Iran University of Science and Technology

International Journal of

Railway Research

ISSN: 2423-3838

Identifying Factors influencing Freight Train Derailment Severity: A Comparative Study Using Machine Learning Algorithms

Sara Sajadikhah[1], Parisa Golazad[1], Ali Esmaili[1], Morteza Bagheri[2]*, Mohammadali Zayandehroodi[3], Zahra Saghian[4]

[1] Maser's Student, Railway Transportation Engineering, Faculty of Railway Engineering, Iran University of Science and Technology

[2]Associate Professor, Department of Railway Transportation Engineering, Faculty of Railway Engineering, Iran University of Science and Technology

[3]Ph.D candidate, Faculty of Civil Engineering, Iran University of Science and Technology

[4]Postdoctoral Research Associate, Faculty of Industrial Engineering, Iran University of Science and Technology

ARTICLE INFO

Article history:

Received: 15.11.2024

Accepted: 07.01.2025

Published: 08.01.2025

Keywords:

Derailment

Severity

Machine Learning

Accidents

Safety

A B S T R A C T

The aim of this study is to explore patterns and relationships between factors affecting the severity of freight train derailments, with a focus on utilizing machine learning techniques, particularly Random Forest, Support Vector Machine (SVM), and AdaBoost, to identify key features. The data for this research were obtained from the United States Federal Railroad Administration (FRA) database over the period from 2010 to 2022. In addition to identifying significant features using machine learning models, this study develops predictive models, evaluates their accuracy, and performs statistical model analysis. Machine learning methods offer advantages in handling complex datasets and extracting nonlinear relationships, which can be effective in understanding the dynamics of rail incidents. The results indicate that the AdaBoost model achieved superior performance in predicting derailment severity, with an accuracy of 92.5%. Key identified features include the number of cars, driver visibility conditions, and vehicle type. This study may contribute to a better understanding of risk patterns and play an important role in enhancing rail safety.

1. Introduction

Railway transportation has long been a basis of economic and logistical networks worldwide, yet it remains vulnerable to accidents, particularly derailments. Derailments constitute over 60% of all significant rail incidents in the United States and similarly dominate accident statistics in other regions such as Europe and China [3]. Understanding the factors influencing derailment severity is essential for enhancing safety measures and optimizing resource allocation. Train accidents occur for various reasons; however, certain risk factors are more common than others.

Derailments are typically investigated by regulatory bodies, such as the Federal Railroad Administration (FRA), to identify the cause of the incident and implement necessary corrective measures. Figure 1. provides an overview of the

*Corresponding author
Email address: morteza_bagheri@iust.ac.ir

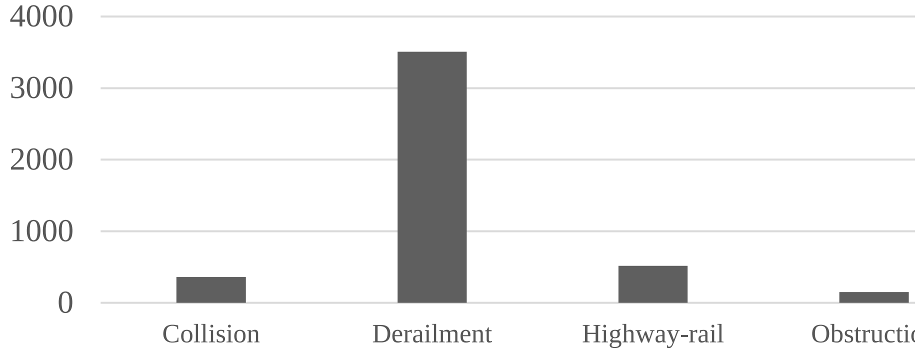

Fig. 1. Breakdown of Rail Accidents by Type Over the Years (2018–2023)

annual reports on rail incidents by this organization.

