Research Article

Journal of Supply Chain Engineering and Logistics Optimization

Identifying Risk Factors and Improving Rail Safety by Using Logistic Regression and Ensemble Learning Approaches

Habeeb Mohammed[1], Rongfang Liu[2]* & Liu Lv[3]

[1]*Department of Industrial & Systems Engineering, North Carolina A&T State University, Greensboro, North Carolina, 27411*
[2]*Department of Marketing and Supply Chain Management, North Carolina A&T State University, Greensboro, North Carolina, USA*
[3]*Transportation Institute, North Carolina A&T State University, Greensboro, North Carolina, 27411*

***Corresponding author:** Rongfang Liu, Department of Marketing and Supply Chain Management, North Carolina A&T State University, Greensboro, North Carolina, USA.

Submitted: 15 December 2025 **Accepted:** 22 December 2025 **Published:** 29 December 2025

https://doi.org/10.63620/MKJSCELO.2025.1008

Citation: *Mohammed, H., Liu, R., Lv, L. (2025). Identifying Risk Factors and Improving Rail Safety by Using Logistic Regression and Ensemble Learning Approaches. J of Sup Cha Eng and Log Opt, 1(2), 01-07.*

Abstract
Train derailments pose significant risks to safety, infrastructure, and the environment. Understanding the factors that contribute to these incidents is crucial for developing effective prevention and mitigation strategies. This study employs logistic regression to investigate the predictors of train derailments using the accident dataset from the Federal Railroad Administration (FRA). The model identified track type and presence of engineers as significant factors influencing derailment risk. Yard tracks, industry tracks and sidings were found to have higher odds of derailments compared to main tracks, emphasizing the need for targeted safety measures in these areas. Also, the presence of engineers was associated with reduced derailment odds, highlighting the importance of skilled crew in ensuring safe operations. This study also employs adaptive boosting, an ensemble learning technique to predict derailment accidents. The model accurately predicts 72% of all instances of derailment and non-derailment accidents. The learning model also identifies the gross tonnage of the train as a key factor in predicting the likelihood of the train derailing. These findings provide valuable insights for developing evidence-based interventions by railroad authorities and safety agencies to mitigate derailment risks and enhance railway safety.

Keywords: Derail, Railroad Safety, Accident Prediction, Machine Learning.

Introduction

Among many modes of transportation, the rail system is one of the safest, with a relatively low rate of serious accidents compared to road transportation. Train accidents per million miles have decreased significantly over the past few decades, demonstrating the efforts of various stakeholders and technological advances to improve rail safety. However, in 2023, there were still 858 fatalities related to railroad linemen in the United States, and 5,481 people were injured in various rail accidents, including passengers, railroad employees, and others. [1] This shows the importance of continuous improvement in rail transportation safety. Among various safety issues, derailment is one of the most serious threats, with more than a thousand derailments occurring each year on the U.S. rail transportation network. Derailment refers to the deviation of a train from the track, which may cause serious damage, injury or death. The financial impact of derailment can be significant, including costs associated with infrastructure repairs, rolling stock damage, and potential environmental cleanup. In 2022, the average cost of each derailment was estimated to be $5 million, but major derailments can result in costs of more than $50 million. Major derailments, while rare, can have serious consequences. For example, the 2015 Amtrak derailment in Philadelphia killed eight people, injured more than 200 people, and caused losses of more than $200 million. [2] The 2023 train derailment in Ohio did not result in direct deaths, but of the 51 derailed cars, 11 were tank cars that dumped 100,000

US gallons (380,000 L) of hazardous materials, including vinyl chloride, benzene residue, and butyl acrylate [3].

The huge property damage, inestimable environmental impact, and psychological damage caused by derailment accidents are not something that anyone or any organization can ignore, requiring us to continue to study and do our best to reduce the occurrence of accidents. Rail safety, especially derailment prevention, is a complex issue that requires a holistic approach. The rail industry can significantly reduce the risk of derailment by addressing track and infrastructure design, maintenance, vehicle monitoring, operation management, and human factors. Adequate training of crew numbers, including the provision of professional engineers, also plays a vital role in ensuring safe and efficient operations.

Literature Review

This literature review synthesizes recent studies on railway derailments, focusing on derailment causes, prediction models, mitigation strategies, and emerging technologies.

