Evaluating Autonomous Vehicle Safety: A Comparative Analysis of ADS and ADAS Incident Data Using Machine Learning Models

Minh Pham

August 2024

Abstract

This paper focuses on Autonomous Vehicles (AVs), which are rising rapidly in road transportation. However, the reliability of AV systems, namely Advanced Driver Assistance Systems (ADAS) and Automated Driving Systems (ADS), remain vague. Therefore, it is essential that we improve their safety measurements. This study aims to evaluate and compare the safety performance of ADAS and ADS by training supervised machine learning models on incident report data, specifically focusing on predicting injury severity, using numerical and categorical data related to crash conditions and environmental factors. We compared different models, for each fine-tuning their respective hyperparameters. The best performance was achieve with Random Forest with hyperparameters `max_depth` = 20, `n_estimators` = 200. This work is expected to provide valuable insights for both the improvement of autonomous systems and overall road safety.

1 Introduction

As autonomous vehicles (AVs) become more integrated into modern transportation, ensuring their safety has become a critical focus of research [4]. One promising approach to enhancing AV safety is using machine learning to predict and prevent accidents. By analyzing data from past incidents, machine learning models can identify patterns and provide valuable insights into how these systems can be improved to reduce accident risks [3].

A comprehensive study by Favarò *et.al.* [2] was published in *PLoS One* aimed to promote safety and transparency by providing an in-depth analysis of autonomous vehicle (AV) accident reports between September 2014 and March 2017. This survey focused on investigating accident types, collision outcomes, damages incurred, accident locations, and accident frequencies. One of the key contributions of this work was the clear categorization of accidents based on AV involvement, offering valuable insights into how these vehicles perform in real-world scenarios. The study also underlined the importance of understanding collision dynamics, particularly in low-speed environments such as city streets, where most incidents occurred. However, this research had certain limitations. Since the data used in the analysis spanned from 2014 to early 2017, it does not capture the more recent developments in AV technology. Given the rapid evolution in autonomous driving systems, this dataset may not accurately represent current accident trends. Nevertheless, the study serves as a strong foundation for evaluating AV safety performance, providing crucial data on AV incident characteristics.

By integrating the before-mentioned accident characteristics, our study aims to build a more accurate and causal machine learning models for predicting injury severity in autonomous vehicle incidents. The Favarò *et.al.* survey directly informed the feature selection process in our analysis. Namely, we decided to integrate the categorization of accident types, including rear-end collisions, head-on collisions, and side-impact crashes. Collision outcomes were categorized by injury severity, ranging from no injuries to minor, moderate, severe injuries, and fatalities. Additionally, we incorporated key environmental factors such as weather conditions (clear, rain, fog), lighting conditions (daylight, dusk, night), and roadway surface conditions (wet, dry, icy) to provide a comprehensive analysis of the incidents.

Moreover, while the dataset used in the PLoS One study may not account for more recent developments in AV technology, it provides critical historical data that still offers valuable lessons in assessing safety trends. This project intends to bridge the gap between past and present AV performance by utilizing more current datasets while adopting the thorough analytical methods seen in previous research. By doing so, we aim to ensure that the machine learning models we develop reflect both historical trends and modern advancements in autonomous driving technology, improving the accuracy and applicability of our predictions.

2 Datasets

The dataset used for this project was obtained from the National Highway Traffic Safety Administration (NHTSA) [1]. It includes a mix of numerical and natural language data, focusing on incident reports involving autonomous vehicles. The dataset contains 1254 rows and 137 features, covering key aspects of each incident, such as vehicle details, incident circumstances, and crash outcomes. Some of the most important features in the dataset include:

- **Reporting entity details**: Name of the entity reporting the incident.
- **Vehicle information**: Make, model, and year of the vehicle involved in the accident.
- **Incident information**: Date, time, and roadway type where the incident occurred.
- **Crash description**: Object with which the vehicle collided (e.g., another vehicle, stationary object, or pedestrian).
- **Post-crash information**: Whether a law enforcement investigation was conducted or any property damage was reported.
- **Narrative**: Detailed descriptions of the pre-crash, crash, and post-crash circumstances, providing deeper context for each incident.