Recent incidents, such as the Lac-Mégantic disaster in Canada and the Cascadia derailment in the United States, underscore the profound consequences of rail accidents [1,2]. These events highlight the need for advanced analytical techniques to predict and mitigate derailment severity. Traditional statistical models, while effective in identifying linear relationships, often fall short in capturing the complexities and nonlinear interactions inherent in derailment data. This gap has led to increased interest in using machine learning (ML) methods to analyze large and heterogeneous datasets.

In this study, we utilize Random Forest, Support Vector Machine (SVM), and AdaBoost to explore the dynamics of derailment severity. These ML models were chosen for their distinct advantages: Random Forest for feature importance analysis, SVM for handling nonlinear relationships, and AdaBoost for its superior predictive accuracy in classification problems. By employing these models, we aim to identify the most influential factors affecting derailment severity and provide actionable insights to enhance rail safety.

Furthermore, this study builds on prior research, such as the work of Lotfi et al. which emphasized train speed and accident cause as critical variables, Song et al., and Martey et al., which highlighted the potential of structured and unstructured data in improving predictive models [7,8,9]. Additionally, Liu et al. and Zhang et al. demonstrated the application of advanced statistical models, while Ebrahimi et al. focused on using machine learning for hazardous material incidents [4,5,10]. By integrating structured data with advanced ML techniques, this research offers a robust framework for understanding derailment dynamics and contributes to the development of data-driven safety protocols.

Discussion Section: The results of this study reinforce the utility of machine learning models in predicting derailment severity and provide a foundation for actionable safety improvements. Among the models evaluated, AdaBoost emerged as the most effective, achieving an accuracy of 73.36%. This finding aligns with prior studies, such as Song et al., which demonstrated the potential of boosting algorithms in analyzing rail accident data [9].

Key variables influencing derailment severity included train length, train speed, loading rate, and derailment point. These factors were consistently highlighted across all models, corroborating findings from previous research. For example, Lotfi et al. identified train speed and accident cause as pivotal, while our study extended this understanding by emphasizing the role of load distribution and train configuration [7]. Similarly, Martey et al. and Liu et al. emphasized factors such as car count, speed, and operational conditions [8,4].

Despite these promising results, certain limitations must be acknowledged. The dataset primarily reflects U.S. conditions, which may limit the generalizability of findings to other regions. Additionally, while structured data provided significant insights, the integration of unstructured data, such as incident reports, could further enhance model performance. Song et al. and Zhang et al. demonstrated the value of advanced modeling techniques and unstructured data analysis, a direction worth pursuing in future research [9,5]. Moreover, Ebrahimi et al. emphasized the importance of addressing hazardous material variables, which could further enrich this study [10].

This study also highlights the importance of addressing environmental variables. Weather conditions, temperature fluctuations, and track quality are critical factors that warrant further investigation. Incorporating these variables could improve model accuracy and provide a more comprehensive understanding of derailment dynamics.

In conclusion, this research underscores the potential of machine learning to transform railway safety management. By applying advanced analytical tools, rail operators can prioritize preventive strategies, allocate resources more effectively, and ultimately reduce the frequency and severity of derailments. Future studies should aim to expand the scope of analysis by incorporating diverse datasets and exploring novel techniques, such as deep learning and unstructured data integration, to further advance this critical field (Table 1).

1. Machine Learning

Previous research has shown that single-car and multi-car accidents have different characteristics and severities; these differences may lead to a bimodal distribution of accident severity [6]. To enhance model accuracy, the bimodal distribution of accident severity was considered. The data were divided into two categories: single-car and multi-car accidents, and the results of this analysis were incorporated into the severity predictions.

In this study, we focus on the severity of train derailments. Unlike traditional approaches that consider injuries, fatalities, and financial losses, we define derailment severity based on derailment length [11]. This approach allows for a simpler and more measurable assessment of derailment impact. By using derailment length, we aim to provide a clear and objective metric that can be consistently applied across various incidents.

