information. The crash data is processed through the following steps. Firstly, the four types of information are combined for each individual crash



(a) Percentage of low-income households for each census tract.

(b) Percentage of urban area for each census tract.

(c) Percentage of senior citizens of each census tract.

Figure 2. Maps presenting the distribution of North Carolina census tracts in terms of (A) Percentage of low-income households, (B) Percentage of urban area, (C) Percentage of senior citizens.

case. Secondly, feature cleaning and selection are performed on the combined dataset. Crash-related features such as severity, alcohol use, involvement of bikes and pedestrians, older driver involvement (65 years and over), over-speeding, airbag use, ejections, gender distribution of drivers, terrain type, and urban/rural classification are selected and cleaned. The data from step 2 is then aggregated into census tract level, resulting in the input variables (see Figure 1). The final dataset for model training includes crash-related variables from all North Carolina census tracts for 2015, 2016, and 2017, along with the number of severe crashes in the following year (i.e., 2016, 2017, and 2018) as the target variable. Severe crashes are defined as those leading to death, serious injury, or minor injury.

Sociodemographic data

The sociodemographic data of North Carolina is included as additional data to the HSIS crash data, collected from the American Community Survey (ACS) 5 year estimates.[18] The socio-demographic data reflects the social economic characteristics of different census tracts. The purpose of using the sociodemographic data is twofold: (i) to identify patterns and factors that can be associated with severe crash occurrences, and (ii) to use these features to define sensitive census tract groups. The socio-demographic related features used in this study and their definitions are the following.

- **Total population**: The total population of a census tract.
- **Urban area percentage**: The percentage of urban land area among total land area of a census tract. To get urban land area percentage of each census tract, GIS map of the urban area in the U.S. is collected from U.S. Census Bureau.[18] The percentage of urban land area can be obtained by overlaying it with the GIS map of 2020 census tracts.[19]
- **Senior citizen population percentage**: The percentage of senior citizen population (65 years and older) among total population of a census tract.
- **Low-income household percentage**: The percentage of low-income household (annual income less than $50,000) among total household of a census tract.[20]

Based on three sensitive demographic features (i.e., low-income household percentage, urban area percentage, and senior citizen percentage), census tracts are grouped into three sensitive group pairs including high-income/low-income, urban/rural, and aging/non-aging census tracts by the following thresholds, and detailed geographical distribution is presented as Figure 2.

- High-income census tracts (Low-income household percentage less than 30 percent)
- Low-income census tracts (Low-income household percentage greater than 30 percent)
- Urban census tract (Urban percentage greater than 50 percent)
- Rural census tracts (Urban percentage less than 50 percent)
- Aging census tracts (Senior citizen population percentage greater than 15 percent)
- Non-aging census tracts (Senior citizen population percentage less than 15 percent)

