IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 25 April 2023, accepted 8 May 2023, date of publication 10 May 2023, date of current version 17 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3275022

RESEARCH ARTICLE

Short-Term Arrival Delay Time Prediction in Freight Rail Operations Using Data-Driven Models

JUAN PINEDA-JARAMILLO[1], FEDERICO BIGI[1], TOMMASO BOSI[2], FRANCESCO VITI[1], AND ANDREA D'ARIANO[2]

[1]Department of Engineering, University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg
[2]Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University, 00154 Rome, Italy

Corresponding author: Juan Pineda-Jaramillo (juan.pineda@uni.lu)

This work was supported by the Luxembourg National Research Fund (FNR) through the Project "ANticipatory Train Optimization with Intelligent maNagEment (ANTOINE)" [The outcome of a collaborative effort between the University of Luxembourg and CFL Multimodal (Société nationale des chemins de fer luxembourgeois)] under Grant BRIDGES2020/MS/14767177/ANTOINE.

ABSTRACT Despite rail's growing popularity as a mode of freight transportation due to its role in intermodal transportation and numerous economic and environmental benefits, optimizing all aspects of rail infrastructure use remains a significant challenge. To address this issue, various methods for developing train disruption prediction models have been used. However, these models continue to struggle with accurately predicting short-term arrival delay times, as well as identifying the causes of delays and the expected impact on operations. The lack of information available to operators makes it difficult for them to effectively mitigate the effects of disruptions. The goal of this study is to investigate a set of data-driven models for the short-term prediction of arrival delay time using data from the National Railway Company of Luxembourg of freight rail operations between Bettembourg (Luxembourg) and other nine terminal stations across the EU, and then investigate the effects of the features associated with the arrival delay time. For our dataset, the lightGBM model outperformed other models in predicting the arrival delay time in freight rail operations, with departure delay time, trip distance, and train composition appearing to be the most influential features in predicting the arrival delay time in the short-term. The National Railway Company of Luxembourg can use the short-term prediction model developed in this study as a decision-support system. For example, knowing a train's arrival delay time allows you to estimate future operational time, providing more support to reduce disruptions and subsequent operational delays via a simple web service.

INDEX TERMS Data-driven models, delays forecasting, freight transport, gradient boosting, rail operation delays.

I. INTRODUCTION

The freight transportation industry is constantly changing, and rail transportation is becoming a more popular option due to its advantages in terms of operational costs, efficiency, reliability, emissions, and safety. This trend has resulted in the gradual integration of rail into intermodal transportation, with public agencies encouraging a shift away from other alternatives [1]. As rail intermodal operations become more important for the efficiency and dependability of the freight transport industry, optimizing all aspects of rail infrastructure is critical. This includes ensuring that the infrastructure is well-maintained, properly managed, and capable of meeting the current transportation system's demands. Furthermore, using technology and data analytics to optimize rail infrastructure and improve overall performance of the rail transport sector is critical [2], [3].

However, due to the complexity of rail networks and the large volume of rolling stock operating on them, train delays are a significant issue that must be addressed. Delays are divided into two types: those caused by the unpredictable time it takes to prepare the train for departure and those caused by variations in the train's performance during its journey [4], [5], [6]. Arrival delay prediction, which involves calculating

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

the difference between the actual arrival time and the scheduled arrival time for a trip between two stations, is critical for rail risk management. When there are disruptions, train dispatchers must assess the impact on the overall schedule and minimize losses by adjusting operations to reduce the chain of delays that could impact overall system operation [7], [8].

Event-based models, which involve procedures with departure, travel, and arrival events, are a common approach for forecasting disruptions and the resulting operational delays in railway operations. Data-driven models, on the other hand, have shown promise in handling and recognizing relationships between nonlinear, multidimensional, and time-based data. These models have been successfully used to uncover interrelationships between various features in rail operations [8], [9], [10], [11], [12], [13], [14], and previous studies have used them to forecast rail operation delays caused by disruptions [15], [16], [17], [18], [19]. These models, however, have failed to predict short-term arrival delay times, as well as the underlying factors that caused the delay and the expected impact on operations. To overcome the limitations of previous studies, the present research has two main objectives:

- Evaluate and compare the effectiveness of various data-driven models in predicting short-term arrival delay times in freight rail operations.
- Determine the significance of features associated with arrival delay time.
- Create a Short-term Decision Support System (STDSS) to evaluate operational interventions aimed at reducing disruptions and their associated delays in real-time freight operations.

The remainder of this article is structured as follows. Section II provides a review of previous studies that have examined methods for modelling delays in rail operations, as well as data-driven models. In Section III, the problem is described in detail. Section IV describes the case study and methodology used to implement data-driven models for predicting short-term arrival delay times in freight rail operations, as well as an examination of the significance of the characteristics associated with arrival delay time. Section V includes the results and discussion of the study. Finally, Section VI summarizes the research's key findings and suggests future research directions.

II. LITERATURE REVIEW

Numerous studies have been conducted to investigate the issue of delay propagation caused by disruptions in rail operations. Barta et al. proposed a Markov chain-based model to investigate the spread of delays among trains that connect intermodal terminals, which can be caused by unforeseen events like traffic congestion or unscheduled maintenance [20]. In another study, a Bayesian networks approach was proposed to address this issue, where evidence of events was used to reduce uncertainty over time for other events [21]. Additionally, researchers assessed the effectiveness of various timetables, including the shuttle timetable, in allowing operations to continue despite disruptions [22], [23].