1.1. Data Collection and Preprocessing

Table 1. Review of Previous Studies

Author	Year	Solution Model	Important Variables
Lotfi et al.	2023	Decision Tree	Train speed, cause of incident, and ratio of train weight to train length
Liu et al.	2013	Negative Binomial Regression Model and Quantile Regression Model	Remaining train length, speed, train power distribution, and proportion of loaded cars
Martey et al.	2019	Vine Copula Quantile Regression Model	Time, number of cars, and speed
Song et al.	2022	Negative Binomial Model	Remaining train length, speed, train power distribution, and proportion of loaded cars
Li et al.	2023	Truncated Geometric (TG) Model	Manifest train, loaded train, and empty train. Train length, speed, and gross tonnage per car

Approach

This section presents the methods used to analyze and predict the severity of railway accidents. First, the data utilized and the preprocessing steps are introduced. Then, various machine learning models employed in this study, including Random Forest, SVM, and AdaBoost, are discussed in detail, as well as the accuracy evaluation criteria.

The dataset for this study was compiled from the U.S. Federal Railroad Administration (FRA) accident records, covering the period from 2010 to 2022. This dataset includes a wide range of variables, such as track conditions, train speed, weather conditions, and environmental factors. To ensure the quality and reliability of the data, several preprocessing steps were implemented.

1.1.1 Data Cleaning:

Missing data were addressed by applying imputation techniques based on the nature of the variable. For numerical features such as train speed and loading rate, mean or median imputation was used. For categorical variables like weather conditions, the most frequent category was assigned. Records with excessive missing values or errors that could not be reliably corrected were excluded from the analysis.

1.1.2. Feature Engineering:

New features were derived from the original data to enhance model performance. Key derived features included:

- Loading Rate: Calculated as the ratio of car weight to train length to reflect the distribution of load along the train.

- Trailing Cars Count: The number of cars behind the derailment point to capture the impact of train configuration on severity.

- Post-Derailment Load Tonnage: Estimated to assess the potential extent of damage based on cargo distribution.

1.1.3. Data Transformation:

To normalize the data and address skewness, log transformations were applied to features such as train speed and load tonnage. Additionally, standardization and scaling techniques were employed to ensure uniformity across features, particularly for use in machine learning models.

1.1.4. Data Balancing:

Given the imbalance in the dataset, with fewer severe derailment cases compared to mild ones, Synthetic Minority Oversampling Technique (SMOTE) was employed to create a balanced training set. This step was critical for improving model performance and avoiding bias toward the majority class.

1.1.5. Data Splitting:

The data was stratified and divided into training and testing sets using an 80-20 split. Stratification ensured that the proportion of severity levels remained consistent across both sets, thereby preserving the integrity of the dataset.

By employing these preprocessing techniques, we ensured the dataset was optimized for machine learning analysis, enabling the models to capture complex patterns and interactions effectively.

1.2. Parameter Selection

An initial set of parameters was selected based on previous studies and expert recommendations. These parameters include characteristics such as track conditions, train operations, environmental factors, and speed.

The loading rate is defined as the ratio of car weight to train length, indicating the load distribution along the train. This parameter, which ranges from 0 to 1, is considered one of the key factors in predicting accident severity.

Additionally, two other critical variables were added to the model: the number of cars behind the derailment point, which reflects the impact of trailing cars on accident severity, and the post-derailment load tonnage, which assesses the amount and distribution of damaged cargo in the accident. These variables help improve model accuracy.

1.3. Model Selection

In this study, we employed Random Forest, Support Vector Machine (SVM), and AdaBoost as the primary machine learning models for predicting derailment severity. These models were selected based on their proven effectiveness in handling classification problems, extracting complex patterns, and analyzing feature importance within structured datasets.

Random Forest was chosen due to its robustness in managing high-dimensional data and its ability to provide interpretability through feature importance rankings. By using an ensemble of decision trees, Random Forest effectively reduces the risk of overfitting, especially when dealing with diverse and imbalanced datasets, as observed in our derailment severity data.