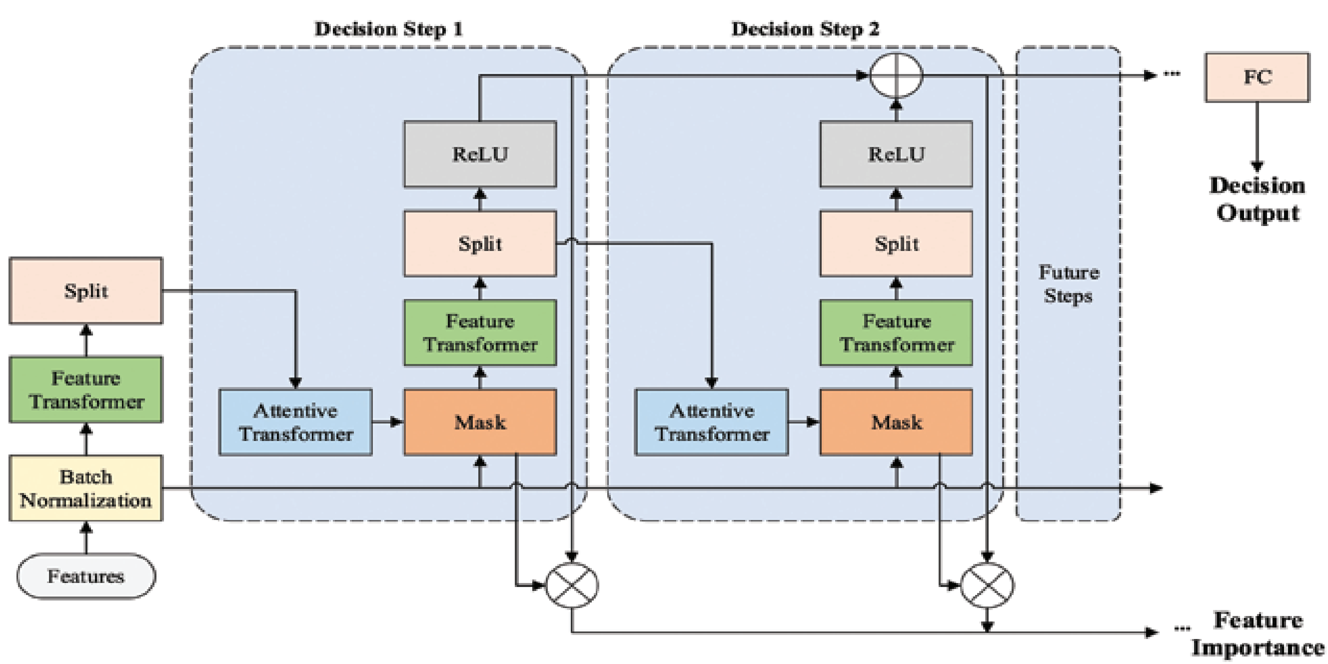


Figure 3. TabNet Architecture.[16]

Methodology

TabNet Model Description

The crash dataset applied in this study is a typical tabular data whose columns represent observations and columns represent features. Tree-based models (e.g., random forest and XGBoost) are popular and powerful technique for modeling tabular data. However, tabular data requires time-consuming feature engineering that require domain knowledge. Therefore, the TabNet model is introduced, which integrate the powerful representation ability from deep learning and the interpretability from tree-based model. The architecture of TabNet is shown in Figure 3. By adopting three key modules (i.e., attentive transformer, mask self-attention, and feature transformer) to determine the most important features at each decision step (*17*), the TabNet model can automate feature engineering and improve model performance. In this study, the traditional method and XGBoost model, a widely used tree-based model, are selected as candidates to compare with the prediction performance and model equality of TabNet.

Performance and equality evaluation metrics

Three metrics are introduced to evaluate model performance and model equality. In terms of prediction performance, root mean square error (RMSE) and mean absolute error (MAE) are used. The RMSE is sensitive to extreme error and the MAE focuses more on average error. Smaller values of these two metrics indicate better performance.

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{N}(y_i\text{-}\hat{y}_l)^2}{N}} \quad (1)$$

$$MAE = \frac{\Sigma_{i=1}^{N}|y_i\text{-}\hat{y}_l|}{N} \quad (2)$$

Where y_i is the actual value of observation i, $\hat{y}_1$ is the predicted value of observation i, and N is the total number of observations.

One way to evaluate model equality is directly comparing the values of RMSE and MAE of different sensitive groups. However, the RMSE and MAE only provide information on the overall performance and cannot provide detail information on the prediction error distribution difference. Ideally, the distribution of the model prediction error should be similar between two group in each sensitive group pair (i.e., high/low income; urban/rural; aging/non-aging), which indicates that the model performance does not bias toward one specific group. Therefore, the Wasserstein distance (WD) is involved to measure the similarity of two probability density functions (PDFs). Larger WD value indicates that the difference between two distributions is more significant.

$$W\,D\,(\mu_1,\mu_2) = \int_{-\infty}^{\infty}|F_1(x) - F_2(x)|dx \quad (3)$$

Where μ_1 and μ_2 is probability measures on $\mathbb{R}$. F_1 is the PDF of μ_1 and F_2 is the PDF of μ_2.