Wen et al. investigated data-driven methods for train dispatching in passenger and freight rail operations, discovering that the use of ML methods is very promising due to the rich data that can be obtained from train operations. For this reason, numerous studies have used data-driven models such as decision trees, support vector machines, random forests, and artificial neural networks to predict and investigate rail operation delays. These studies' findings have been mixed, with some revealing a strong relationship between train delays and dwell times and others revealing a weaker relationship between running times and departure delays [24], [25]. Peters et al., for example, built a neural network based on rules between dependent trains to forecast delays for real-time delay monitoring [26], whereas Pongnumkul et al., used the moving average of historical travel times and travel times of the k-nearest neighbors (k-NN) to predict passenger train arrival times [27]. In another interesting study in this field, Oneto et al. created a dynamic data-driven train delay prediction system for large-scale railway networks using weather data from national services [28].

Several data-driven approaches have been used to forecast rail operation delays caused by disruptions using regression models. For example, Kecman and Goverde created a decision tree model and a least-trimmed squares robust linear regression model to predict train running and dwell times [15]. Li et al. used linear regression and K-Nearest Neighbor algorithms to predict the duration of station stops in a different study [16]. Barbour et al. used a support vector regression model to forecast estimated arrival times for freight trains based on train, network, and traffic congestion features [29]. Meanwhile, other researchers [17], [18] used data-driven methods to investigate the characteristics of rail service interruptions and the resulting delays in High-Speed Railway Systems. Minbashi et al. also proposed a machine learning-based framework to improve predictability in freight rail operations by focusing on yard arrivals and departures and employing a random forest algorithm [30].

While previous research has studied the problem of predicting train disruptions, our study addresses a significant gap in the literature. Specifically, we focus on the short-term prediction of arrival delay times in freight rail operations and identify the root causes of delay and their expected impact on operations. While previous studies have struggled to predict the arrival delay time in the short-term, particularly after the train departs from the previous control station, our study develops a consistent data-driven model using supervised Machine Learning (ML) that surpasses other models in predicting the arrival delay time within this context. Additionally, we use the Shapley Additive exPlanation method (SHAP) to thoroughly analyze the impact of the features such as departure delay time, trip distance, and train composition on arrival delay time. Our findings allow us to develop a STDSS that can evaluate operational interventions aimed at reducing delays in freight rail operations. In general, our research makes an important addition to the existing literature by addressing a specific gap in short-term prediction and identifying the root causes of delay in freight rail operations, with the aim of creating this useful STDSS for a specific company.

In a previous study [31], some of the authors of this research used a binary classification approach to predict whether a train will be delayed or not in the long-term (days) and to identify the features that cause those delays. However, in this current study, the authors focus on developing a short-term prediction model for real-time decision making using a regression approach.

TABLE 1 summarizes recent and representative studies on railway delay propagation in chronological order. It discusses the type of railway investigated as well as the major contributions.

III. PROBLEM DESCRIPTION

Time and distance charts are a common and standardized method for evaluating train performance and detecting intra-journey schedule deviations. FIGURE 1 shows an example where the cumulative distance traveled on the vertical axis and cumulative time on the horizontal axis can be easily identified, allowing for easy identification of critical points where delays accumulate and evaluating the train's performance compared to the schedule at any point during the journey. This method aids in providing a comprehensive view of the journey and is useful for evaluating train performance [46].

Given that the orange line in FIGURE 1 represents the scheduled time-spatial trajectory and the blue line represents the train's actual ones, the goal of this research is to predict the arrival delay time after the train has departed from the previous control station (so that the departure delay time is known), and we use data-driven models to identify the data behavior responsible for the variety of intra-journey possibilities.

Equations (1)-(5) are defined based on FIGURE 1, where the departure delay time (Dep_{Delay}) is a function of the actual departure time (Act_{Dep}) and the scheduled departure time (Sch_{Dep}). Similarly, the arrival delay time (Arr_{Delay}) is determined by the actual arrival time (Act_{Arr}) as well as the scheduled arrival time (Sch_{Arr}). In contrast, the scheduled travel time (Sch_{Time}) is a function of Sch_{Arr} and Sch_{Dep}.

$$Dep_{Delay} = Act_{Dep} - Sch_{Dep} \tag{1}$$

$$Arr_{Delay} = Act_{Arr} - Sch_{Arr} \tag{2}$$

$$Sch_{Time} = Sch_{Arr} - Sch_{Dep} \tag{3}$$

$$Arr_{Delay} = Act_{Arr} - Sch_{Time} + Act_{Dep} - Dep_{Delay} \tag{4}$$

$$Arr_{Delay} = \sum X_i - Sch_{Time} + Act_{Dep} - Dep_{Delay} \tag{5}$$

Given that this problem is dealing with the short-term prediction of the arrival delay time once the train has departed from the previous station, the arrival delay time is a function of known features, except for the actual arrival time, for which data-driven models are implemented to predict its value.

Data for train journey segments can be sourced from a schedule of specific waypoints defined by the National

TABLE 1. A summary with studies related to delay prediction.

