Injury Severity Risk Reduction Modeling Strategies for South Carolina Work Zone CMV Crashes

***Abdullah Al Mamun, Gurcan Comert, Judith Mwakalonge, and Mashrur Chowdhury Doctoral Student, Clemson University, Associate Professor, Benedict College, Associate Professor, South Carolina State University, and Professor, Clemson University**

Keywords

Work Zone Crash Severity, Feature Selection, Data Balancing, Crash Severity Risk Modeling, Statistical Models

Abstract

Work zones on roadways pose significant safety challenges due to the dynamic environment and increased interaction between vehicles and construction activities. The risk associated with large commercial motor vehicles (CMVs) is heightened due to operations such as lane merging and the presence of diminished lane widths. Accurate prediction of injury severity in work zone crashes could aid in developing targeted safety strategies and interventions. This could involve delivering warning messages about the risk of injury severity to connected and automated vehicles, based on vehicle and roadway sensor data, in the event of a likely crash. This study proposes the development of a crash injury severity prediction model using a comprehensive dataset of work zone crashes involving large trucks, buses, and vans in South Carolina (SC) from 2014-2018. This dataset has been compiled by the SC Department of Transportation (SCDOT) based on the crash police reports. Various feature selection methods, including Pearson's correlation coefficient, Chi-squared test statistics, feature importance using Random Forest, and recursive feature elimination using Logistic Regression are applied to the dataset. These methods aim to select the most predictive variables to enhance the performance metrics of the model. One challenge with this dataset is the class imbalance: the Low Severity (i.e., no apparent injury/property damage only) datapoints outnumbered the High Severity (i.e., possible injury, suspected minor injury, suspected major injury, and fatality) class by four to one. To address this, several data balancing methods are employed, including assigning weights to the High Severity class, a variety of oversampling techniques for the High Severity class using methods such as Synthetic Minority Oversampling Technique (SMOTE), KMeansSMOTE, Adaptive Synthetic Minority Oversampling Technique (ADASYN), Random Over-Sampling Examples (ROSE), and Generative Adversarial Network (GAN), as well as a couple of undersampling techniques for the Low Severity class using NearMiss and ROSE. Combinations of oversampling and undersampling using ROSE, and synthetic data generation for both Low and High Severity classes using ROSE, are also explored. Various sets of model training datasets are prepared using different combinations of feature selection and data balancing methods. Subsequently, a variety of statistical, machine learning, and deep learning models are trained with each of the training datasets. These models include Bayesian Mixed Ordered Logit Model, CatBoost, XGBoost, Extra Trees, Random Forest, LightGBM, KNeighbours, NeuratNetTorch, and NeuralNetFastAI. The accuracy of these models varies between 0.68 and 0.82. For the Low Severity class,

precision ranges from 0.71 to 0.88, while for the High Severity class, it ranges from 0.28 to 1.00. The recall for the Low Severity class lies between 0.51 and 1.00, and for the High Severity class, it ranges from 0 to 0.72. The F1-score for the Low Severity class is between 0.64 and 0.90, and for the High Severity class, it varies from 0 to 0.49. The area under the Receiver Operating Characteristic Curve (ROC AUC) for these models falls between 0.50 and 0.77. These results underscore the challenge of distinguishing between Low and High Severity incidents, particularly given the high precision yet low recall in the High Severity class. Considering all the model performance metrics, the Bayesian Mixed Ordered Logit Model provides the overall best fit for a dataset with all the feature selection methods and ROSE oversampling technique applied. This model achieved an accuracy of 0.70, an ROC AUC of 0.75, precisions of 0.88 (Low Severity) and 0.38 (High Severity), recalls of 0.72 (Low Severity) and 0.63 (High Severity), and F1-scores of 0.79 (Low Severity) and 0.48 (High Severity), respectively. Furthermore, the model identifies First Deformed Area, Primary Contributing Factor, Location After Impact, and Occupant Seating Location as key factors significantly associated with increased crash injury severity. While the model showed promising results in predicting injury severity in work zone crashes involving CMVs, further research is needed to improve the model's ability to identify High Severity crashes with greater accuracy.