In general, two criteria can be applied to detect the model inequality issue: (1) significant differences in RMSE and MAE between two sensitive groups, and (2) larger WD value.

SMOTE

Model inequity arises from two sources: (1) model bias and (2) dataset bias. Model bias occurs when the model becomes overfitted to one group, prioritizing overall performance while minimizing the cost function during training. Dataset bias occurs when the training data primarily consists of samples from a specific group, leading to a biased model. To address data imbalance in the training set, this study employs SMOTE, a statistical method that increases the number of samples in minority groups to match the majority group (*16*). As shown in Figure 4, The minority groups are oversampled by creating synthetic samples in the feature space defined by the instance and its K nearest neighbors. The detailed steps of creating a synthetic minority sample are as followed: (1) select a sample instance $X_i \in \mathbb{R}^d$ and its K nearest neighbors in the minority groups, where d is the feature dimension; (2) randomly select a neighbor $X_j \in \mathbb{R}^d$ from the K nearest neighbors; and (3) create the synthetic minority sample as $X_s = X_i + \gamma|X_j - X_i|$, where $\gamma \in [0,1]$. The above steps are repeated until the number of samples in each minority group is matched with that in the majority group. Since this study has selected three binary sensitive variables, during the oversampling process, each sample group is uniquely defined by a combination of these three variables such as a sample group consisting of all rural, low-income, and aging census tracts. Thus, there are a total of 2^3=8 groups.

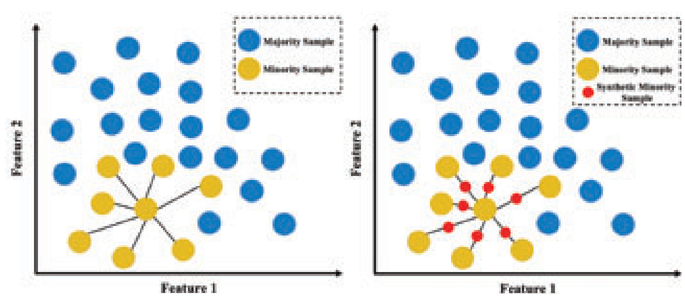


Figure 4. Illustration of the SMOTE oversampling process.[21]

Results and Discussions

The data from 2015 and 2016 serve as the training set, while the data from 2017 is used for testing. The performance and equality evaluation results of state-of-the-art models and the proposed model are presented. The models compared include the traditional method, XGBoost learning model, XGBoost learning model with SMOTE, TabNet learning model, and the proposed TabNet learning model with SMOTE (see Table 1). RMSE and MAE are used to assess overall model performance, and these metrics are separately calculated for two groups in each sensitive group pair to evaluate model equality. The WD metric is used solely for evaluating model equality. The traditional method performs the poorest in terms of overall model performance, with the highest RMSE and MAE values. AI-based XGBoost and TabNet models improve overall performance with lower RMSE and MAE values. However, these AI-based models can exacerbate biases, worsening model equality. For example, the WD value for high-income/low-income census tracts increases from 0.81 to 0.88 and 1.26 when using these AI-based models, indicating a larger difference in prediction error distributions between the two income groups. To address model equality, oversampling is applied to the training set. For both XGBoost and TabNet learning models, applying oversampling decreases WD values for all three sensitive census tract group pairs, except for a slight increase in the WD value for the aging/non-aging pair with the TabNet model. This demonstrates the effective alleviation of model biases through oversampling. Overall, the proposed model (TabNet learning model with oversampling) achieves the best overall model performance with RMSE and MAE values of 3.75 and 2.65. Furthermore, compared to the traditional method, the proposed model successfully improves model equality within each sensitive group pair. The WD values for low-income/high-income, urban/rural, and aging/non-aging census tracts are 0.75, 0.60, and 0.25, respectively, all smaller than those of the traditional method. In summary, the proposed model effectively addresses model biases while enhancing overall prediction performance.

The proposed model significantly enhances overall prediction performance for next year's severe crash numbers (see Table 1). To further demonstrate its capability, Figure 5 illustrates the geographical distribution of prediction performance improvement across census tracts. Improvement is measured by relative changes in absolute prediction errors (i.e., positive changes indicate

Table 1. Overall and Conditional Performance comparison of different approaches.

