# Neural Network based Safety Metric for Autonomous Vehicles

Haojie Zhu
*Department of Civil and Environmental Engineering*
*University of Michigan*
Ann Arbor, Michigan, USA
zhuhj@umich.edu

***Abstract*—Safety metrics are crucial for autonomous vehicles (AVs) as they provide early warnings to human drivers and aid in the development of AV technologies. Numerous safety metrics exist in the literature; however, their underlying assumptions are often overly simplistic, which impacts their precision and accuracy. To address this issue, I developed a neural network-based safety metric that uses collected trajectory data without assuming specific driving behaviors. In a case study involving a three-lane highway, the proposed neural network-based safety metric was compared with four existing safety metrics: Safety Metric based on the Assessment of Risk, Model Predictive Instantaneous Safety Metric, Pegasus Criticality Measure, and time-to-collision. The results indicate that the neural network-based safety metric achieves superior precision and accuracy.**

***Index Terms*—Safety metrics, Autonomous vehicles, Machine Learning**

## I. Introduction

Autonomous vehicles (AVs) have consistently garnered public interest due to their potential to significantly improve convenience and safety in daily life. However, there remains a pronounced gap between the current performance of AVs and public expectations. A notable example is a 2023 incident involving a Cruise Robotaxi that struck and dragged a pedestrian, raising public concerns about the safety performance of AVs. Safety metrics play a crucial role in enhancing the safety performance of AVs by serving as safeguards that alarm human drivers and disengage automated modes. Additionally, these metrics can identify driving scenarios essential for the further advancement of AV technologies.

Most existing safety metrics assess danger levels by predicting the trajectories of background vehicles (BVs) and AVs based on various driving behavior assumptions. Commonly used safety metrics include the Safety Metric based on the Assessment of Risk (SMAR) [1], Model Predictive Instantaneous Safety Metric (MPrISM) [2], Pegasus Criticality Measure (PCM) [3], and time-to-collision (TTC) [4]. SMAR uses naturalistic driving data to construct a realistic probabilistic behavior model for human drivers, allowing it to predict collision probabilities for any traffic state by considering real-world driving characteristics. However, developing such a realistic behavior model is challenging, especially in safety-critical scenarios, due to data limitations.

MPrISM evaluates the safety status of the ego vehicle by considering the worst-case scenario for a given traffic snapshot, assuming that background vehicles will take aggressive actions to cause a crash, and the ego vehicle will try its best to avoid the accident. It then assesses danger based on the calculated time to collision. The PCM algorithm uses Model Predictive Control (MPC) [5] to manage the ego vehicle and calculate the optimal acceleration decision, assuming that background vehicles maintain their speed and heading. A scenario is deemed safety-critical when the required acceleration surpasses a certain threshold. The TTC method assumes the ego vehicle maintains its current state, such as velocity and heading, and estimates the minimum time until a collision with other vehicles occurs. These methods' effectiveness heavily relies on the underlying assumptions; for instance, MPrISM might be conservative as other vehicles are unlikely to always take the most aggressive actions. Both PCM and TTC assume that background vehicles will maintain their current states, which is not always the case in real-world scenarios.

To address these challenges and enhance the accuracy of safety metrics, this project investigates the application of deep learning algorithms to develop a neural network-based safety metric. This work focuses on multi-lane highway scenarios, employing a collision-inclusive simulation with Simulation of Urban Mobility (SUMO) [6] to construct the dataset. The developed safety metric shows superior accuracy and precision compared to existing rule-based metrics.

## II. Method

### *A. Neural network structure*

Given the suitability of Multilayer Perceptrons (MLPs) for common classification and regression tasks [7], the neural network based safety metric uses MLP as the backbone. The model comprises a frequency encoding layer, four layers of MLP with Batch Normalization, a prediction layer, and a sigmoid function layer, as depicted in Fig. 1.

