Domain Knowledge-Guided Machine Learning for Enhanced Crash Dynamics Prediction

Niranjan Ballal, Dr. Thomas Soot, Dr. Michael Dlugosch
(Fraunhofer-Institute for High-Speed Dynamics, Ernst-Mach-Institut, EMI, Germany)

Abstract

Machine learning (ML) is emerging as a key tool for predicting crash dynamics in near real-time, driving significant advancements in automotive safety engineering. In this domain, high-fidelity synthetic data, such as finite element (FE) crash simulation data, plays a pivotal role. Real-world crash data often presents challenges such as limited availability, noise, and incomplete information, making it difficult to train robust ML models. By contrast, simulation data provides detailed insights into the underlying physical phenomena, offering a controlled environment to generate diverse datasets. This makes simulation data an indispensable resource for the development of predictive models that aim to improve vehicle safety and occupant protection.

Traditional ML approaches in this field often rely on scenario-level input parameters, such as impact velocity, and collision angles, to predict outcomes like intrusion or injury levels. While these approaches are effective to a degree, they frequently fall short of leveraging the rich, granular information embedded within simulation data. This limitation can result in suboptimal predictive accuracy, particularly when dealing with complex crash dynamics involving multiple interacting factors.

A domain knowledge-guided methodology is introduced to address this limitation, segmenting the problem domain into smaller, homogeneous sub-domains based on specific physical phenomena. By reducing the required ML model complexity within each sub-domain, tailored ML models are developed to optimize predictions for specific crash dynamics, enhancing both data efficiency and prediction model accuracy. This approach leverages hierarchical model trees informed by domain expertise, ensuring that each sub-domain's unique traits are effectively captured without relying on a singular, overly complex model.

Embedding domain knowledge through segmentation not only allows for the customization of sub-model architectures but also facilitates adaptive data generation. The proposed methodology improves predictive performance and reduces training sample requirements while preserving fidelity. Comparative evaluations demonstrate superior accuracy and robustness in capturing nuanced

relationships within crash simulation data, positioning this approach as a significant step forward in the application of ML to safety-critical domains.

1. Introduction

The automotive market is undergoing a transformative shift, driven by software, connectivity and electrification [1, 2]. Within this transformation, autonomous driving technologies are reshaping urban mobility, driving towards safer and more convenient transportation systems. In 2022, the passenger autonomous vehicle market was valued at about $40 billion and is expected to reach around $300 billion in 2035, with a significant contribution from conditional (SAE Level 3) and high (SAE Level 4) automation vehicles [3]. Currently, Europe is the second-largest market for autonomous vehicles (AVs), with China following. In 2023, electric vehicles (EVs), which often serves as a platform for AV technologies, comprised around 14.6% of all new car sales in the European Union (EU) [4]. This expansion is supported by strong government support, particularly in countries like Germany.

AVs are starting to become more prevalent in urban settings. While several automotive manufacturers such as BMW, Mercedes-Benz, and Honda have achieved even Level 3 autonomy, the focus for many is now toward the higher levels, with very strong emphasis on Vulnerable Road Users (VRUs), such as pedestrians, cyclists, and micro-mobility users [5, 6, 7]. With this shift, cities are transitioning towards mixed traffic ecosystems where AVs coexist with human-driven vehicles and VRUs [8]. It has also emerged from studies that, while Connected and Autonomous Vehicles (CAVs) are believed to have a number of safety benefits, there is also significant concern over the interaction with VRUs and their safety [9].

Recent research underlines that attention to VRU safety issues shall become increasingly crucial with increasing market penetration of AVs. For instance, Waymo has conducted various studies analysing real-world accidents involving VRUs and human drivers, gaining deeper insights into the injury risks VRUs face and to enhance injury mitigation systems [10]. Another systematic review indicated that it is challenging task to eliminate accidents involving VRUs, emphasizing the need for improved machine learning (ML) prediction methods, sensor technologies and a need for reliable data source [11]. Thes limitations are compounded by the insufficiency of real-world crash injury data involving VRUs, which is often sparse, noisy or incomplete, hindering the development of robust ML models for such safety-critical applications. To overcome this limitation, Finite Element (FE) simulation data can be used as a reliable high-fidelity synthetic data source. FE simulation data provides detailed insights into the underlying crash dynamics, offering a controlled and scalable environment to generate diverse datasets.