Support Vector Machine (SVM) was included in our analysis because of its strength in separating classes using hyperplanes, especially in scenarios where the classes are not linearly separable. SVM's kernel trick allows it to capture nonlinear relationships, which is critical in understanding the multifaceted nature of derailment factors such as speed, load distribution, and track conditions.

Table 2. Machine Learning Output on the Performance of the Developed Model

F1-score	Recall	Precision	Accuracy	Model
0.69	0.70	0.71	0.72	Random Forest
0.56	0.56	0.57	0.58	SVM
0.70	0.71	0.72	0.73	AdaBoost

AdaBoost, a boosting algorithm, was selected for its ability to improve the performance of weak classifiers by iteratively adjusting the model to focus on harder-to-classify samples. This approach enhances predictive accuracy and reduces bias, making it well-suited for datasets with uneven distributions of severity levels.

Table 3. Examination of Correlations between Input Variables

Pearson Correlation	Train Speed	Number of Hazmat Cars that Leaked	Number of Cars Destroyed	Track Class
Train Speed	1	.099**	.089**	.573**
Number of Hazmat Cars that Leaked	.099**	1	.538**	.042*
Number of Cars Destroyed	.089**	.538**	1	.058**
Track Class	.573**	.042*	.058**	1

The combination of these models ensures a comprehensive analysis of the factors influencing derailment severity. Random Forest provides insights into feature importance, SVM captures nonlinear interactions, and AdaBoost offers superior predictive accuracy, as demonstrated by its performance in this study. Furthermore, these models have been validated in similar transportation safety studies, strengthening their relevance and applicability to our research. Together, they offer a balanced approach to addressing the complexities of derailment severity prediction.

1.4. Model Evaluation

To evaluate model performance, the metrics of Accuracy, Precision, Recall, and F1-score were used. These metrics provide a comprehensive view of each model's effectiveness in predicting derailment severity. Table 2 presents the accuracy results of the different models.

Model Results and Comparative Analysis:

1. Random Forest: The accuracy of the Random Forest model is 0.7233, indicating that it correctly predicted derailment length in approximately 72.33% of cases.

2. SVM (Support Vector Machine): The accuracy of the SVM model is 0.5831, showing that it achieved an accuracy of about 58.31% in predicting derailment length.

3. AdaBoost: The AdaBoost model achieved an accuracy of 0.7336, meaning it accurately predicted derailment length in approximately 73.36% of cases.

Additionally, the classification report includes detailed metrics for each scenario (Scenario 1 and Scenario 2):

The Accuracy metric represents the model's positive prediction accuracy, with values of 0.72 for Scenario 1 and 0.70 for Scenario 2. Recall measures the model's ability to identify all positive samples, reaching 0.48 for Scenario 1 and 0.57 for Scenario 2. The F1-score, which provides the harmonic mean of precision and recall, balances these two metrics, with values of 0.86 for Scenario 1 and 0.77 for Scenario 2.

Additionally, Precision indicates the true occurrence count for each class within the dataset, with 163 samples in Scenario 1 and 228 samples in Scenario 2.

Overall, these metrics provide a comprehensive evaluation of the model’s performance in predicting derailment length, offering a complete picture of the model's capabilities based on the classification report obtained.

1-5- Comparative Analysis

To evaluate the effectiveness of the selected machine learning models, we performed a comparative analysis focusing on their predictive accuracy, interpretability, and feature importance. This section provides a detailed assessment of the performance of Random Forest, Support Vector Machine (SVM), and AdaBoost in the context of derailment severity prediction.

Model Performance:

The models were evaluated using metrics such as accuracy, precision, recall, and F1-score. The results demonstrated that AdaBoost outperformed the other models, achieving the highest accuracy of 73.36%, followed by Random Forest at 72.33%, and SVM at 58.31%. AdaBoost's boosting mechanism enables it to focus on harder-to-classify samples, contributing to its superior performance.