The causes of railway derailments include track and infrastructure factors, vehicle factors, and human factors. The study by Zhu et al. [4] emphasized the impact of track irregularities on vehicle stability, and Li et al. [5] demonstrated that regular maintenance and timely repairs can significantly reduce the risk of derailment. It shows that the design and construction defects of railway tracks and improper routine maintenance will increase the risk of derailment. Zhai et al. [6] analyzed the impact of wheel-rail contact force on the probability of derailment. The study by Chen and Zeng [7] emphasized the importance of vehicle suspension systems in maintaining stability. These studies reveal the relationship between train vehicles and derailment events, emphasizing the interaction between wheels and tracks and the impact of vehicle design on the occurrence of derailment.

Wang et al. [8] studied the over speeding behavior in railway operation, and a study found a correlation between over speeding and derailment events. Zhang et al. [8] explored operational errors (signal errors and braking errors) to show the relationship between daily management and train operation and railway derailment. In the field of predictive simulation, machine learning, and various emerging technology environments, Xu et al. [10] used advanced simulation tools such as multibody dynamics and finite element analysis to simulate derailment scenarios. Liu et al. [11] used machine learning techniques to predict derailment events based on historical data. Automated track inspection technologies such as ultrasonic and laser scanning can enhance the detection of defects [12]. Wang et al. [9] discussed the benefits of on-board monitoring systems in preventing derailments. The implementation of a real- time vehicle monitoring system can detect anomalies in advance. Huang et al. [13] showed the application of AI in anomaly detection and risk assessment. AI and machine learning algorithms can enhance predictive maintenance and derailment prevention. Li et al. [14] explored the integration of the Internet of Things into railway systems to improve safety.

Railway derailment research covers a wide range of factors, from track and vehicle conditions to operational practices and human factors. Advances in predictive models, simulations, and emerging technologies offer promising avenues for improving rail safety. Continuous research and innovation are essential to developing effective derailment prevention and mitigation strategies and ensuring the safety and reliability of rail systems. Current research lacks a complete and comprehensive analysis of railway safety accident data. In FRA's nearly fifty years of accident records, there is still a large amount of data available for research that has not been fully mined, such as the relationship between accidents and speed, and whether the requirement for the number of on-the-job engineers has a positive impact on the occurrence and handling of accidents. This is exactly the focus of this article.

Methodology

Selected variables with categorical responses are transformed into dummy variables with numerical responses, considering the statistical method that is being adopted. Factors that influence derailment accidents are obtained using the maximum likelihood estimates from the logit function. The next step involves using adaptive boosting ensemble leaning technique to predict the occurrence of derailment accidents. Based on this, recommendations are made to aid in the management and control of rail accidents.

Logistic Regression

This statistical technique allows the prediction of a discrete outcome from a set of predictors that may be continuous, discrete, dichotomous, or mixture of them [15]. Discrete variables represent countable values, often whole numbers, for instance the number of engineers on a train. Continuous variables, on the other hand, can take on any value within a range, including decimals, such as gross tonnage of a freight train. Dichotomous variables are a specific type of discrete variable with only two possible categories, like whether the train has a caboose, or whether the train is running on a main line. In this case a yes/no response is required. According to Tabachnick and Fidell [15], logistic regression (LR) is also suitable when there is a nonlinear relationship between the responses of the dependent variables (DV) and at least one of the independent variables (IVs). For instance, in this case, the probability of the occurrence of a derailment accident may be a little affected by a 10-mph difference when a train is travelling at a low speed (e.g. 30 vs. 40), but the probability may change more significant with an equivalence difference at a high travel speed (e.g. 120 vs. 130).

Logistic regression would also help us investigate which of the variables predict the outcome as well as how each of them affects, i.e. increase or decrease the probability of the outcome. Higher order interactions (two or more) between predictors are not considered in our model to avoid the complications that come with it. Plus, there's no guarantee that the model performance will significantly be improved. The model is evaluated using the likelihood ratio estimates of the predictors in the full model and is run in Python using Google Collaboratory® environment. The likelihood ratio test is used to test the significance of each predictor to the model at alpha (a) value of 0.05.