In this paper, we explore how FE simulation data can be exploited using ML and domain expertise to enhance predictive injury modelling for crash dynamics. We propose a domain knowledge-guided method that segments the problem domain into small, homogeneous sub-domains, optimizing model accuracy and data efficiency. By leveraging the FE simulation data, we aim to improve predictive performance while addressing the limitations of limited availability of real-world data. This approach feeds into the broader goal of enhancing VRU safety as autonomous driving transitions into complex urban traffic environments

2. Proposed »Hierarchical Model-Tree« Approach

As shown in Figure 1: traditional ML approaches for injury prediction in vehicle-to-VRU crash scenarios typically map input parameters, such as collision velocity and impact angle, directly to injury severity. While this approach is straightforward and provides some predictive capability, it often requires a large number of training samples to capture the complex physical interactions involved. As a result, generating the necessary high-fidelity crash simulations becomes computationally expensive and resource intensive. The methodology proposed in this paper addresses this limitation by introducing a structured, domain-informed approach that enhances prediction accuracy while efficiently leveraging the detailed characteristics of simulation data.

Figure 1: Training and prediction pipeline for the global modelling approach, where a single ML model is used to map inputs to outputs.

As shown in the **Error! Reference source not found.**the proposed method implements a two-staged process, first classifying the crash impact region before applying sub-section-specific regression models for injury prediction. This segmentation is based on the underlying physical phenomena. For example, in the case of head impacting the windshield region, impact of head on the A-pillar would differ significantly compared to the impact of head in the middle of the windshield. The classification model classifies impact locations using key output channels derived from simulation data, which is determined by an expert. Then, dedicated regression models, trained separately for each sub-section, predict injury severity with greater accuracy with relatively lower number of samples due to the homogeneity of the sub-sections. This approach is particularly relevant for VRU crashes, where VRUs may impact with different vehicle structures, such as windshields, motor hood or side pillars, with varying injury outcomes.

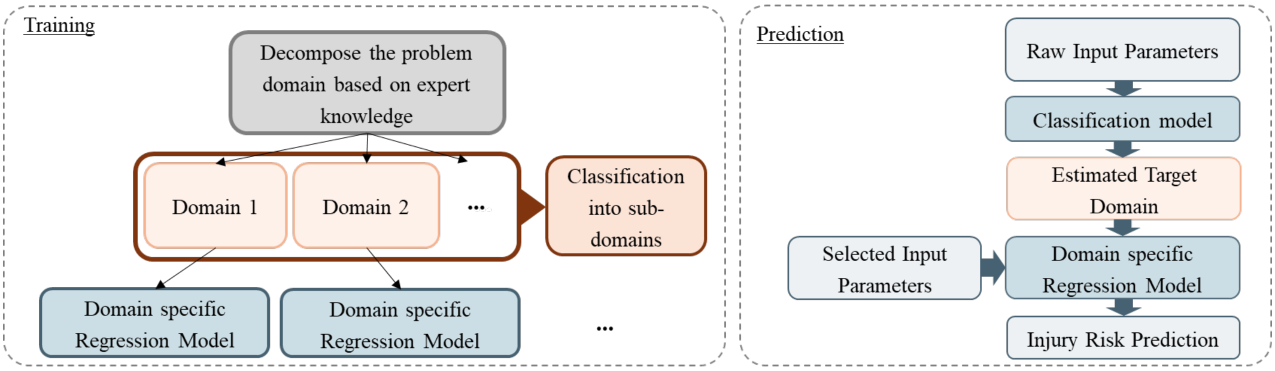

Figure 2: Training and prediction pipeline for the »Hierarchical Model-Tree« approach, where the problem is decomposed into sub-domains to leverage the detailed characteristics of FE simulations

3. Implementation

The proposed method is applied to the problem involving vehicle-to-cyclist crash. An exemplary simulation setup is as shown in the Figure 3: Different scenario parameters, such as vehicle velocity, cyclist velocity, impact angle and impact location along the vehicle bumper are varied. For this study, the focus is on predicting head injuries, head injury criteria (HIC) [12], with the first step involving the selection of the sub-region where the head first contacts the windshield. This subset of data is visualized in Figure 4: , where each scatter point represents a simulation. The position of each point on the plot indicates the location of the head at the moment of initial contact with the windshield, which is outlined using the blue boundary.

