SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Deep Learning, Machine Learning, or Statistical Models for Weather-related Crash Severity Prediction

Abimbola Ogungbire, M.S.
Hardik Gajera, Ph.D.
Panick Kalambay, Ph.D.
Srinivas Pulugurtha, Ph.D., P.E., F.ASCE

Project 2320
December 2023

Introduction

In the realm of traffic safety, weather-related road traffic crashes pose a significant public health concern, leading to numerous injuries and fatalities. Past research has explored crash severity prediction using statistical machine learning (ML) and deep learning (DL) models, each with its strengths and limitations. This study strategically selected a diverse set of models, including ordered logit and ordered probit models (OLM and OPM), random forest (RF), XGBoost, multi-layer perceptron neural network (MLP), and TabNet, to comprehensively assess their effectiveness in predicting crash severity while considering weather conditions. By including this array of models, this research offers valuable insights for traffic safety professionals to predict crash severity levels, enabling them to make informed decisions and allocate resources effectively to reduce the impact of weather related crashes on road safety.

Study Methods

Three main categories of models, each with its distinct characteristics and strengths, were explored. The statistical models, particularly the OLM/OPM, were selected. These models are designed to preserve the ordered nature of severity classes and have proven to be highly valuable in regression analysis involving three or more ordered categories. Building upon the statistical foundation, the study delved into ML models, including RF and XGBoost. RF, an ensemble learning method, constructs multiple decision trees to make predictions, while XGBoost iteratively corrects errors made by previous trees, optimizing prediction accuracy. The DL techniques explored the MLP and TabNet models. To assess model performance comprehensively, a range of key metrics, including precision, recall, F1 score, and accuracy, were employed. The chosen metrics were instrumental in uniformly comparing and evaluating the effectiveness and reliability of all models. With more emphasis placed on the F1 score, the best models in evaluating different crash severity levels were identified.

Findings

The study's findings regarding crash severity prediction reveal variations in model performance across different severity levels. For predicting severe-injury crashes, the TabNet model was identified as the most effective, albeit with relatively low precision (8.45%) and an F1 score of 28.78, suggesting some challenges in severe-injury classification. This contrasts with previous research, where models such as XGBoost have outperformed others, possibly due to their proficiency in handling non-linear relationships in crash data. However, the choice of the best model should be context-driven, considering dataset characteristics and problem complexities.

In the case of moderate-injury classification, the RF model exhibited the best performance, with a notably higher F1 score of 57.03%, indicating a better balance between precision and recall. Achieving this balance is essential, given the varying importance of false positives and false negatives in different traffic safety applications.

For PDO crash prediction, the TabNet model outshone others, with a high precision of 79.62% and an impressive F1 score of 92.1, signifying an excellent balance of precision and recall. These findings indicate that the model's performance varied significantly across severity levels, with PDO predictions being the most accurate, moderate-injury predictions demonstrating a good balance, and severe-injury predictions posing more significant challenges. Several factors, including class imbalance, complexity differences, and data distribution nuances, contribute to these variations in model performance across severity levels.

Policy/Practice Recommendations

The models developed in this study have practical implications and offer valuable recommendations for enhancing traffic safety. The RF model's balanced precision and recall make it suitable for allocating resources to moderate-injury crashes, potentially reducing injury severity. Traffic safety authorities and organizations are advised to consider implementing RF to optimize emergency responses. Additionally, the TabNet model's exceptional precision and recall for PDO crashes can streamline insurance processes and resource allocation. It is recommended that insurance agencies leverage TabNet for accurate PDO crash identification. Furthermore, these predictive models can inform targeted interventions and infrastructure improvements in areas prone to severe crashes, contributing to proactive traffic safety measures. Integrating models such as RF and TabNet into decision-making processes empowers stakeholders to make informed, data-driven decisions, ultimately reducing injuries and fatalities on roadways.

About the Authors

Abimbola Ogungbire, M.S., is pursuing his Ph.D. in the Infrastructure and Environmental Systems (INES) program at the University of North Carolina at Charlotte. His areas of interest are traffic data analytics and visualization.

Panick Kalambay, Ph.D., recently graduated with his doctoral degree in Civil Engineering at the University of North Carolina at Charlotte. His areas of interest are traffic operations & safety, and transportation planning.

Hardik Gajera, Ph.D., recently graduated with his doctoral degree in Civil Engineering at the University of North Carolina at Charlotte. His areas of interests include traffic operations & safety, traffic simulations, and connected & automated vehicles.

Srinivas S. Pulugurtha, Ph.D., P.E., F.ASCE, is currently working as a Professor & Research Director of the Department of Civil and Environmental Engineering at the University of North Carolina at Charlotte.

To Learn More

For more details about the study, download the full report at
transweb.sjsu.edu/research/2320

MTI is a University Transportation Center sponsored by the U.S. Department of Transportation's Office of the Assistant Secretary for Research and Technology and by Caltrans. The Institute is located within San José State University's Lucas Graduate School of Business.