Models	RMSE (cases/year)	MAE (cases/year)	WD (cases/year)
Traditional Method			
All	4.13	3.01	-
High Income (Low Income)	3.47(4.29)	2.39(3.18)	0.81
Rural (Urban)	4.37(3.86)	3.32(2.70)	0.65
Aging (Non-Aging)	4.13(4.13)	3.07 (2.94)	0.38
XGBoost			
All	3.82	2.72	-
High Income (Low Income)	2.92(4.02)	2.05(2.89)	0.88
Rural (Urban)	3.91(3.73)	2.93(2.49)	0.56
Aging (Non-Aging)	3.74(3.92)	2.70(2.73)	0.42
XGBoost with SMOTE			
All	3.76	2.69	-
High Income (Low Income)	2.92(3.95)	2.09(2.85)	0.81
Rural (Urban)	3.88(3.63)	2.90(2.47)	0.50
Aging (Non-Aging)	3.71(3.83)	2.70(2.67)	0.36
TabNet			
All	3.91	2.76	-
High Income (Low Income)	3.00(4.12)	2.17(2.91)	1.26
Rural (Urban)	4.04(3.78)	2.93(2.57)	0.99
Aging (Non-Aging)	3.83(4.02)	2.76(2.74)	0.16
TabNet with SMOTE*			
All	3.75	2.65	-
High Income (Low Income)	2.96(3.93)	2.06(2.81)	0.75
Rural (Urban)	3.84(3.66)	2.89(2.40)	0.60
Aging (Non-Aging)	3.65(3.88)	2.67(2.63)	0.25

** The proposed framework*

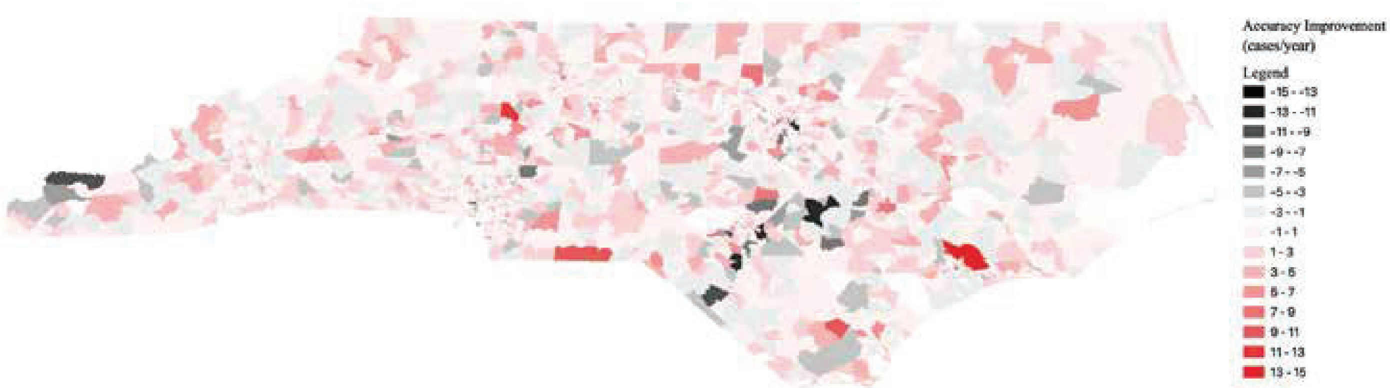
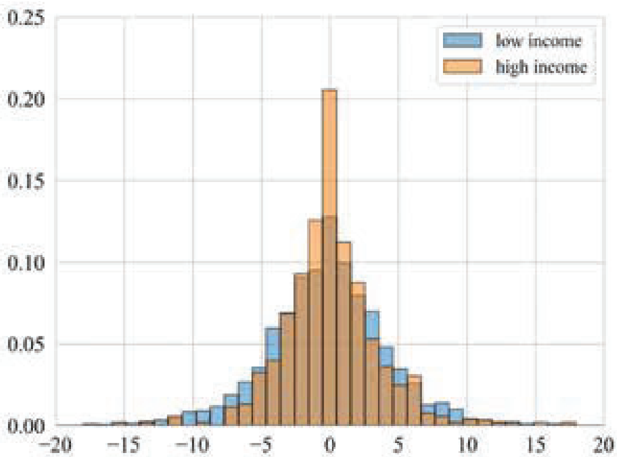
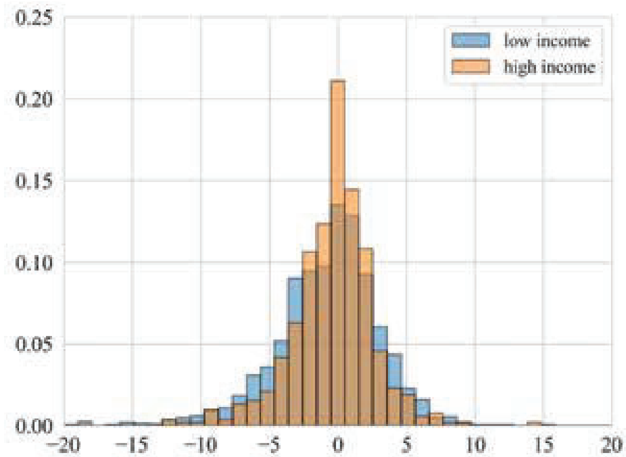