RAILWAY	CONTRIBUTIONS	YEAR
Freight rail network between Luxembourg and nine stations in Belgium, France, Germany, Poland, and Italy	Development of a short-term predictive data-driven model for predicting arrival delay time in freight rail operations. Determination of the impact of the features associated with arrival delay time. Creation of a STDSS to evaluate operational interventions aimed at reducing disruptions and their associated delays in real-time freight operations.	Current study, 2023
Freight rail network in various EU countries	A gradient boosting model was implemented to identify whether a trip was delayed or not [31]	2023
Two main yards of Sweden, Malmö and Hallsberg	A Machine learning-based framework to improve predictability in freight rail operations [30]	2023
A Shunting Yard in Hallsberg in Sweden	Tree-based methods were used to predict the status of departing trains from shunting yards [32]	2021
High-Speed rail system in a national corridor (Wuhan–Guangzhou, China)	Bayesian networks were used to predict the primary delay, the total delay time and the number of affected trains during train operations[33]	2020
High-speed railway lines: Wuhan-Guangzhou, and Xiamen-Shenzhen (China)	A deep learning approach was used to predict train delays [34]	2020
High-speed rail system in four national corridors in China	Analyzed operational data to determine attributes related to delays such as causes, frequencies, space-time distributions, and number of affected trains [35]	2019
High-Speed rail system in a national corridor (Wuhan–Guangzhou, China)	Analyzed train disturbances, the total delayed time, and the number of affected trains [17]	2019
High-Speed rail system in a national corridor (Wuhan–Guangzhou, China)	Three Bayesian network-based models were trained to predict train delays, capturing superposition and interaction effects of those delays [36]	2019
Passenger railways in a national corridor (Eindhoven-Amsterdam, the Netherlands)	A random forest model was trained for predicting passenger train delays [37]	2019
National passenger railway network (Germany)	An ensemble prediction model was used for predicting train delays [38]	2019

TABLE 1. ***(Continued.)*** **A summary with studies related to delay prediction.**

High-Speed rail system in a national corridor (Wuhan–Guangzhou, China)	Built an analytical model using train operation records to evaluate the quality of service at individual track sections [39]	2018
Freight rail network in a regional corridor (Tennessee, USA)	Compared various regression models to predict arrival times [40]	2018
National mixed-traffic railway line (Stockholm-Norrköping, Sweden)	Predicted train delay propagation using Bayesian Networks [22]	2018
Section of freight rail network in Tennessee, US	Estimated times of arrival were predicted, using data from trains, the network, and conflicting traffic in the network [29]	2018
National railway network (Italy)	A deep learning model was developed to predict train delays in large-scale railway networks [41]	2018
Specific zone of a passenger railway line in a city (Copenhagen, Denmark)	A clustering model was implemented to identify various recurrent delay patterns [42]	2018
High-Speed rail system in a national corridor (Wuhan–Guangzhou, China)	Conducted a statistical analysis on delay causes, frequencies, temporal/spatial occurrences, and recovery patterns [18]	2017
National railway line (Istanbul-Ankara, Turkey)	Predicted train departure and arrival delays using Markov-chain based models [43]	2017
National railway network (Italy)	Deep learning machines were utilized to predict dynamic delays in large-scale railway networks [28]	2017
Passenger railways in a national corridor (Utrecht–Eindhoven, The Netherlands)	Developed a model to predict dwell times at intermediary stations without using passenger demand data [16]	2016
Passenger railways in a national corridor (Copenhagen - Roskilde, Denmark)	Used statistical analysis to analyze train delays, considering minimum and scheduling running times [44]	2016
National railway network (The Netherlands)	A disruption length prediction model was constructed, aiming at reducing the uncertainty in the lengths of railway disruptions [45]	2016

Railway Company of Luxembourg (*Société Nationale des chemins de fer Luxembourgeois* or CFL).

To better control the delays accumulated along the train's route, a short-term forecast must be performed once the train passes through an intermediate station as a "checkpoint" to create a STDSS that makes accurate real-time predictions that allow for the implementation of strategies such as train rescheduling, reordering, rerouting, and other strategies to optimize freight rail operations.

TABLE 2. **Characteristics contained in the data.**

TYPE	ATTRIBUTES
Train-related characteristics	Train ID, incoterm, train length and weight, wagon count, maximum TEU (twenty-foot equivalent unit), TEU count.
Wagon-related characteristics	Information about each wagon included in the trains, including the wagon ID, model, order, maximum speed, tare weight of the model, tare weight, and type.
Station-related characteristics	Information about the stations that the trains pass through, including the order, name, city, country, scheduled and actual departure and arrival times, as well as the locations of any intermediate stations along the route.
Operational characteristics	Scheduled and actual departure times, scheduled and actual arrival times, month of operation, whether the train arrived at night or during rush hour, weight per length and per wagon, overall trip mileage, elapsed time between control stations, weekday of departure and arrival, average speed over the course of each trip, and route.

IV. METHODS AND PROCEDURES

In this section, we describe the process of creating a short-term predictive data-driven model to predict the arrival delay time of a train that has already departed from the previous control station. FIGURE 2 depicts the steps in the methodology used in this study, from data collection to the development of the predictive model for further analysis. All the steps depicted are discussed in depth below.