Additional details such as **weather conditions**, **lighting**, **roadway surface**, and **posted speed limits** are also included, helping to explain the conditions under which each incident occurred. These features, especially **Highest Injury Severity Alleged**, play a crucial role in predicting accident outcomes and are central to the machine learning models used in this study.

2.1 Dataset Characteristics

- **Number of samples (rows)**: 1254
- **Number of features (columns)**: 137
- **Number of null values**: 17,834

The dataset contains a mix of categorical and numerical data, with **Highest Injury Severity Alleged** being one of the most critical features. This feature categorizes the severity of injuries in each incident, with values such as ”Unknown,” ”No Injuries Reported,” ”Moderate,” ”Minor,” ”Serious,” and ”Fatality.” Below is the distribution of injury severity in the dataset:

- **Unknown**: 63 samples
- **No Injuries Reported**: 1007 samples
- **Moderate**: 32 samples

- **Minor**: 126 samples
- **Serious**: 21 samples
- **Fatality**: 2 samples
- **Null Values**: 3 samples (in the 'Highest Injury Severity Alleged' column)

This feature is essential for predicting accident outcomes, making it a key focus of the model's classification process.

2.2 Feature Selection and Data Preprocessing

Out of the 137 features available, we selected the most relevant ones for predicting accident outcomes. These features include both environmental and vehicle-specific variables, such as:

- Weather
- Lighting
- Roadway Type
- Roadway Surface
- Incident Date
- Incident Time (24:00)
- Posted Speed Limit (MPH)
- Crash With
- Highest Injury Severity Alleged
- CP Pre-Crash Movement
- SV Pre-Crash Movement
- Latitude and Longitude
- City and State

Rows containing null values in key features were removed to avoid introducing bias into the model. Filling these missing values with averages could skew the results, so we opted for removal. Additionally, categorical variables were encoded using **Label Encoding** to convert them into numerical form, ensuring compatibility with the machine learning models employed.

To evaluate model performance, the dataset was split into training and testing sets, using an 80/20 split. This means that 80% of the data (1003 rows) was used to train the machine learning models, while the remaining 20% (251 rows) was reserved for testing. This split ensures that the models are trained on a sufficient amount of data while preserving a portion for independent validation.

By focusing on the most relevant features and employing a robust data splitting method, we aim to build accurate models that can predict the severity of injuries in autonomous vehicle incidents.

3 Methodology/Models

For the weather feature, the dataset originally separated different weather conditions into individual columns. Each column was labeled 'Y' if that weather type was present. To simplify this, we created a new column that combined these separate columns into one, labeling the weather condition accordingly. We then used a Label Encoder to transform the weather data into numerical values so the model could process it.

Next, we reviewed the dataset and selected the most relevant features for predicting injury severity. Rows containing null values were removed to avoid bias, as auto-filling with the mode or mean could imbalance the model. Categorical columns, such as roadway type and vehicle movement, were also transformed into numerical values using a Label Encoder.

For training the models, we set X (input) as all the selected features except Severity, and y (output) as the Severity. The dataset was split, with 80% used for training and 20% for testing. We defined a function to evaluate different models, including Random Forest, K-Nearest Neighbors, Decision Tree, Multilayer Perceptron, and Logistic Regression. After comparing the accuracy, classification reports, and confusion matrices, Random Forest emerged as the optimal model. Hyperparameter tuning was conducted for each model, but Random Forest consistently provided the best performance.

4 Results and Discussion

This section presents the accuracy of different machine learning models applied to both ADS and ADAS datasets, along with a comparison of performance, hyperparameter tuning results, and relevant visualizations.