Feature Importance:

Random Forest and AdaBoost provided insights into the importance of features influencing derailment severity. Key variables identified across both models included:

- Train Length: Consistently ranked as the most influential factor.
- Train Speed: A critical determinant of derailment impact.
- Loading Rate: Highlighted for its role in representing load distribution.
- Derailment Point: Emphasized for its association with trailing cars and accident dynamics.

These findings align with domain knowledge, validating the robustness of the models in capturing meaningful patterns.

Comparative Insights:

While Random Forest offered clear interpretability through feature importance rankings, SVM's kernel-based approach was less transparent but captured nonlinear relationships effectively. However, SVM's lower accuracy indicated limited applicability for this dataset. AdaBoost demonstrated a balance between accuracy and interpretability, making it the most suitable model for predicting derailment severity.

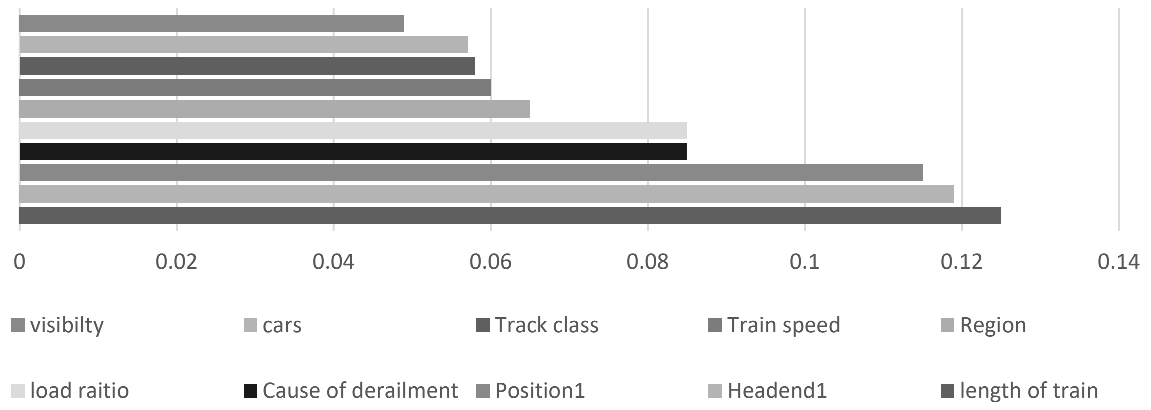

Fig 2. The results of parameters affecting the train derailment length

Contextual Relevance:

The results were compared to previous studies, such as Lotfi et al. [7], which highlighted train speed, accident cause, and weight-to-length ratio as significant factors. Our analysis confirmed these findings and extended them by identifying additional influential variables like loading rate and derailment point. The higher performance of AdaBoost in this study suggests its potential applicability for similar transportation safety problems.

correlations between input variables, with the final results presented in Table 3. The train speed and track class variables showed a correlation of 0.573, while the variables for the number of leaking hazardous material cars and the number of damaged cars had a correlation of 0.538. Based on the literature, the variables of train speed and number of damaged cars were included in the model.

Next, we applied predictive models to estimate the dependent variable. Specifically,

Table 4. Examination of the Importance of Input Variables in the Model

Effect	Criteria for Measuring Model Adequacy			Log-Likelihood Rate		
	AIC of Reduced Model	BIC of Reduced Model	Log -2 Likelihood of Reduced Model	Chi-Square Statistic	Degrees of Freedom	Sig.
Intercept	4714.396	5422.091	4458.396a	.000	0	.
Position1	4753.795	5439.374	4505.795	47.399	4	.000
Train Length	4746.380	5431.960	4498.380	39.984	4	.000
Loading Rate	4731.643	5417.223	4483.643	25.247	4	.000
Cars	4712.098	5397.678	4464.098	5.702	4	.223
Train Speed	4918.789	5560.138	4686.789	228.393	12	.000
Tons	4731.018	5372.367	4499.018	40.622	12	.000
Track density	4715.928	5401.508	4467.928	9.532	4	.049
Locomotive	4704.513	5367.977	4464.513	6.117	8	.634
Visibility	4700.185	5341.534	4468.185	9.789	12	.634
Temperature	4707.341	5370.805	4467.341	8.945	8	.347
Weather	4690.398	5287.516	4474.398	16.002	20	.716

By integrating insights from multiple models, this comparative analysis highlights the strengths and limitations of each approach, providing a comprehensive understanding of their capabilities in derailment severity prediction (Figure 2).