A. DATA COLLECTION

The study used data from the National Rail Company of Luxembourg - CFL Multimodal, which was collected over a 17-month period, from November 2019 to April 2021. The datasets contain information on their freight rail operations conducted between this period of time between Bettembourg (Luxembourg) and other nine stations within the EU (Boulou, Champigneulles and Lyon in France; Zeebrugge and Antwerp in Belgium; Kiel and Rostock in Germany; Poznan in Poland; and Trieste in Italy). This data was provided by CFL Multimodal, which contained a wide variety of attributes related to trains, wagons, stations, and operations, as shown in TABLE 2. The datasets were meticulously analyzed and combined to ensure that all freight rail operations along the various routes depicted in FIGURE 3 were considered.

To ensure high quality, the dataset used in this study underwent various data pre-processing procedures such as feature

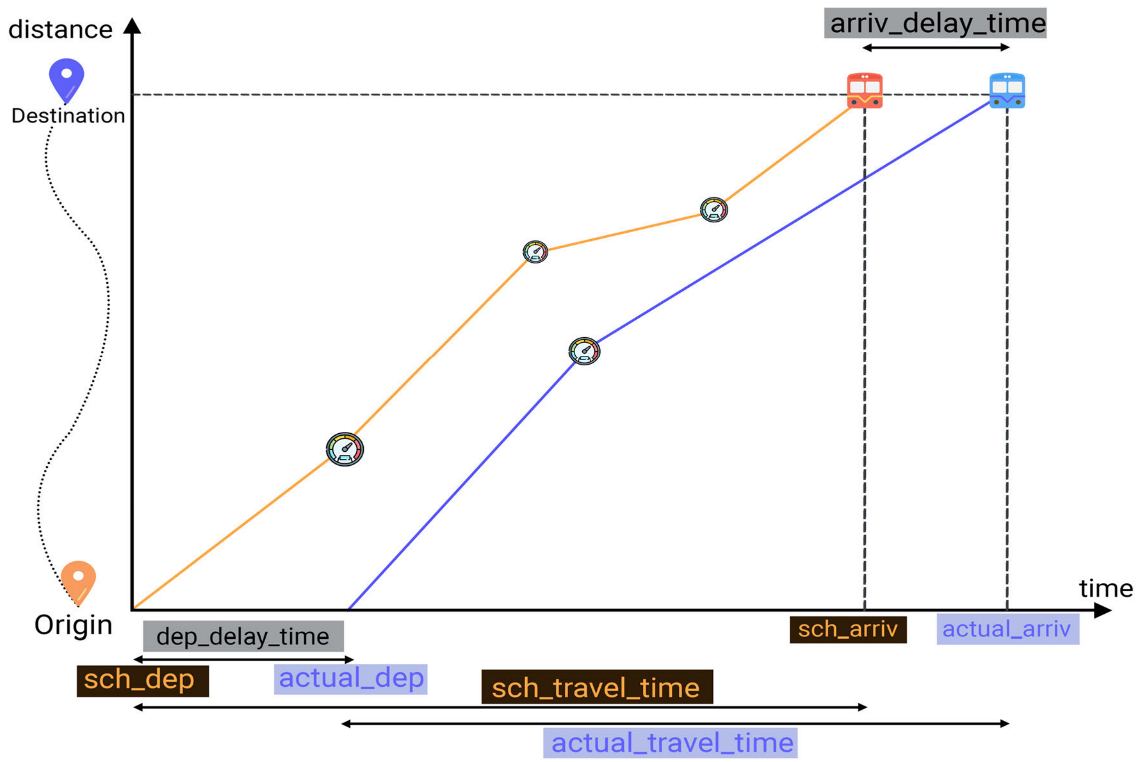


FIGURE 1. Time and Distance chart – Intra-journey characteristics.

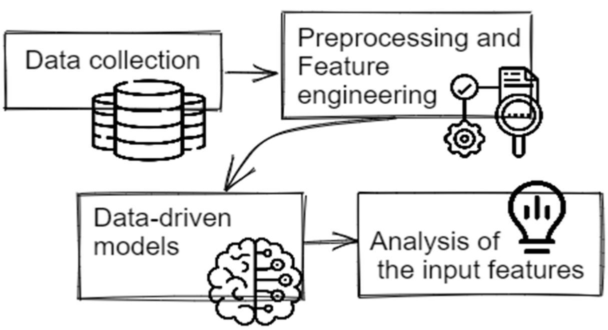


FIGURE 2. Flow diagram of the methodology used in this study.