The raw input consists of the ego vehicle and the nearest six vehicles' states, including position, velocity, and heading. Before feeding the state vectors into the MLP with a Batch Normalization backbone, we employ the frequency encoding method outlined in [8], which uses a set of sine and cosine basis functions to project the vectors into a high-dimensional space, thereby improving the capture of high-frequency variations in the state spaces. The state value $s$ after the mapping function $M(s)$ can be written as follows:

$$M(s) = [s, s^2, \sin(s), \cos(s)]$$

Batch Normalization is integrated into MLP to stabilize the training process and enhance performance [9]. Finally, a linear prediction layer and a sigmoid function layer are used to output the safety criticality of the current scenario.

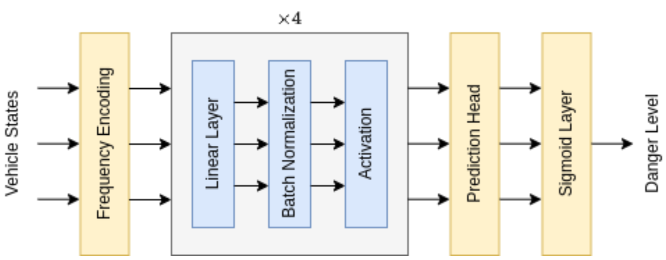


Fig. 1. Overall structure of the neural network.

### B. Trajectory dataset

To replicate a naturalistic driving environment (NDE), the behavior of BVs is modeled using real-world naturalistic driving data (NDD). Detailed information about the NDE modeling method used in this study can be found in [11]. To efficiently generate safety-critical trajectories, the naturalistic and adversarial driving environment (NADE) [12] is used to produce trajectories that include collisions. For evaluating the generalizability of safety metrics, the ego vehicle is controlled using three different algorithms. One algorithm combines Intelligent Driving Model (IDM) [13] and Minimizing Overall Braking Induced by Lane Change (MOBIL) model [14], while the other two algorithms are based on reinforcement learning methods [1] designed for safety and mobility.

In total, 1,291 hours of simulation data are collected with a time resolution of 0.1 seconds, during which 3,797 crashes are recorded. The dataset is then divided into three parts: 20% for validation, 20% for testing, and 60% for training. Using the logged trajectory data, the dataset is annotated according to the evaluation framework proposed in [10]. Specifically, states that will inevitably cause a crash are labeled as 1, while all other steps are labeled as 0.

Fig. 2. The studied three-lane highway with a length of 1200 meters. The green symbol represents a background vehicle, while the red symbol denotes the ego vehicle. The length and width of all vehicles are 5m and 1.8m, respectively. The lane width is 4m.

## III. Results

The efficacy of the proposed neural network-based safety metric is evaluated on a three-lane highway, as depicted in Fig.2. The training process, illustrated in Fig. 3, demonstrates rapid convergence. Furthermore, this neural network-based metric is compared with SMAR, MPrISM, PCM, and TTC, using the test dataset. Detailed implementation information for these safety metrics can be found in [1] and [10].

The Receiver Operating Characteristic (ROC) curve and Precision-Recall (PR) curve are presented in Fig. 4. According to the statistical analysis, the neural network-based safety metric achieves the highest precision and recall, along with the largest area under the curve (AUC), indicating superior performance in classifying safety-critical scenarios.

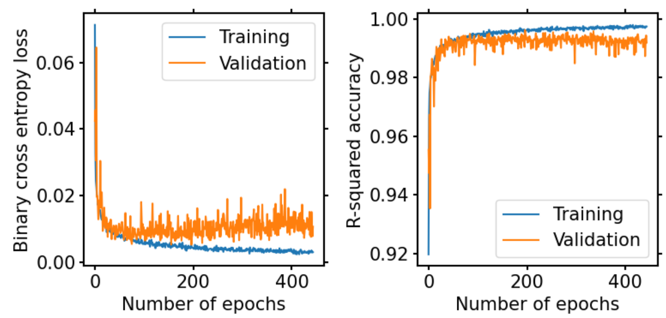


Fig. 3. Loss and accuracy for the training and validation process. Binary cross entropy loss is used to calculate the gradient and update the parameters of the neural network, and R-squared accuracy is used to illustrate the performance.