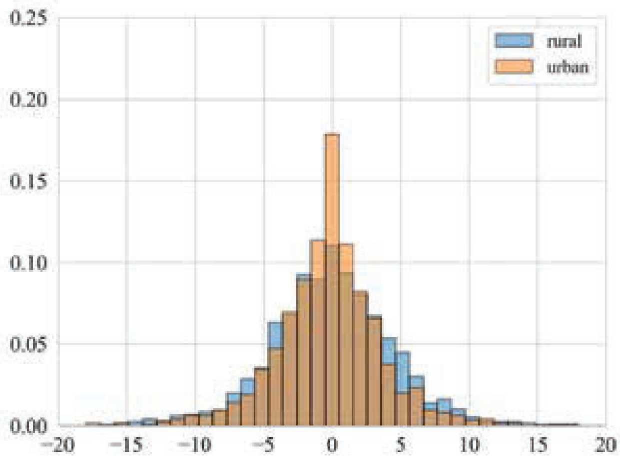
Figure 5. Geographical distribution of the improvement in prediction performance obtained by the proposed model compared with the traditional method.



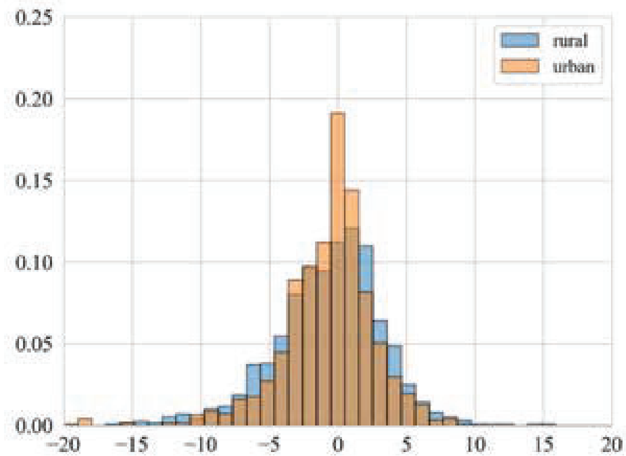
(a) low/high-income (traditional method)