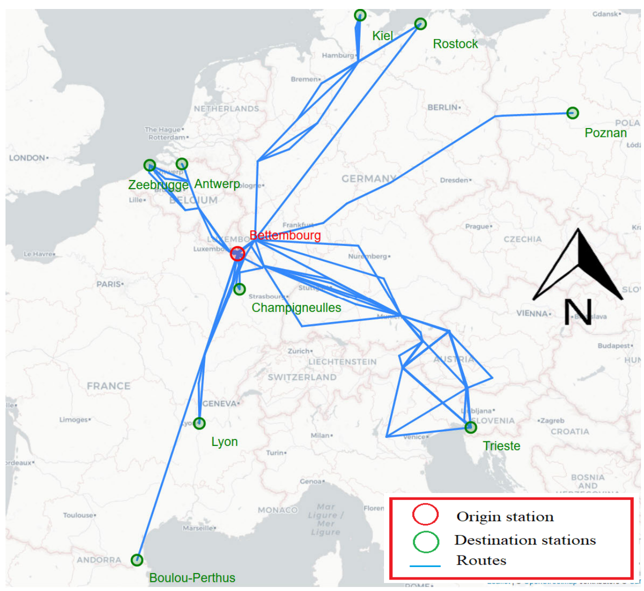


FIGURE 3. Freight rail routes included in the study, including all control stations between origins and destinations.

engineering, data cleaning, and data transformation. Section IV-B describes the resulting dataset in detail, including its size, descriptive statistics, and the data pre-processing procedures used. The data-driven models were then developed and trained using this refined dataset, as described in Section IV-C, with the goal of predicting freight rail arrival delay times. The goal of these models is to predict arrival delay times in freight rail operations in order to provide valuable insights that can be used to improve the reliability of freight rail transport.

B. DATA PROCESSING

Following the organization and combination of the datasets listed in TABLE 2, single dataset was processed to ensure that each row represents the trips between a pair of control stations, which are the stations (junctions) that the train passes through between the starting and destination stations, as shown in FIGURE 3.

Data imputation techniques were used to fill any missing values in the merged dataset. The median value was used to fill in numerical features, and the most common class was used to fill in categorical features. Furthermore, by utilizing the dataset's available features such as train weight, train length, and train wagon count, we used feature engineering to create two new features to improve the predictive capability of our models: train weight per length and train weight per wagon. Additionally, we used one-hot encoding to convert categorical features into dummy variables and the z-score standardization method to rescale numerical features to ensure that the data is on the same scale [47], [48]. The interquartile range method was used to remove outliers from the numerical features as well.

As part of our correlation analysis, we used a 0.7 Pearson correlation coefficient threshold to eliminate any features that were highly correlated with one another. This threshold was chosen in accordance with standard data analysis practice, which states that a correlation coefficient greater than 0.7 indicates a strong linear relationship between two variables [49]. We were able to reduce redundancy in our dataset and improve model performance by removing these highly correlated features.

The goal of this study was to predict the arrival delay time, which is the numerical difference between the actual arrival time and the scheduled arrival time for trips between two control stations (as explained in Section III, equations (1)-(5)). Regression approach was chosen as the best data-driven approach because the target feature is a numeric value. Following extensive data pre-processing and feature engineering, a total of 10,265 trips between control stations were identified for analysis.

C. DATA-DRIVEN MODELS

Predicting arrival delay times in rail operations is a difficult task due to the numerous factors that can affect train schedules. Machine Learning models are increasingly being used for this purpose because of their ability to effectively analyze large amounts of data and learn from it in order to make accurate predictions. Several studies have demonstrated the effectiveness of various machine learning models in predicting arrival delay times in rail systems, including linear regression, logistic regressions, k-nearest neighbors, random forests, gradient boosting machines, and artificial neural networks [39], [50], [51], [52], [53]. These models can consider a variety of factors, such as weather, passenger volume, and train speed, to provide more accurate predictions of arrival delays.

To effectively train and evaluate machine learning (ML) models for predicting arrival delay time, the original dataset was randomly divided into two sets, namely a training set and a testing set, with a 70% to 30% ratio [13]. It is worth noting that both the training and testing data are part of the data used in this study, and as such, they were subject to the same preprocessing and cleaning steps to ensure their consistency and quality. The proportions of independent input features and the target feature, which in this case is the arrival delay time, were the same in both subsets. In order to avoid bias in the results, it is also critical to ensure that the distribution of values for all independent features is similar for both groups.

The arrival delay time was then predicted using a set of machine learning models that had previously been widely and efficiently applied to a variety of regression problems. These models are as follows:

- Linear regression is a machine learning algorithm that forecasts numerically continuous output with a constant slope. This model is typically used to predict values within a continuous range rather than categorizing them into different classes [54].
- The K-nearest neighbors regressor is a non-parametric ML algorithm that approximates the relationship between independent features and continuous outcomes by averaging observations in the same neighborhood [55].
- Random forest regressor, which is a tree-based ensemble ML model that generates many regressors in parallel and aggregates their results by combining a sampling method and an ensemble approach to improve model building [54].
- Light gradient boosting machine, an open-source framework developed by Microsoft for training gradient boosting models [56]. This is another tree-based ensemble ML model that works in a sequential order, with each subsequent model attempting to improve on the errors of the previous model. As a result, each model improves ensemble performance [57].

The performance of the data-driven models was assessed using several metrics, including the Root Mean Squared Error (RMSE), the Coefficient of Determination (R2), the Mean Absolute Percentage Error (MAPE), and the Mean Absolute Error (MAE), which are commonly used metrics to assess the performance of regression machine learning models [52], [58], [59], [60], [61].