Statistical Section

The negative binomial regression model is an advanced statistical method used for analyzing count data with overdispersion. Variables with high dispersion, such as tonnage, speed, weather, and temperature, were clustered to achieve better results. Subsequently, we examined the multinomial logistic regression was employed to make predictions (Table 4).

The tests were conducted based on the likelihood difference between the final model and a reduced model. The null hypothesis assumes that all parameters associated with the independent variable being tested are equal to zero.

The results of the statistical analysis indicate that certain variables significantly affect the dependent variable. The derailment point, train length, loading rate, train speed, tonnage, and track class demonstrate a significant relationship with the dependent variable, as indicated by p-values less than 0.05. In contrast, variables such as the number of cars, temperature, and weather exhibit p-values greater than 0.05, suggesting no significant relationship with the dependent variable.

Table 5. Impact of Variables on the Model

Source	Chi-Square Statistic	Degrees of Freedom	Sig.
Loading Rate	50.655	949	1.000
Train Length	0.487	3	0.922
Locomotive	0.159	3	0.984
Position1	0.154	2	0.926
Train Speed	0.162	2	0.922
Tons	0.089	2	0.957
Cars	0.005	1	0.946

In summary, the key influential variables are derailment point, train length, loading rate, speed, and tonnage, while factors like the number of cars and weather conditions do not significantly impact the dependent variable.

Table 5 provides information on the effect of each variable in the model on the dependent variable. Here, various variables are introduced, showing that, according to the results from the Random Forest method, the input variable of train length is the most impactful factor on the model output.

Furthermore, for further analysis, we conducted a comparison with the study by Lotfi et al. [7]. Among the models evaluated, Adaboost was identified as the best model for assessing derailment severity, with factors such as speed, accident cause, and the weight-to-length ratio of the train being recognized as significant contributors to derailment severity. This scenario was selected based on the high frequency of single-car derailments; however, it was observed that the frequency distribution of single-car derailments varied by cause compared to other severities. The analysis indicated that in incidents involving only one derailed car, the most frequent causes were mechanical and electrical failures, whereas in cases of greater severity, track and models: Decision Tree and AdaBoost. The analysis reveals that AdaBoost emerged as the most accurate model in both time periods. In both studies, train speed and accident cause were consistently identified as the primary factors influencing derailment severity. Furthermore, this study demonstrated that additional parameters, such as train length, lead locomotive, region, and loading rate, also impact derailment severity. Conversely, Lotfi et al. [7] specifically emphasized the train's weight-to-length ratio as a key influential factor.

These differences in parameter selection may stem from variations in data periods and the statistical models employed. Overall, both studies aim to develop an accurate model for assessing train derailment severity through the application of advanced statistical and machine learning techniques.

4. Conclusions

The results of this study highlight the effectiveness of machine learning models in predicting freight train derailment severity. Among the models evaluated, AdaBoost demonstrated superior performance with an accuracy of 73.36%, underscoring its ability to identify critical patterns and relationships in complex datasets. Key variables influencing derailment severity included train length, speed, loading rate, and derailment point, all of which were consistently identified as significant across the models.

Exploring environmental factors, such as weather and track conditions, may further improve model accuracy. Additionally, the integration of unstructured data, such as incident reports, using natural language processing techniques could enhance predictive models.