An Adversarial Attack-Resilient Traffic Sign Classification System for Autonomous Vehicles Leveraging a Generative Adversarial Network

***M Sabbir Salek, Abdullah Al Mamun, and Mashrur Chowdhury**

Senior Engineer, USDOT National Center for Transportation Cybersecurity and Resiliency (TraCR), Greenville, SC; Ph.D. Student, Glenn Department of Civil Engineering, Clemson University, Clemson, SC; Eugene Douglas Mays Chair and Professor, Glenn Department of Civil Engineering, Clemson University, Clemson, SC

Keywords: Autonomous vehicle, Image classification, Traffic sign classification, Generative adversarial network, and Adversarial attack.

Abstract

Autonomous vehicles (AVs) rely on deep neural network (DNN)-based classification systems to recognize traffic signs. However, DNN-based classification systems have some cybersecurity vulnerabilities. For example, an adversarial attack can introduce slight perturbations to the input images fed to a traffic sign classification system and cause the underlying DNN models to misclassify the signs on the roadway. These perturbations can be so minimal that they are imperceptible to regular human eyes. However, they can be effective in deceiving the DNN models used in AVs' traffic sign classification systems. To this end, this study aims to develop an AV traffic sign classification system resilient to such adversarial attacks.

Much work has been done on DNN-based traffic sign classification systems for AVs in the literature [1-3]. However, to the best of our knowledge, none of the existing studies utilized a GAN-based adversarial defense method for AV traffic sign classification systems. In this study, the authors developed an adversarial attack-resilient traffic sign classification system based on GAN for AVs, which we refer to as the AR-GAN defense method. The novelty of the AR-GAN defense method lies in (i) assuming zero knowledge of adversarial attack models and samples and (ii) providing consistently high traffic sign classification performance under various adversarial attack types. The AR-GAN method utilizes a Wasserstein GAN-based loss function with gradient penalty [4] to overcome the typical convergence issues with GANs, such as mode collapse and vanishing gradient. The AR-GAN classification system consists of a generator that denoises an image by reconstruction, and a classifier that classifies the reconstructed image. The generator in the AR-GAN method is based on the deep convolutional GAN (DCGAN) architecture [5] and trained to generate unperturbed samples from adversarial samples before feeding them to the classifier. The classifier in the AR-GAN method is based on the residual network (ResNet) architecture [6] and trained on traffic sign images reconstructed by the generator. In addition, the AR-GAN uses a particular training framework to ensure the performance of the models used in the AR-GAN traffic sign classification system.

We tested the AR-GAN under no-attack and under various adversarial attacks, such as Fast Gradient Sign Method (FGSM), DeepFool, Carlini and Wagner (C&W), and Projected Gradient Descent (PGD). This study considered two forms of these attacks, i.e., (i) black-box attacks (assuming the attackers possess no prior knowledge of the classifier), and (ii) white-box attacks (assuming the attackers possess full knowledge of the classifier). The classification performance

of the AR-GAN was compared with several benchmark adversarial defense methods. The results showed that both the AR-GAN and the benchmark defense methods are resilient against black-box attacks and could achieve similar classification performance to that of the unperturbed images. However, for all the white-box attacks considered in this study, the AR-GAN method outperformed the benchmark defense methods. In addition, the AR-GAN was able to maintain its high classification performance under varied white-box adversarial perturbation magnitudes, whereas the performance of the other defense methods dropped abruptly at increased perturbation magnitudes. This shows the potential of the AR-GAN method to be deployed as a robust AV traffic sign classification system to achieve resiliency against various types of adversarial attacks.

The AR-GAN defense method developed in this study could help thwart adversarial attacks on an AV's perception module that would disrupt safe AV operations. Our future work will focus on expanding the AR-GAN defense method to include all types of traffic signs in the United States.