- R2 measures the proportion of variance in the dependent variable that is explained by the independent variable(s). It is a measure of how well the model fits the data, with higher values indicating a better fit.
- RMSE measures the difference in absolute terms between predicted and actual values, giving higher weights to larger errors. It provides a measure of how far off the model's predictions are from the actual values, with lower values indicating a better fit. A lower RMSE suggests that the predicted model is closer to the underlying distributions of the actual data, or in other words, a more accurate model.
- MAE is another measure of the difference between predicted and actual values. Unlike RMSE, MAE gives equal weights to all errors and provides a measure of the average magnitude of the errors. A lower MAE indicates a more accurate model.

- MAPE measures the percentage difference between predicted and actual values. It is often used in forecasting and provides a measure of the average magnitude of the errors as a percentage of the actual values. A lower MAPE indicates a more accurate model

Overall, these metrics are useful for evaluating the performance of regression machine learning models because they offer different perspectives on the model's prediction accuracy. Using multiple metrics can help ensure that the model performs well in various areas.

The equations for calculating these metrics are shown in (6)-(9), where: y_i is the actual value of the observation i (target); $\hat{y}_i$ is the predicted value of the observation i (model's output); $\bar{y}_i$ is the average value of all observations i, and n is the number of observations. A good model will typically have a high R2 value as well as low RMSE, MAPE and MAE values.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{6}$$

$$\mathrm{R2} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{7}$$

$$\mathrm{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \bar{y}_i}{y_i} \right| \tag{8}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - \hat{y}_i|}{n} \tag{9}$$

The tuning of hyperparameters is an important step in optimizing the performance of data-driven models. To that end, the random search method was used to find the models' optimal hyperparameters [62]. Furthermore, the models were evaluated using the k-fold cross-validation method to ensure that their performance is robust and generalizable. To do this, the training set was divided into K subsets, with the classes in each subset represented in the same proportions as the entire dataset, and the learning model was then applied to the remaining subsets [63]. This method is commonly used to mitigate any bias introduced by the holdout method, which uses a fixed amount of data for training and the remainder for testing.

Following the selection of the best ML model for predicting arrival delay time, it is critical to assess the model's learning curves to ensure that they are accurate. The learning curves depict the trend of the model's training and cross-validation scores as a function of training sample count. This allows us to detect possible problems of overfitting or underfitting as well as determine whether adding more observations to the training set improves model performance [64].

The models were trained and validated in Python 3.8.5 on an Intel Core i9-10885H CPU @ 2.40 GHz with 32 GB DDR4 memory ram, a Hard Disk SSD 1TB NVMe class 40, and a GPU NVIDIA Quadro P620 DDR5. This hardware configuration enables quick and efficient model training and validation, reducing the time and resources required for the analysis.

D. ANALYSIS OF THE INPUT FEATURES

Following the identification of the best data-driven model, the impact of the features associated with arrival delay time is calculated using the model's coefficients for each input feature. Following the conditional dependence theory [65], the model's coefficients represent the relationship between the given input feature x_i and the target y (i.e., arrival delay time), with the assumption that all other features x_j remain constant. These coefficients represent the impact of each input feature on the model's output, allowing us to evaluate the effect of each individual feature on the arrival delay time.

After that, the Shapley Additive exPlanation method (SHAP) is used to generate feature dependence plots. This method ensures that the results are better interpreted because it reveals the direct impact of each feature on the model [61], [66], allowing for the discovery of correlations between two variables and their impact on freight rail arrival delay times. SHAP feature dependence plots depict the interaction effect of two combined features from the same observation, as well as their impact on the model-predicted feature: the arrival delay time.

V. RESULTS AND DISCUSSION

This section presents the results and analysis of data-driven models for short-term prediction of arrival delay times in freight rail operations, which is divided into two parts: (a) an examination of the performance of the trained data-driven models, with the goal of identifying the model that performed the best based on evaluation metrics such as RMSE, R2, MAPE and MAE, and (b) an analysis of the features that have the most significant impact on delays in freight rail operations, which are then used to gain insights into how the features interact in the output of the best data-driven model.

A. DATA-DRIVEN MODELS

Initially, several feature combinations were tested to determine the most effective set of attributes for data-driven models. Using the Pearson method for correlation analysis, less relevant attributes were removed, resulting in a refined set of features that did not compromise the models' performance. The final dataset's composition is shown in TABLE 3, and a validation process was carried out to ensure that the distribution of values for all features was consistent between the training and test groups by carrying out a consistency test of the data, as shown in FIGURE 4. This approach was used to ensure that the models were trained on a representative sample of the data while avoiding overfitting and underfitting risks.

As described in in Section IV-C, five data-driven models were initially analyzed and evaluated based on the proposed evaluation metrics and the best performing model was then selected for further analysis, where we can see that the lightGBM model performed better than the other models.