4.1 Model Performance on ADS and ADAS

The accuracy of each model on the ADS and ADAS datasets is respectively shown:

Acronyms used:

- **RF**: Random Forest
- **KNN**: K-Nearest Neighbors
- **DT**: Decision Tree
- **MLP**: Multilayer Perceptron
- **LR**: Logistic Regression

RF	KNN	DT	MLP	LR
91.88%	78.63%	86.75%	79.06%	78.21%

Table 1: Evaluation Accuracy for different ML Models

RF	**KNN**	**DT**	**MLP**	**LR**
95.98%	85.43%	89.45%	86.93%	88.94%

Table 2: Evaluation Accuracy for ADAS Dataset across different ML Models

The accuracy differences between ADS and ADAS for each model are displayed in Figure 1. As shown, the ADAS models generally outperform the ADS models by a margin ranging from 2% to 11%.

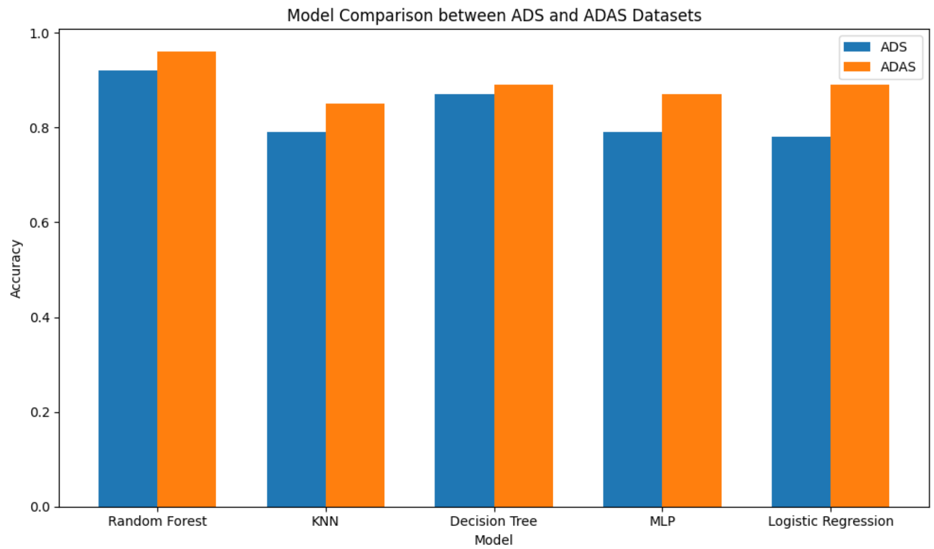

Figure 1: Accuracy comparison between ADS and ADAS models

4.2 Hyperparameter Tuning

To further optimize the models, hyperparameter tuning was conducted. The tuned results for the most relevant models are as follows:

- **Random Forest (Tuned)**: Best hyperparameters: {'max_depth': None, 'n_estimators': 200}, Accuracy: 91.88%
- **K-Nearest Neighbors (Tuned)**: Best hyperparameters: {'n_neighbors': 9, 'weights': 'distance'}, Accuracy: 89.74%
- **Decision Tree (Tuned)**: Best hyperparameters: {'max_depth': None, 'min_samples_split': 2}, Accuracy: 86.75%
- **MLP Classifier (Tuned)**: Best hyperparameters: {'alpha': 0.01, 'hidden_layer_sizes': (100,)}, Accuracy: 79.06%
- **Logistic Regression (Tuned)**: Best hyperparameters: {'C': 0.01, 'solver': 'liblinear'}, Accuracy: 78.21%

Figure 2 shows the Random Forest model's performance after tuning. Despite the increased complexity from tuning, Random Forest consistently emerged as the top-performing model across both ADS and ADAS datasets.

4.3 Discussion of Results

Overall, the Random Forest model achieved the highest accuracy for both datasets, indicating its robustness in predicting injury severity. The high accuracy could be partly attributed to the large number of "No Injury" entries in the dataset, which may have made the prediction task easier. However, this also suggests potential class imbalance, as the model may struggle to predict rarer, more severe outcomes (e.g., fatalities).

Additionally, while hyperparameter tuning improved the performance of K-Nearest Neighbors and Decision Tree models, Random Forest remained the most reliable option. One possible explanation is its ability to handle non-linear interactions between features, which are prevalent in complex datasets like those involving ADS and ADAS.