Adaptive Boosting Ensemble Learning

Ensemble learning is a machine learning meta-approach that combines predictions from several models to improve predictive performance. Boosting algorithms essentially integrate multiple weak classifiers into a strong classifier typically using decision trees [16]. Boosting methods provide good results for both classification and regression problems in supervised learning. According to Meng et al. [17], AdaBoost works better in classification problems hence it is used as a classifier to determine predict the occurrence of a derailment accident.

AdaBoost is a learning model that adjusts each instance by applying more weight to erroneously categorized instances. Our data is structured and large enough for the model to learn and be able to predict derailment accidents while reducing bias and variation. The prepared data is split into a training set and a testing set in an 80:20 ratio. The algorithm learns the data using the training set and then makes predictions. This is achieved by fitting a series of weak classifiers to several weighted training data [18]. Incorrectly predicted observations are assigned a greater weight in the next iteration. This process goes on until the specified number of models is reached. The final model is the weighted sum or a linear combination of the various weak learners thereby creating a stronger and more robust classifier [19]. The framework for AdaBoost is illustrated in Figure 1. To prevent overfitting and assess the generalizability of the model, the data is divided into training and testing set (80:20). Also, cross-validation and stratification is used in the sampling process which repeatedly splits the data into training and validation sets to obtain a more robust estimate of model performance.

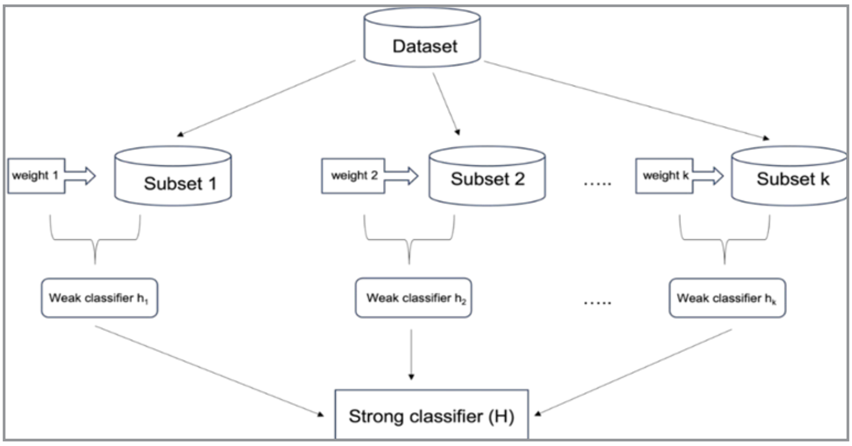

Figure 1: Framework for AdaBoost method

Analysis and Interpretation

In this section, logistics regression and adaptive boosting are used to predict the occurrence of derailment accidents. Factors influencing the likelihood of these accidents are also examined independently to obtain insights into the relationship between the accident and the train's operating conditions.

Data Description

The data used in the analysis is obtained from the FRA Rail Equipment Accident/Incident Database from 1975 to 2022 [20]. All classes of railroads are considered across all states including freight and passenger trains. The data obtained contains 216,141 observations/incidents and 160 features. The dataset is complex with records predominantly in text. Also, it is imbalanced, and variables are mostly categorical with high dimensions. The data also contain unstructured text used in the narration of incidents. The more descriptive representation of the dataset is shown in Figure 2.

Feature Elimination

Based on the criteria used by Meng et al. [17], correlated, redundant, and sparse features are removed from the dataset. For instance, correlated features are those that are more than 80% correlated with one another, e.g. loaded freight cars and gross tonnage. Redundant are those whose information is inherent in another feature, e.g. state and county. Features with more than 80% of their values missing are considered sparse and are therefore removed from the dataset e.g. Adjunct code 2. Of the 160 features in the dataset, four are selected for this analysis.

Data Preparation

Outliers within the data are first identified using boxplots. ENGINEERS is identified to have several outliers therefore those observations were removed from the dataset. The deletion approach was used to handle missing responses within the features selected. The justification for this approach is that it doesn’t significantly affect the sample size for the analysis. Lastly, Features such as WEATHER, are categorical and are therefore transformed into dummy variables with binary encoding.