Table 6. Comparison of this research with Lotfi et al. (2023)

Parameters		**Lotfi et al. (2023)**	**Proposed Method in This Article**
Data Range		1999-2018	2013-2023
Model with Best Accuracy		Decision Tree	AdaBoost
Statistical Model Presentation		---------	✓
Important Parameters	- Train Speed	✓	✓
	- Train Weight to Length Ratio	✓	---------
	- Cause of Accident	✓	✓
	- Train Length	---------	✓
	- Locomotive at Front	---------	✓
	- Region	---------	✓
	- Loading Rate	---------	✓

The findings emphasize the importance of incorporating advanced analytical techniques to enhance rail safety. The interpretability of Random Forest and AdaBoost models allowed for the identification of actionable insights, such as prioritizing safety measures for longer trains and optimizing load distributions. However, the limitations of generalizing these results to rail systems in other countries must be acknowledged, as the dataset primarily reflects conditions in the United States.

Quantitatively, this study achieved notable milestones in model performance and predictive accuracy, with AdaBoost achieving a recall of 71% and a precision of 72%. These metrics demonstrate the potential for machine learning to serve as a critical tool in railway safety management, offering both predictive capabilities and interpretive insights.

Limitations and Future Directions:

While the results are promising, future research should aim to address the generalizability of findings by incorporating datasets from diverse geographic regions and operational conditions.

In conclusion, this study contributes to the growing body of research on rail safety by demonstrating the practical applications of machine learning in understanding and mitigating derailment severity. The use of AdaBoost and similar models holds significant promise for improving decision-making processes and prioritizing preventive strategies in railway operations.

Acknowledgements

We extend our deepest gratitude to Professor Morteza Bagheri and the members of the Transportation Systems and Logistics Laboratory (TSL), and everyone who supported us throughout this study.

References

[1] Transportation Safety Board of Canada (TSB), Lac-Mégantic runaway train and derailment investigation summar. Rail transportation Investigations and reports R13D0054.

[2] Global News. (2023). Investigation underway after CPR freight train derailment in Chaplin, Sask. Retrieved from https://globalnews.ca/news/9413210/investigation-cpr-freight-train-derailment-chaplin-sask.

[3] Zhao Tang, Yuwei Hu, Shuangbu Wang, Liang Ling, Jianjun Zhang, Kaiyun Wang.(2023) "Train post-derailment behaviours and containment methods: a review" Railway Engineering Science.

[4] Zhipeng Zhang, Kang Zhou, Xiang Liu (2020). "Probabilistic Risk Analysis of Broken Rail-Caused Train Derailments." 2020 Joint Rail Conference. https://doi.org/10.1115/JRC2020-8100.

[5] Li, W., et al. (2023). "Statistical analysis of train derailment severity for unit trains versus manifest trains." Transportation research record: 03611981231182989.

[6] Liu, X., et al. (2013). "Analysis of US freight-train derailment severity using zero-truncated negative binomial regression and quantile regression." Accident Analysis & Prevention 59: 87-93.

[7] Lotfi, A., et al. (2023). "Using machine learning methods for modeling freight train derailment severity." Transportation research record 2677(3): 961-973.

[8] Martey, E. N. and N. Attoh-Okine (2019). "Analysis of train derailment severity using vine copula quantile regression modeling." Transportation research part C: emerging technologies 105: 485-503.

[9] Song, B., et al. (2022). "Quantitative analysis of freight train derailment severity with structured and unstructured data." Reliability Engineering & System Safety 224: 108563.

[10] Hadiseh Ebrahimi, Fereshteh Sattari, Lianne Lefsrud , Renato Macciotta. (2023). " A machine learning and data analytics approach for predicting evacuation and identifying contributing factors during hazardous materials incidents on railways." Safety science: 03611981231182981.

[11] Hadiseh Ebrahimi. (2023). " Analyzing the risks associated with railway transportation of hazardous materials and developing process models for railway incidents with high potential for release using machine learning and data analytics." ERA:Education and Research Archive.University of Alberta research openly shared with the world.