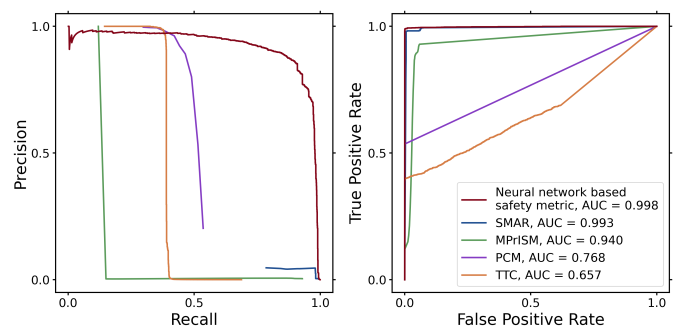


Fig. 4. ROC curve and PR curve for the proposed neural network based safety metric, SMAR, MPrISM, PCM, and TTC.

## IV. Conclusion

In this project, I develop a neural network-based safety metric using trajectory data collected through NADE and three different autonomous driving algorithms. The proposed safety metric employs an MLP with Batch Normalization as its backbone. Compared to four representative safety metrics (i.e., SMAR, MPrISM, PCM, and TTC), the neural network-based safety metric demonstrates the best performance.

For future research, additional neural network architectures, such as Transformers, which are well-suited for sequential data, could be explored. Moreover, the scenarios can be enriched to improve the representation of real-world driving environments, thereby allowing for further development of neural network-based safety metrics.

## References


[1] S. Feng, H. Sun, X. Yan, H. Zhu, Z. Zou, S. Shen, and H. X. Liu, "Dense reinforcement learning for safety validation of autonomous vehicles," Nature, pp.620-627, 2023.

[2] B. Weng, S.J. Rao, E. Deosthale, S. Schnelle, and F. Barickman, "Model predictive instantaneous safety metric for evaluation of automated driving systems," IEEE Intelligent Vehicles Symposium (IV), pp. 1899-1906, 2020.

[3] P. Junietz, F. Bonakdar, B. Klamann, and H. Winner, "Criticality metric for the safety validation of automated driving using model predictive trajectory optimization," International Conference on Intelligent Transportation Systems (ITSC), pp. 60-65, 2018.

[4] D. N. Lee, "A theory of visual control of braking based on information about time-to-collision," Perception, 5(4):437–459, 1976.

[5] E. Camacho and C. Alba, "Model predictive control," Springer Science and Business Media, 2013.

[6] D. Krajzewicz, "Traffic simulation with SUMO–simulation of urban mobility," Fundamentals of traffic simulation, pp. 269-293, 2010.

[7] F. Murtagh, "Multilayer perceptrons for classification and regression," Neurocomputing, no. 5-6, pp. 183-197, 1991.

[8] B. Mildenhall, et al., "NeRF: representing scenes as neural radiance fields for view synthesis," In: European conference on computer vision, pp.405–421, 2020.

[9] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proceedings of the 32nd International Conference on Machine Learning, 2015.

[10] X. Yan, S. Feng, D. J. LeBlanc, C. Flannagan and H. X. Liu, "Evaluation of Automated Driving System Safety Metrics With Logged Vehicle Trajectory Data," in IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 8, pp. 8913-8925, 2024.

[11] X. Yan, S. Feng, H. Sun and H. X. Liu, "Distributionally consistent simulation of naturalistic driving environment for autonomous vehicle testing," arXiv:2101.02828, 2021.

[12] S. Feng, X. Yan, H. Sun, Y. Feng, and H. X. Liu, "Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment," Nature communications, vol. 12, no. 1, pp. 748, 2021.

[13] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," Physical review E, 62(2), 1805, 2000.

[14] A. Kesting, M. Treiber, and D. Helbing, (2007). "General lane-changing model MOBIL for car-following models," Transportation Research Record, 1999(1), 86-94, 2007.