

Title: Analyzing Pedestrian Injuries in Pedestrian-Vehicle Crashes using Emerging Hotspots Analysis and XGBoost

Abstract: As one of the most vulnerable entities within the transportation system, pedestrians might face more dangers and sustain severer injuries in traffic crashes than others. However, the inherent heterogeneity of the traffic crash data can cause incorrect conclusions in many ways. Also, the crash data has inherent patterns related to both space and time. Crashes that occurred in locations with highly aggregated uptrend patterns should be worth exploring to examine the most recently deteriorative factors affecting the pedestrian-injury severities in crashes. Therefore, developments and applications of proper modeling approaches are needed to identify causes of pedestrian-vehicle crashes to better ensure the safety of pedestrians. In this study, an emerging hotspot analysis is firstly utilized to identify the most targeted hotspots, followed by a proposed XGBoost model that analyzes the most recently deteriorative factors affecting the pedestrian- injury severities. Variable importance and partial dependence of the top 15 contributing factors are identified and discussed to interpret the models and evaluate the significance of each independent variable. The overall accuracy of the best model on the hotspot dataset is 94.49%, which shows a relatively high performance compared to conventional models. Results of recent hotspots with the uptrend of crash occurrences in this research could give a solid reference for the identifications of contributing factors affecting the pedestrian-injury severities to policymakers and researchers.