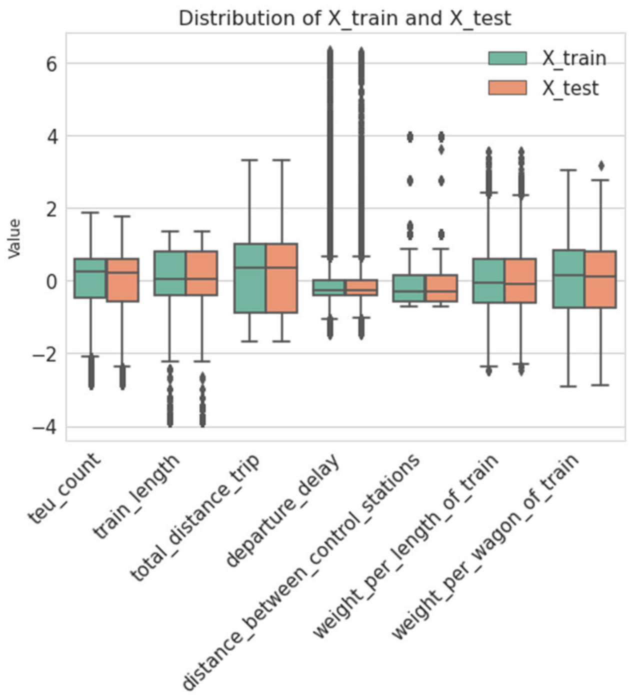


FIGURE 4. Similarity in the distribution of the values of independent features for training and testing data.

TABLE 3. Composition of the final dataset.

FEATURE	DESCRIPTION	FEATURE DISTRIBUTION AND STATISTICS
Target: y	Arrival delay time [min]	mean: 74.35, median: 12.0, std: 241.98, min: -281.0, max: 1570.0
x1	Number of TEU (Twenty-foot Equivalent Unit)	mean: 60.75, median: 64.0, std: 18.0, min: 0.0, max: 98.36
x2	Train length [m]	mean: 537.61, median: 544.0, std: 108.64, min: 34.0, max: 720.0
x3	Distance of the TOTAL trip [km]	mean: 486.69, median: 322.2, std: 280.08, min: 84.51, max: 1454.14
x4	Departure delay time [min]	mean: 79.54, median: 17.0, std: 241.7, min: -260.0, max: 1562.0
x5	Distance between control stations [km]	mean: 53.9, median: 50.18, std: 37.56, min: 1.46, max: 131.06
x6	Train weight over train length [t/m]	mean: 2.24, median: 2.2, std: 0.53, min: 0.91, max: 4.08
x7	Train weight over quantity of wagons [t/wagon]	mean: 66.49, median: 68.97, std: 16.62, min: 20.93, max: 100.69

To improve its performance even further, the random search method was used in conjunction with popular ML Python libraries such as Pycaret, Scikit-learn, and lightGBM [56], [67], [68]. The evaluation metrics of the lightGBM model after tuning their parameters are also shown in TABLE 4.

TABLE 4. Results.

MODEL	R2	MAPE	MAE	RMSE
INITIAL RESULTS				
Light Gradient Boosting Machine (lightGBM)	0.937	1.240	26.230	59.313
Linear Regression (LR)	0.932	1.390	27.925	61.876
Random Forest Regressor (RF)	0.925	1.367	27.908	65.349
K Neighbors Regressor (KNN)	0.915	1.700	37.620	69.548
RESULTS AFTER TUNING THE PARAMETERS OF THE BEST MODEL				
Tuned lightGBM	0.938	1.273	26.378	59.021

Considering the results in TABLE 4, even though the results for some models are quite similar, the tuned lightGBM model slightly outperforms for predicting the arrival delay time (even if any of those are valid options). As a result, this model is chosen to assess the impact of the input features on the model output as well as to investigate the relationship between disruptions and their subsequent delays. For both training and test data, the errors between the best model's prediction of the arrival delay time and the actual arrival delay time of the operations performed by CFL Multimodal were estimated (see FIGURE 5). The scatter plots and the corresponding equations provided for the training and test data show that the model performs well in predicting the arrival delay time of operations made by CFL Multimodal. The R2 scores of 0.96 for the training data and 0.89 for the test data indicate a strong correlation between the predicted and actual values of arrival delay time. The equation (10) for the training data and (11) for the test data reveal that the model's predictions are consistent with the actual data, with only slight deviations from the ideal line (y=x), demonstrating an overall performance of the model being satisfactory.

$$y_{train} = 0.96x + 3.22 \tag{10}$$
$$y_{test} = 0.94x + 6.40 \tag{11}$$

LightGBM is an open-source gradient boosting framework that improves prediction accuracy in regression and classification problems by utilizing decision tree algorithms. It is based on the gradient boosting framework, which combines multiple weak learners to create a strong learner capable of making more accurate predictions. The framework constructs trees in depth and computes gradient and hessian values using a histogram-based approach, which speeds up training and reduces memory usage [56], [57]. LightGBM also has data parallelism, which enables faster training on large datasets, and regularized parameter learning, which reduces overfitting.

