Transfer Learning with data augmentation to Enhance Injury Severity Prediction in Car-to-Truck Rear-End Crashes

Peilong Yang, Rui Yang, Huida zhang, Yeheng Liu, Hang Qiu, Yu Liu

I. INTRODUCTION

Data-driven algorithms based on machine learning (ML) and deep learning (DL) have demonstrated significant predictive accuracy for occupant injury assessment in traffic accidents [1]. However, specific accident types exhibit systematic deviations in injury patterns compared to conventional accidents. Analysis of the Future Mobile Traffic Accident Scenario Study (FASS) database developed by China Automotive Engineering Research Institute (CAERI) reveals a critical disparity: passenger car-to-truck rear-end crashes, though representing only 5% of total cases, exhibit severe injury rates (AIS3+) exceeding 58%—double the baseline for general crashes. This rare yet high-risk crash type poses critical challenges to conventional prediction models due to data scarcity and distributional mismatch. To address this, we proposed a two-stage framework integrating transfer learning [2] with data augmentation to improving the injury severity prediction in passenger car to truck rear-end crashes.

II. METHODS

The study analyzed 1,671 passenger vehicle crash cases (4,187 occupants) from CAERI's FASS database after rigorous filtering, retaining only cases with critical crash-related parameters and occupant injury severity (AIS levels). 92 car-to-truck rear-end crashes (207 occupants) were isolated as the subset.

Data Preparation

A strict case-level stratified split was implemented to ensure independence between datasets, car-to-truck rear-end cases were split into stratified training (70%), validation (15%), and test sets (15%), while other cases followed an 80%-10%-10% split.

Training Framework

A two-stage transfer learning framework was developed to train this enhanced model: (1) Pretraining: An XGBoost model [3] was pretrained on mixed dataset to capture generalized injury patterns across diverse crash types. The pretraining dataset contains all non-truck training set (i.e., 80% of total non-truck cases) and 30% subset of car-to-truck rear-end training set (i.e., 30% of the 70% allocated to training) (2) Fine-tuning: The remaining 70% of the car-to-truck rear-end training set was augmented via SMOTE (Synthetic Minority Oversampling Technique) [4] for fine-tuning, and the pretrained model was frozen to preserve existing knowledge, with incremental adjustments to specialize in truck rear-end crash.

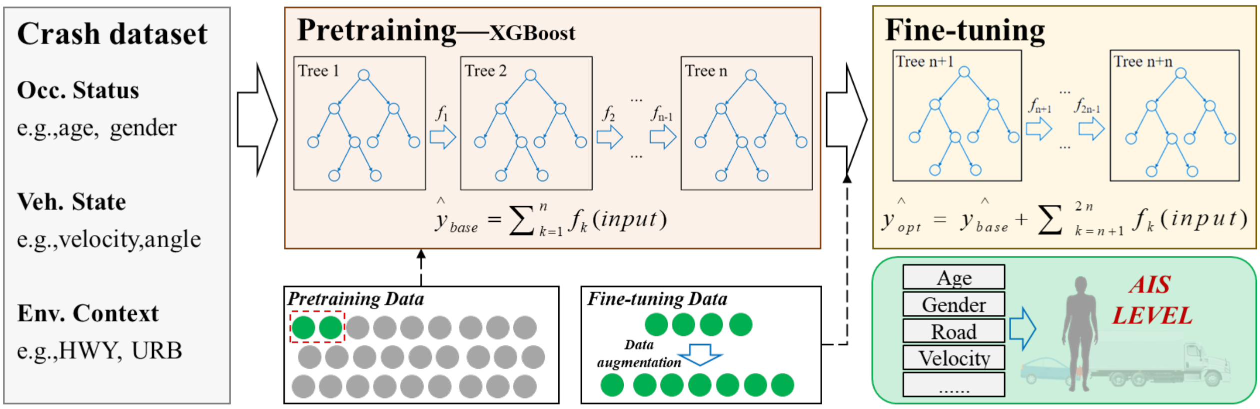

Fig.1. Two-stage transfer learning framework

P. L. Yang (First author), R. Yang, H. D. Zhang, H. Qiu, Y. Liu (Corresponding author, e-mail:liuyu@caeri.com.cn, tel:+86 17843761276) are Researchers in China Automotive Engineer Research Institute Co., Ltd., Chongqing, China. Y.H. Liu, is a postgraduate student in Chongqing University of Technology, Chongqing, China.

III. INITIAL FINDINGS

To validate the effectiveness of the proposed transfer learning framework, we compared three key models on independent test sets using standard metrics: accuracy (overall correctness), recall (sensitivity to severe injuries), and AUC (discrimination power). We additionally trained a one-stage model using conventional end-to-end training on the full mixed training set for comprehensive comparison. As shown in Table 1, on the truck rear-end test set, the fine-tuned model maintained equivalent recall (80.0%) to both pretrained model and one-stage model, while delivering substantial improvements in accuracy and AUC: Its 57.7% accuracy and 0.62 AUC exceeded the pretrained model by 3.9% accuracy/5.1% AUC, and surpassed the one-stage model by 3.2% accuracy/3.3% AUC. This demonstrates a significant enhancement in its predictive performance for passenger car rear-end truck crash injuries. On the global test set (Table II), despite a controlled decrease of 1.5% in accuracy than pretrained, the fine-tuned model achieved the highest AUC (0.74) and recall (64.4%), outperforming the one-stage (0.71 AUC, 62.5% recall) and pretrained (0.73 AUC, 63.8% recall). These results clearly indicate that, the optimized model exhibits robustness across diverse crash types without forgetting of general patterns, ensuring reliable performance in both specialized and broad accidents.

TABLE I
PERFORMANCE ON TRUCK REAR-END TEST SET

	Accuracy	Recall	AUC
One-Stage	54.5%	80.0%	0.60
Pretrained	53.8%	80.0%	0.59
Fine-tuned	57.7%	80.0%	0.62

TABLE II
PERFORMANCE ON GLOBAL TEST SET

	Accuracy	Recall	AUC
One-Stage	67.8%	62.5%	0.71
Pretrained	71.6%	63.8%	0.73
Fine-tuned	70.1%	64.4%	0.74

IV. DISCUSSION

Accurate injury severity prediction in car-to-truck rear-end crashes remains challenging due to data scarcity and distributional discrepancy. By integrating mixed-data pre-training and scenario-specific fine-tuning—augmented with SMOTE, the proposed framework enhances prediction robustness, achieving improved accuracy and AUC performance while preserving generalizability. Moreover, this approach enables scalable adaptation to other rare, high-risk crash types (e.g., rollovers [5]). However, the reliance on moderate-sized crash datasets and binary injury classification (MAIS3+ vs. non-MAIS3+) may oversimplify biomechanical reality. Our future work will extend this framework to large-scale crash datasets and deep learning models for continuous biomechanical metrics.

V. REFERENCES

[1] Santos, K., et al., J Saf Res, 2022.
[2] Pan, S. J., et al., IEEE Trans. Knowl. Data Eng, 2010.
[3] Chen, T., et al., SIGKDD, 2016.
[4] Chawla, N. V., et al., JAIR, 2002.
[5] Riexinger, L. E., et al., IRCOBI, 2022.