Logistic Regression

Assumptions

Before proceeding with the logistic regression, we must ensure that all the assumptions related to the procedure are met:

i. Linearity: presence of a linear relationship between the log-odds of the outcome (Derailment) and the predictors.

ii. Independence of observations: the outcome of one observation does not influence that of another.

iii. No multicollinearity: this assumption is checked using the VIF value and all of them were below 5 suggesting just a moderate correlation which will not be of major concern.

iv. No outliers: to prevent outliers from unduly influencing the model estimates, all observations with outlier values were removed from the dataset.

v. Large sample size: sample size is sufficiently large so estimates can be more reliable and ensures that the model generalizes well to new data

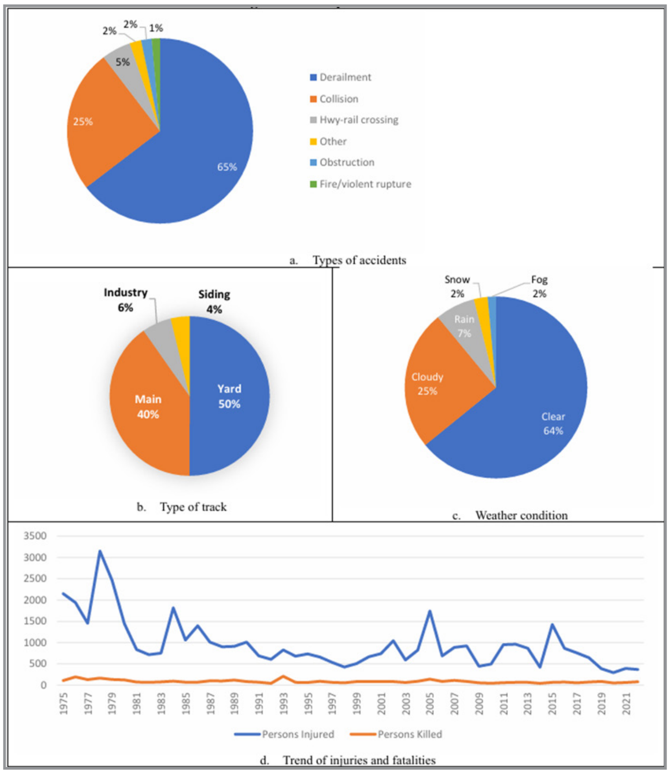

Figure 2: Description of dataset

The logistic regression procedure was run in Python using the Google Collaboratory platform. All predictors are entered into the model at the initial stage to improve the overall fit of the model to the data, and to capture more of the underlying relationships between the predictors and the outcome variable. The parameter estimates of all the variables as well as their odds ratio is presented in TABLE 1 . The coefficients represent the change in the log-odds of derailment associated with a one-unit change in the predictor, holding other predictors constant. For example, the coefficient for "Engineers" is -0.2209, indicating that an increase in the number of engineers is associated with a decrease in the log- odds of derailment.

Table 1: Maximum likelihood and Odds Ratio Estimates

Parameter	Maximum likelihood estimates		Odds ratio of estimates		
	Coefficient	P>|z|	Odds Ratio	95% Confidence limits	
const	0.7151	0.000	2.044	1.893	2.208
Clear weather	-0.0955	0.000	0.909	0.886	0.932
Main track	-0.2163	0.000	0.805	0.785	0.827
Gross Tonnage	0.0002	0.000	1.000	1.000	1.000
Engineers	-0.2209	0.000	0.802	0.747	0.860

The association of predicted probabilities and observed responses is assessed using various measures like concordance, Somers' D, and Gamma. The relatively high percentage of concordant pairs (69.75%) and the positive values of Somers' D (0.34) and Gamma (0.35) indicate a good discriminatory ability of the model.

AdaBoost Ensemble Method

Hyperparameter Tuning

There are several parameters that influence the prediction performance of any ensemble model. Hyperparameter tuning is a prior step used to find optimal hyperparameter values that produce the best classification accuracy. Grid search was used to find the optimal values for these parameters. The parameters of interest in the AdaBoost classifier model are:

i. Base estimator: the algorithm used to train the model. A decision tree classifier is used as base estimator for our model.

ii. Number of estimators: the number of models that the algorithm trains. This was set at 300, based on the results of the grid search.

iii. Learning rate: this is the contribution of each model to the weight of the final model. This was set to 1.5, based on the results of the grid search.