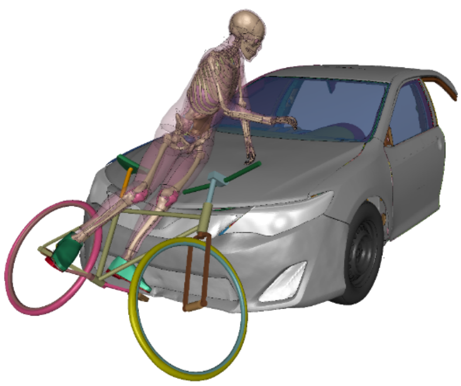

Figure 3: Exemplary FE simulation setup involving vehicle-to-cyclist crash

The windshield region is divided into sub-sections based on the varying stiffness, as shown in Figure 4: . The region is primarily sub-divided into A-pillar, windshield center and a transition regions. The A-pillar region has the highest stiffness, while the windshield center is comparatively less stiff. The transition region shows a mixture of properties from these two sub-regions. As shown in Figure 4: a classification model is first trained to classify the simulations into these sub-regions first before predicting the HIC. To predict the injuries, different ML models and architectures, specifically tailored to each of these sub-regions, are trained.

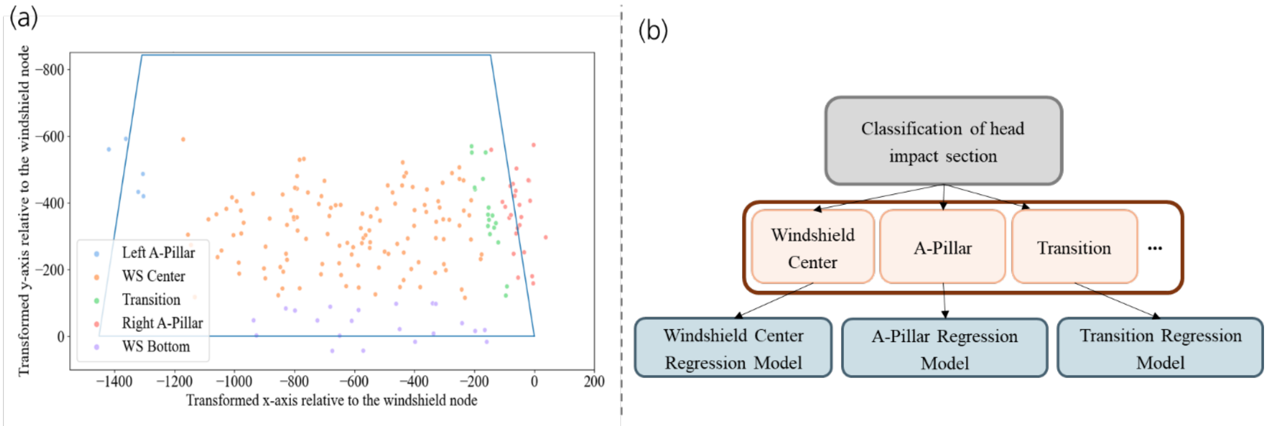

Figure 4: (a) Scatter plot of all simulations having the first head contact with the windshield, with the impact location along the windshield. The color indicates the corresponding sub-section to which each sample belongs. (b) »Hierarchical Model-Tree« approach pipeline for vehicle-to-cyclist crash scenario, with only simulations having the first head contact with the windshield.

4. Results and Discussion

As shown in Table 1: the global modelling approach, where a single ML model is trained to predict the HIC yields an R^2 score of 0.478. In this context, an R^2 of 0 indicates poor performance, while an R^2 of 1 represents excellent performance. In comparison to the global modelling approach, »Hierarchical Model-Tree« approach achieves an average R^2 score of 0.850. The individual R^2 scores for each sub-regions - windshield center, A-pillar and transition region - are 0.904, 0.803 and 0.519 respectively. To compute the average, these individual scores are weighted by the number of samples in each of these sub-regions before calculating the overall average. The proposed »Hierarchical Model-Tree« approach is observed to perform significantly better in comparison to the global modelling approach by approximately 77%.