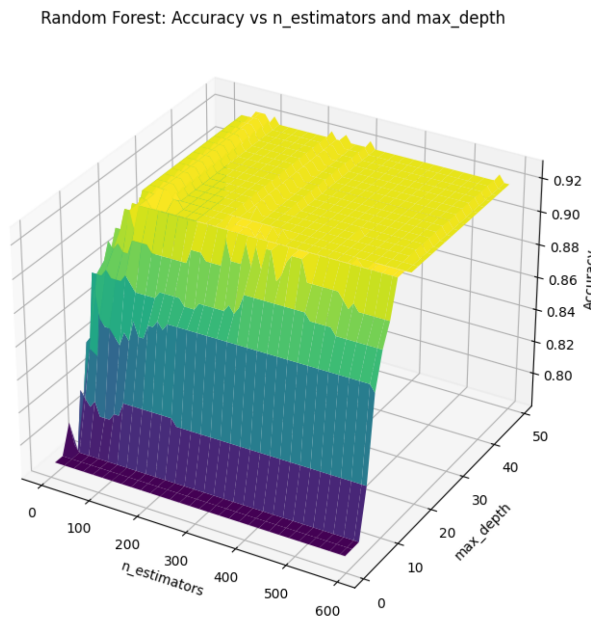

Figure 2: Random Forest performance after hyperparameter tuning

Potential limitations of this approach include the imbalance in injury severity categories, with most incidents classified as "No Injury," potentially skewing model accuracy. Future work could explore re-sampling techniques or cost-sensitive learning to address this issue. Unexpected outcomes, such as the relatively lower performance of the MLP Classifier, may be due to its sensitivity to feature scaling and the need for further tuning.

5 Conclusions

In this study, we set out to analyze the safety performance of Advanced Driver Assistance Systems (ADAS) and Automated Driving Systems (ADS) using machine learning models. By applying models such as Random Forest, K-Nearest Neighbors, Decision Tree, MLP Classifier, and Logistic Regression, we aimed to predict injury severity in vehicle incidents based on various features.

The results showed that the Random Forest model consistently performed the best, with an accuracy of 91.88% for ADS and 95.98% for ADAS. This suggests that Random Forest is highly effective at handling complex data, particularly for this type of safety analysis. Other models, while still useful, did not perform as well, likely due to their sensitivity to specific features or hyperparameter configurations.

One limitation we identified is the imbalance in the dataset, where "No Injury" cases dominate, possibly leading the models to over-predict this outcome. To address this, future research could explore re-sampling techniques or adding more data to balance the classes better.

In the future, we could also try other models, such as the Transformer, or experiment with more advanced hyperparameter tuning methods to improve accuracy. Overall, this project demonstrates the potential of machine learning in improving the safety of autonomous driving systems, but further refinements are necessary for better predictions, especially in severe cases.

6 Acknowledgements

This work was supported by the Inspirit AI program, which provided valuable guidance in enhancing my AI skills and research experience. I would also like to thank my mentor, Edoardo Contente, for his continuous support and insightful feedback throughout the project.

References

[1] National Highway Traffic Safety Administration. *Standing General Order on Crash Reporting*. Accessed: 2024-09-02. 2024. URL: `https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting#data`.

[2] Francesca M. Favarò et al. "Examining accident reports involving autonomous vehicles in California". In: *PLOS ONE* 12.9 (Sept. 2017), pp. 1–20. DOI: `10.1371/journal.pone.0184952`. URL: `https://doi.org/10.1371/journal.pone.0184952`.

[3] Sina Mohseni et al. "Practical Solutions for Machine Learning Safety in Autonomous Vehicles". In: *CoRR* abs/1912.09630 (2019). arXiv: `1912.09630`. URL: `http://arxiv.org/abs/1912.09630`.

[4] Jun Wang et al. "Safety of Autonomous Vehicles". In: *Journal of Advanced Transportation* 2020.1 (2020), p. 8867757. DOI: `https://doi.org/10.1155/2020/8867757`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1155/2020/8867757`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1155/2020/8867757`.