Results

Training and testing of the data is done with a 5-fold cross-validation while stratifying the response variable into homogeneous subgroups to ensure representative sampling and reduction of bias.

Four performance criteria are used to assess the model's performance, namely: Accuracy, Precision, Recall, and F1- score. Accuracy is the base metric often used to evaluate model performance. It is the ratio of correct predictions to the total number of predictions. Precision measures the rate of positive prediction. It is the ratio of true positives to the sum of true positives and false positives. Recall measures the number of positives classified correctly out of the total number of positives in the dataset. It is also known as Sensitivity. F1-score combines the Precision and Recall values into a mean value and is given by:

$$2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The results of the four evaluation criteria are summarized in TABLE 2. The confusion matrix in Figure 3 shows the model's performance in predicting train derailment. AdaBoost performs well in predicting derailment accidents with an accuracy of 72%.

Table 2: Classification report of AdaBoost model.

	Precision	Recall	F1-score	Support
Derailment	0.73	0.96	0.83	22859
Non-Derailment	0.63	0.16	0.25	9816
Accuracy	0.72	32675		
Macro average	0.68	0.56	0.54	32675
Weighted average	0.70	0.72	0.65	32675

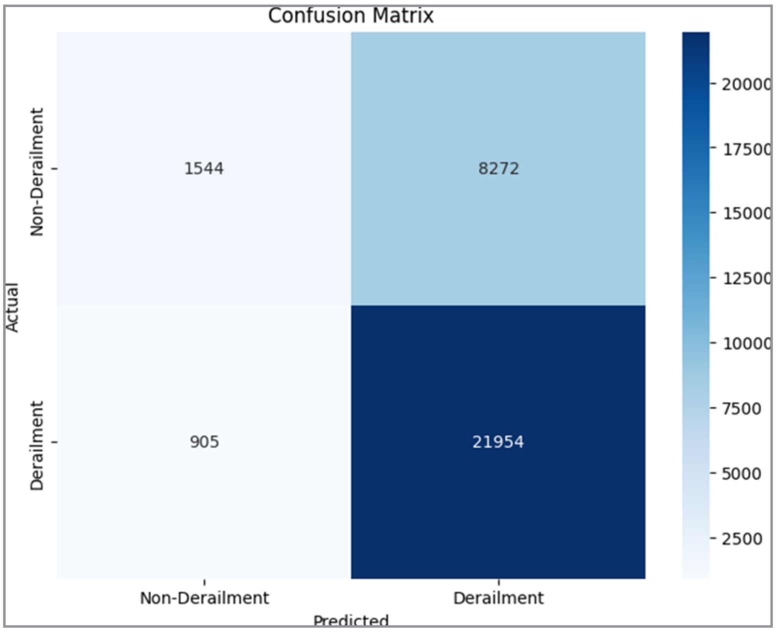

Figure 3: Confusion matrix

The Area Under the Curve (AUC) in Figure 4 summarizes the model's overall discriminatory power. In this case, the AUC is 0.68, which is considered a moderate to good performance. It suggests that the model has a 68% chance of correctly classifying a randomly chosen derailment instance higher than a randomly chosen non-derailment instance.

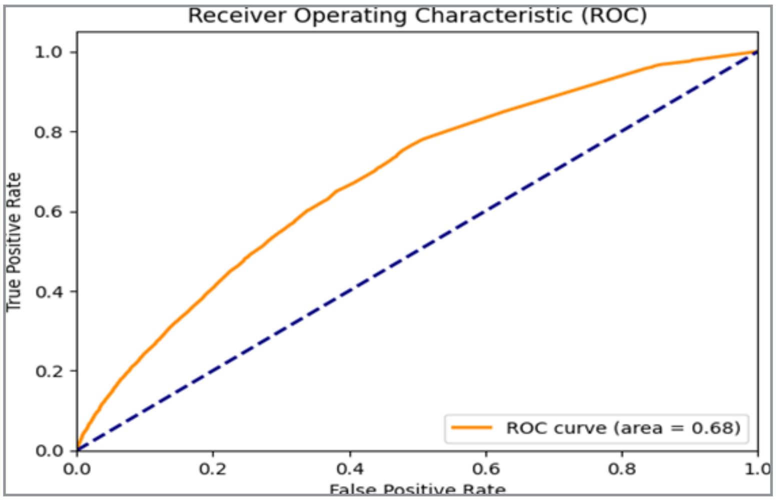

Figure 4: ROC curve for derailment prediction

The key factors that influence the prediction of derailment accidents are also investigated. The features and their contribution to the AdaBoost prediction are obtained and ranked according to their importance. TABLE 3 shows the features and their importance to the prediction model.