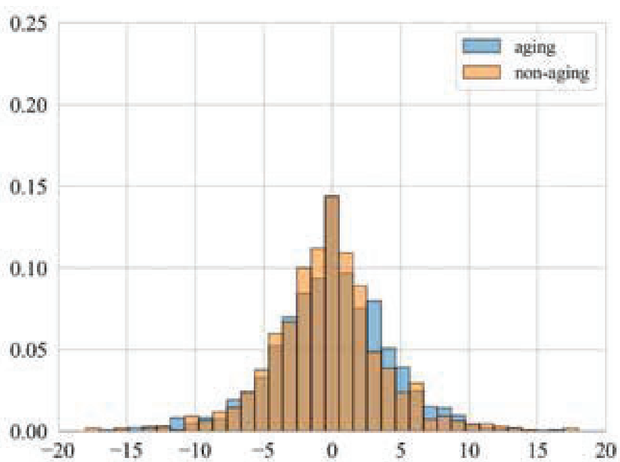
(b) low/high-income (proposed model)



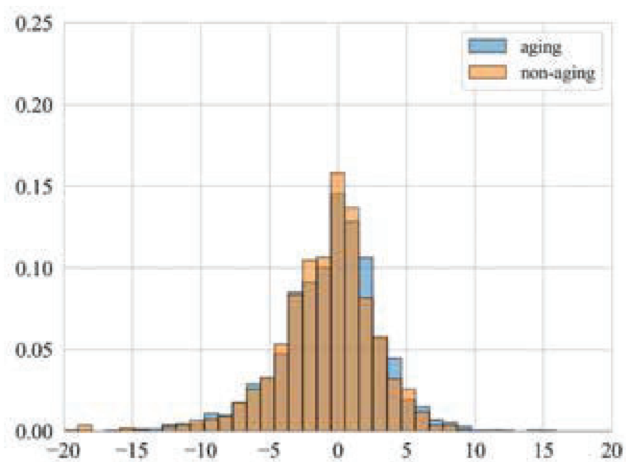
(c) rural/urban (traditional method)



(d) rural/urban (proposed model)



(e) aging/non-aging (traditional method)



(f) aging/non-aging (proposed model)

Figure 6. Comparison of prediction error distributions between low-income and high-income census tract for (a) traditional method and (b) proposed method; Comparison of prediction error distributions between rural and urban census tract for (c) traditional method and (d) proposed method; Comparison of prediction error distributions between aging and non-aging census tract for (e) traditional method and (f) proposed method.

improvement). Figure 5 highlights that the proposed model improves prediction performance for the majority of census tracts. Among 2484 census tracts, only 764 exhibit reduced prediction performance, while 1096 benefit from improved prediction performance with the proposed model.

Figure 5 demonstrates the overall prediction performance improvement achieved by the proposed model. In Figure 6, the distributions of model prediction errors are compared within each sensitive group pair for the traditional method and the proposed model. This comparison highlights that the proposed model not only enhances model performance but also improves model equality (see Figure 6 (b), (d), and (e)) compared to the traditional method (see Figure 6 (a), (c), and (e)). Model equality is improved when the prediction error probability density functions (PDFs) of the two groups in each sensitive group pair become more similar. For instance, the traditional method exhibits a significant performance advantage in high-income census tracts compared to low-income census tracts. However, applying the proposed model reduces the difference in prediction error distributions between these two groups, indicating an improvement in model equality. Similar improvements are observed for the other two sensitive group pairs.

Conclusions

Traffic crash prediction model is a vital supportive tool for governors to allocate safety improvement grants. The traditional method directly uses the number of severe crashes in previous year as reference to allocate next year's safety improvement grants. However, this method cannot accurately represent the number of crashes in next year due to the time varying nature of crash occurrences. AI-based crash prediction models can improve the overall prediction performance while, the issue of equality in AI-based crash prediction models has been neglected. Applying model with biases for the allocation of safety improvement grants may exacerbate social inequality. Therefore, in order to facilitate safety grants allocation, this study purposes an equitable framework for predicting the number of severe crashes happened in next year by incorporating oversampling technique with AI-based models. The HSIS crash data and sociodemographic data from North Carolina are utilized as a study case. Specially, this study applies XGBoost model and TabNet model to improve the overall model performance. However, albeit the improved performance, these models increase the model inequality. In order to alleviate the inequality issue, this study applies the SMOTE to balance the training dataset, thereby reducing dataset bias. The results show that for both XGBoost and TabNet learning models, the SMOTE can improve the model equality. Moreover, the proposed framework of TabNet learning model with SMOTE is proven to improve the overall prediction performance as well as the model equality within three sensitive group pairs including low-income/high-income, rural/urban, and aging/non-aging census tracts.