The results highlight the effectiveness of integrating domain-knowledge, leveraging the simulation data to significantly improve the accuracy of engineering predictions. As the solution is motivated using the domain knowledge, it also brings in greater confidence in the prediction and the architecture. The success of the »Hierarchical Model-Tree« approach in the selected vehicle-to-cyclist crash scenario suggests that applying this method to other crash types could lead to enhanced prediction outcomes. Additionally, exploring the impact of this approach on the other challenges, such as efficient data generation [13], could prove to be valuable.

	Global Modelling Approach	»Hierarchical Model-Tree« Approach
Global/ Weighted Average	0.478	0.850
Sub-section: WS Center	-	0.904
Sub-section: A-Pillar	-	0.803
Sub-section: Transition	-	0.519

Table 1: R^2 score comparison for global modelling and »Hierarchical Model-Tree« approach

5. References

[1] 'Gartner Identifies Key Automotive Trends for 2025', Gartner. Accessed: Jan. 28, 2025. [Online]. Available: https://www.gartner.com/en/newsroom/press-releases/2025-01-16-gartner-identifies-key-automotive-trends-for-2025

[2] 'Digital transformation in the automotive industry', ericsson.com. Accessed: Jan. 28, 2025. [Online]. Available: https://www.ericsson.com/en/industries/automotive

[3] 'The future of autonomous vehicles (AV) | McKinsey'. Accessed: Jan. 28, 2025. [Online]. Available: https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/autonomous-drivings-future-convenient-and-connected

[4] 'The future is near: when will Europe see self-driving cars and is society ready?' Accessed: Jan. 28, 2025. [Online]. Available: https://gurtam.com/news/self-driving-cars

[5] 'Level 3 highly automated driving available in the new BMW 7 Series from next spring.' Accessed: Jan. 28, 2025. [Online]. Available: https://www.press.bmwgroup.com/global/article/detail/T0438214EN/level-3-highly-automated-driving-available-in-the-new-bmw-7-series-from-next-spring?language=en

[6] M.-B. Group, 'Mercedes-Benz increases top speed of its Level 3 automated driving system to 95 km/h', Mercedes-Benz Group. Accessed: Jan. 28, 2025. [Online]. Available: https://group.mercedes-benz.com/innovations/product-innovation/autonomous-driving/drive-pilot-95-kmh.html

[7] 'Japan: Honda startet Sicherheitssystem mit Fahrfunktionen nach Level 3'. Accessed: Jan. 28, 2025. [Online]. Available:

https://www.honda.de/cars/world-of-honda/news-events/2021-03-04-japan-honda-startet-sicherheitssystem-mit-fahrfunktionen-nach-level-3.html

[8] 'Autonomous driving behaviors for vulnerable road users - KPMG Japan', KPMG. Accessed: Jan. 28, 2025. [Online]. Available: https://kpmg.com/jp/en/home/insights/2024/11/auto-intelligence-vru.html

[9] A. P. Morris *et al.*, 'Autonomous Vehicles and Vulnerable Road-Users—Important Considerations and Requirements Based on Crash Data from Two Countries', *Behav Sci (Basel)*, vol. 11, no. 7, p. 101, Jul. 2021, doi: 10.3390/bs11070101.

[10] 'Protecting Vulnerable Road Users: New Insights from Waymo's Safety Research', Waymo. Accessed: Jan. 28, 2025. [Online]. Available: https://waymo.com/blog/2024/11/waymos-research-vru

[11] A. Deshmukh *et al.*, 'A systematic review of safety-critical scenarios between automated vehicles and vulnerable road users'.

[12] E. Hertz, 'A NOTE ON THE HEAD INJURY CRITERION (HIC) AS A PREDICTOR OF THE RISK OF SKULL FRACTURE', in *Publication of: ASSOCIATION FOR THE ADVANCEMENT OF AUTOMOTIVE MEDICINE*, 1993. Accessed: Jan. 31, 2025. [Online]. Available: https://trid.trb.org/View/408468

[13] B. Niranjan, S. Thomas, D. Michael, T. Niclas, and F. Dirk, 'A Method for Efficient Generation and Optimization of Simulation-Based Training Data for Data-Driven Injury Prediction in VRU-Vehicle Accident Scenarios', presented at the 27th International Technical Conference on the Enhanced Safety of Vehicles (ESV)National Highway Traffic Safety Administration, 2023. Accessed: Jan. 31, 2025. [Online]. Available: https://trid.trb.org/View/2211667