Table 3: Important features of derailment accidents

Feature	Description	Importance
TONNAGE	Gross tonnage of the train	0.975
MAIN TRACK	Type of track the train was travelling on	0.015
CLEAR WEATHER	Visibility at the time of accident	0.005
ENGINEERS	No. of Engineers on the train	0.005

According to TABLE 3, TONNAGE is the most important factor in determining the likelihood of a train to derail while travelling, followed track types, which is a distant second. Visibility and the number of engineers on the train have little impact in determining whether a train will be involved in a derailment accident.

Discussion

The logistic regression analysis reveals insightful patterns regarding factors associated with train derailments. The model, demonstrating good fit, highlights significant influences of weather, type of track, and presence of engineers. The odds ratio estimates provided insights into the direction and magnitude of the predictors' effects. Notably, 'Clear weather' and 'Main track' were associated with decreased odds of derailment, suggesting that derailments are more likely to occur in adverse weather conditions or on tracks other than the main track. Conversely, 'Gross Tonnage' exhibited a positive association, implying that heavier trains are at a higher risk of derailment.

The presence of engineers is linked to a significant decrease in derailment odds, suggesting the importance of skilled crew members in mitigating risk. This finding might warrant further investigation into the specific mechanisms through which engineers contribute to safety, potentially informing crew training and operational procedures. Gross tonnage presents a complex relationship with derailment risk. It can be inferred that higher gross tonnage elevates derailment odds, probably due to the increased momentum and inertia of heavier trains. This suggests a need for balancing speed and cargo weight to optimize safety.

In the second part of the analysis, the ensemble learning model demonstrates a clear proficiency in predicting derailments (class 1) compared to non-derailments (class 0). For the "Derailment" class, the model boasts a precision of 0.73, signifying that 73% of instances predicted as derailments were indeed correct. The recall of 0.96 for this class indicates that the model successfully identified 96% of actual derailments. The F1-score of 0.83 provides a balanced measure of precision and recall, suggesting a good overall performance in predicting derailments.

In contrast, for the "non-derailment" class, the model shows a precision of 0.63, meaning 63% of instances classified as non-derailments were accurate. However, the recall of 0.16 indicates that the model only captured 16% of actual non-derailment events. The F1-score of 0.25 for this class reflects a substantially lower overall performance compared to the "Derailment" class. The overall accuracy of 0.72 implies that the model correctly classified 72% of all instances, encompassing both derailments and non-derailments. This is further supported by the confusion matrix in Figure 3. However, the macro average and weighted average metrics, which account for class imbalance, reveal a discrepancy in performance between the two classes. The weighted average, which considers the support of each class, highlights the model's bias towards predicting derailments due to their higher frequency in the dataset.

Summary

To further improve the safety of rail transportation, it is important to base decisions on data. This is especially important given the significant progress that has been made in big data analytics. This study examined the FRA accident data collected over a 47-year period and a dominant occurrence of derailment accidents is discovered. Researchers carry out two main studies, first to identify the factors influencing derailment accidents and secondly apply machine learning techniques to predict the occurrence of derailment accidents. Logistic regression revealed that the number of engineers on board and the type of track are the most influential factors in assessing the likelihood of a train derailing. Further analysis using adaptive boosting to predict its occurrence reveals that the gross tonnage carries the largest share of information in predicting the occurrence of a derailment accident.