To enhance the robustness and comprehensiveness of this study, there are several directions for future research. Firstly, while this study attempts to improve model equality by oversampling the training dataset, further efforts can be made to redesign the cost function for model training. Secondly, more sensitive variables, such as race and education level, should be incorporated, beyond the three variables considered in this study. Finally, integrating other datasets with HSIS data may further improve the model's overall performance and equality. **itej**

Acknowledgement

The authors would like to acknowledge Dr. Kristin Kersavage, Manager at the Highway Safety Information System (HSIS) Laboratory, for providing accident data used in this study. The authors would also like to thank Dr. Yang Zhou and Dr. Yunlong Zhang from Texas A&M University for advising on this work.

References

1. White House. President Biden's Bipartisan Infrastructure Law. Washington, DC, 2021. https://www. whitehouse. gov/bipartisan-infrastructure-law (Accessed August 14, 2023).
2. Kar, K., and T. K. Datta. Development of a Safety Resource-Allocation Model in Michigan. *Transportation Research Record*, Vol. 1865, No. 1, 2004, pp. 64–71.
3. Chimba, D., G. Shao, E. Kidando, K. Hager, and S. Patel. Funding Allocation Criteria for Local Traffic Safety Enforcement Agencies. Presented at the Transportation Research Board 96th Annual Meeting, Transportation Research Board, 2017.
4. Shirani-Bidabadi, N., N. Mallipaddi, K. Haleem, and M. Anderson. Developing Bicycle-Vehicle Crash-Specific Safety Performance Functions in Alabama Using Different Techniques. *Accident Analysis & Prevention*, Vol. 146, 2020, p. 105735.
5. Srinivasan, R., and D. Carter. *Development of Safety Performance Functions for North Carolina*. North Carolina. Dept. of Transportation. Research and Analysis Group, 2011.
6. Aguero-Valverde, J., and P. P. Jovanis. Bayesian Multivariate Poisson Lognormal Models for Crash Severity Modeling and Site Ranking. *Transportation Research Record*, Vol. 2136, No. 1, 2009, pp. 82–91.
7. Tarko, A. P., K. C. Sinha, and O. Farooq. Methodology for Identifying Highway Safety Problem Areas. *Transportation Research Record*, Vol. 1542, No. 1, 1996, pp. 49–53.
8. Labi, S. Efficacies of Roadway Safety Improvements across Functional Subclasses of Rural Two-Lane Highways. *Journal of Safety Research*, Vol. 42, No. 4, 2011, pp. 231–239. https://doi.org/10.1016/j.jsr.2011.01.008.
9. Iranitalab, A., and A. Khattak. Comparison of Four Statistical and Machine Learning Methods for Crash Severity Prediction. *Accident Analysis & Prevention*, Vol. 108, 2017, pp. 27–36.