Overall, the ensemble learning model appears to be more adept at predicting derailments than non- derailments. There's a significant room for improvement in identifying non-derailment events, as indicated by the lower recall for this class. This imbalance could be attributed to the data distribution, where derailments might be over-represented compared to non-derailments. The model's overall performance is decent, with respectable accuracy and strong performance in detecting derailments, which is arguably a more critical aspect given the potential consequences of such events. Also, the logistic regression model successfully identified several factors significantly associated with train derailment. The findings highlight the importance of weather conditions, track type, train weight, and crew composition in influencing derailment risk. These insights could inform targeted interventions and strategies to enhance railway safety and reduce the incidence of derailments. Further research could explore additional predictors, potential interactions between variables, and the development of more refined predictive models to support proactive risk management in the railway industry.

Acknowledgments

Habeeb Mohammed: Investigation, Methodology, Analysis and interpretation of results, Writing-original draft.
Rongfang Liu: Conceptualization, Funding acquisition, Supervision, Writing -review & editing, Validation.
Liu Lv: Analysis, Validation, Writing - original draft. All authors reviewed the results and approved the final version of the manuscript.

References

1. Federal Railroad Administration. (2024). Railroad safety statistics annual report 2023. U.S. Department of Transportation. https://www.fra.dot.gov

2. Federal Railroad Administration. (2023). Rail equipment accident/incident data (Form 54). U.S. Department of Transportation. https://data.transportation.gov/Railroads/Rail-Equipment-Accident-Incident-Data-Form-54-/85tf-25kj/about_data
3. Gao, Y., Liu, H., & Wang, J. (2019). Operational best practices for derailment prevention. Transportation Safety Journal, 28(3), 165–178.
4. Huang, S., Zhang, Y., & Wang, X. (2020). Application of artificial intelligence in railway safety. Journal of Intelligent Transportation Systems, 14(2), 99–111.
5. Li, J., Wang, Y., & Zhang, H. (2018). Track maintenance practices and derailment risk reduction. Railway Maintenance Journal, 22(1), 47–58.
6. Li, X., Chen, Y., & Huang, J. (2020). IoT-based monitoring systems for railway safety. Sensors and Actuators Journal, 13(2), 89–102.
7. Liu, Q., Zhou, W., & Feng, S. (2018). Derailment prediction using machine learning techniques. Journal of Transportation Engineering, 144(5), 123–134.
8. Meng, H., Tong, X., Zheng, Y., Xie, G., Ji, W., & Hei, X. (2022). Railway accident prediction strategy based on ensemble learning. Accident Analysis & Prevention, 176, 106817. https://doi.org/10.1016/j.aap.2022.106817
9. National Transportation Safety Board. (2016). Derailment of Amtrak passenger train 188, Philadelphia, Pennsylvania, May 12, 2015 (Report No. RAR-16/02). https://www.ntsb.gov/investigations/AccidentReports/Reports/RAR1602.pdf
10. National Transportation Safety Board. (2024). Norfolk Southern Railway train derailment and hazardous materials release, East Palestine, Ohio, February 3, 2023 (Report No. RIR-24/05). https://www.ntsb.gov/investigations/AccidentReports/Reports/RIR2405%20CORRECTED.pdf
11. Sarkar, D., & Natarajan, V. (2019). Ensemble machine learning cookbook: Over 35 practical recipes to explore ensemble machine learning techniques using Python. Packt Publishing.
12. Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Pearson.
13. Walker, K. W., & Jiang, Z. (2019). Application of adaptive boosting (AdaBoost) in demand-driven acquisition prediction: A machine-learning approach. Journal of Academic Librarianship, 45(3), 203–212. https://doi.org/10.1016/j.acalib.2019.03.004
14. Xu, G., Liu, X., & Wu, D. (2019). Simulation techniques for derailment analysis. Railway Dynamics Journal, 31(3), 207–221.
15. Yang, L., Zhang, R., & Wang, T. (2018). Automated track inspection technologies for derailment prevention. Journal of Railway Technology, 29(1), 77–90.
16. Zhai, W., Zhu, S., & Chen, J. (2019). Wheel–rail interaction and derailment mechanisms. Vehicle System Dynamics, 57(6), 803–822. https://doi.org/10.1080/00423114.2018.1488978
17. Zhang, H., Wang, S., & Li, X. (2020). Human factors in railway operations and derailment risk. Human Factors Journal, 32(4), 297–314.
18. Zhu, X., Liu, P., & Wang, Y. (2017). Impact of track geometry on railway safety. Transportation Research Record, 2563(1), 103–111. https://doi.org/10.3141/2563-12