10. Theofilatos, A., C. Chen, and C. Antoniou. Comparing Machine Learning and Deep Learning Methods for Real-Time Crash Prediction. *Transportation Research Record*, Vol. 2673, No. 8, 2019, pp. 169–178.
11. Zheng, Y., S. Wang, and J. Zhao. Equality of Opportunity in Travel Behavior Prediction with Deep Neural Networks and Discrete Choice Models. *Transportation Research Part C: Emerging Technologies*, Vol. 132, 2021, p. 103410. https://doi.org/10.1016/j.trc.2021.103410.
12. Angwin, J., J. Larson, S. Mattu, and L. Kirchner. Machine Bias. In *Ethics of data and analytics*, Auerbach Publications, pp. 254–264.
13. Lin, L., F. Shi, and W. Li. Assessing Inequality, Irregularity, and Severity Regarding Road Traffic Safety during COVID-19. *Scientific Reports*, Vol. 11, No. 1, 2021, pp. 1–7.
14. Anagnostou, M., O. Karvounidou, C. Katritzidaki, C. Kechagia, K. Melidou, E. Mpeza, I. Konstantinidis, E. Kapantai, C. Berberidis, I. Magnisalis, and V. Peristeras. Characteristics and Challenges in the Industries towards Responsible AI: A Systematic Literature Review. *Ethics and Information Technology*, Vol. 24, No. 3, 2022, p. 37. https://doi.org/10.1007/s10676-022-09634-1.
15. Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority over-Sampling Technique. *Journal of Artificial Intelligence Research*, Vol. 16, 2002, pp. 321–357.
16. Arik, S. Ö., and T. Pfister. Tabnet: Attentive Interpretable Tabular Learning. No. 35, 2021, pp. 6679–6687.
17. Bureau, U. C. American Community Survey 5-Year Data (2009-2021). *Census.gov*. https://www.census.gov/data/developers/data-sets/acs-5year.html. (Accessed Februrary 16, 2023).
18. Bureau, U. C. Urban and Rural. *Census.gov*. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html (Accessed February 16, 2023).
19. Bureau, U. C. 2020 Census - Census Tract Reference Map. *Census.gov*. https://www.census.gov/geographies/reference-maps/2020/geo/2020pl-maps/2020-census-tract.html. (Accessed February 16, 2023).
20. Bennett, J., R. Fry, and R. Kochhar. Are You in the American Middle Class? Find out with Our Income Calculator. Pew Research Center.
21. Vijayvargiya, A., C. Prakash, R. Kumar, S. Bansal, and J. M. R.S. Tavares. Human Knee Abnormality Detection from Imbalanced SEMG Data. *Biomedical Signal Processing and Control*, Vol. 66, 2021, p. 102406. https://doi.org/10.1016/j.bspc.2021.102406.

Zihang Wei *is a Ph.D. student at the Zachry Department of Civil and Environmental Engineering at Texas A&M University. He is specialized in transportation engineering and he also works as a Research Assistant at Texas A&M Transportation Institute (TTI). He received his master of science degree in Civil Engineering from Texas A&M University in 2021 and received his bachelor's degree in Transportation Engineering from Nanjing University of Aeronautics and Astronautics in 2019. His research interests are connected and autonomous vehicles, transportation safety, data science, and the application of artificial intelligence in transportation engineering.*

Zihao Li *is a Ph.D. student at the Zachry Department of Civil and Environmental Engineering at Texas A&M University. He also works as a Research Assistant in Mobility Division at Texas A&M Transportation Institute (TTI). He received his bachelor's degree in Safety Engineering from Northeastern University, Shenyang, China in 2017, and received his master of science degree in Safety Engineering from Tsinghua University, Beijing, China, in 2020. His research interests include connected and automated Vehicle, transportation resilience, safety data analysis, and the application of artificial intelligence in transportation.*

Mihir Kulkarni *is a Ph.D. student at the Zachry Department of Civil and Environmental Engineering at Texas A&M University. He specializes in transportation engineering, and also works as a Research Assistant at Texas A&M Transportation Institute (TTI) where his works focus on modeling vehicular emissions. He received his bachelor's degree in Civil Engineering and master's degree in Transportation Engineering from Indian Institute of Technology Madras in 2022. His research interests are modeling of heterogeneous and non-lane-based traffic, autonomous vehicle control, and the application of connected vehicles.*

Ximin Yue *is a Ph.D. student at the Zachry Department of Civil and Environmental Engineering at Texas A&M University. He holds a master of science degree in Industrial Engineering with a specialization in data science from Texas A&M University. Additionally, he earned his bachelor's degree in Industrial and System Engineering from Shanghai Maritime University and University of Newhaven. With a deep-rooted passion for innovation, Ximin's research focuses on connected and autonomous vehicles, data science, and the application of deep reinforcement learning. Through his dedication, he aspires to make a profound impact on transportation engineering, contributing to the advancement of the field.*

Answer to "Where in the World" on page 14: Lake Pocotapaug, East Hampton, CT, USA. Photo submitted by Joseph Balskus, P.E., PTOE, RSP1 (M).