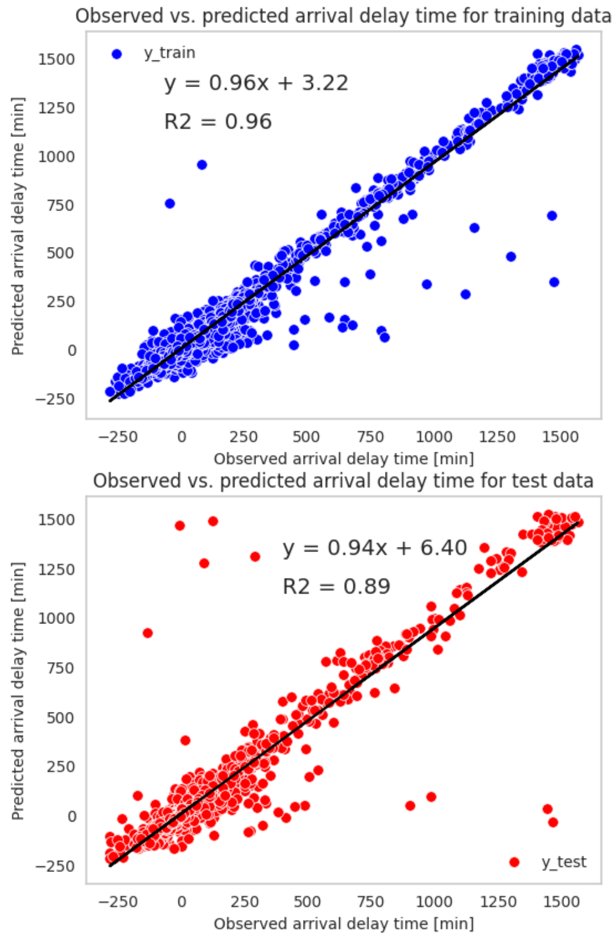


FIGURE 5. Scatter plots of errors for training and test data between the arrival delay time predicted by the model and the arrival delay time of the operations made by CFL Multimodal.

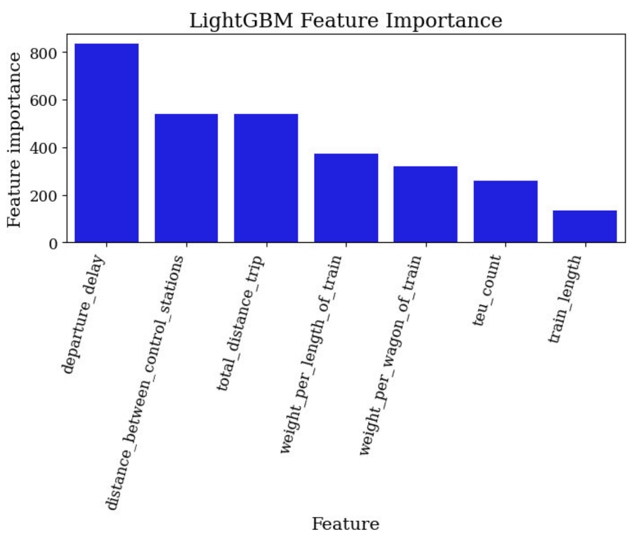


FIGURE 6. Plot showing the input feature importance on arrival delay time resulting from the tuned lightGBM model.

B. ANALYSIS OF THE INPUT FEATURES

FIGURE 6 depicts the effect of each input feature on the magnitude of the output from the tuned lightGBM model.

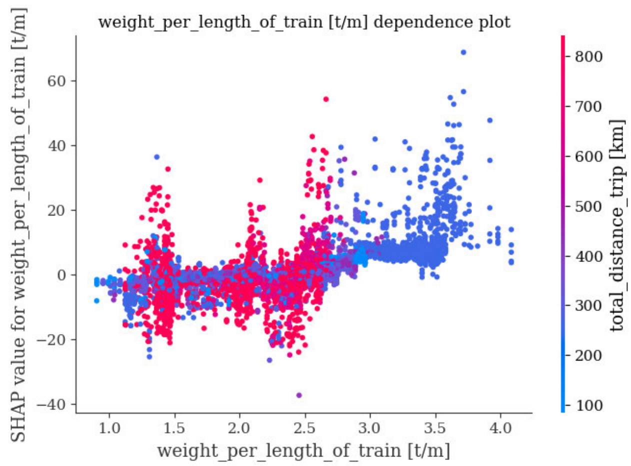


FIGURE 7. Feature dependence plot of weight per length of the train vs total distance trip.

The input features are arranged in descending order by the magnitude of their impact. The greater the value of the feature, the more important it is in predicting the arrival delay time. The departure delay time, as shown in FIGURE 6, is the most important factor in predicting the arrival delay time, followed by the distance traveled (between the previous and destination control stations) and the train composition (in terms of weight, length, and number of wagons).

The SHAP method was used to construct the feature dependency plot between each pair of the seven available features, allowing the discovery of greater interaction effects between each pair of features with a higher SHAP value, and thus a higher incidence in the predicted feature. FIGURE 7 and FIGURE 8 depict the strongest interactions discovered in the feature dependence scatter plots, which show the effect of a single feature on the predictions of the lightGBM model. The following considerations must be made:

- Each point represents a single prediction (observation) from the dataset.
- The x-axis represents the value of the specified feature.
- The y-axis displays the SHAP value for that feature, which indicates how much the model's prediction of the arrival delay time for that sample is influenced by knowing the feature's value.
- The color corresponds to the second feature, which interacts significantly with the feature on the x-axis.

FIGURE 7 depicts the variability of the train's weight per length in predicting the arrival delay time, with a growing trend in the impact of this variable on predicting the arrival delay time, and it is also observed that trains with a higher weight per length of the train have a lower total distance of the trip in general. FIGURE 8, on the other hand, depicts the roughly linear and positive trend between the departure delay time and its SHAP value, or the direct correlation between the departure delay time and the arrival delay time.

This emphasizes the significance of departure delay time in predicting arrival delay time in freight rail operations. These

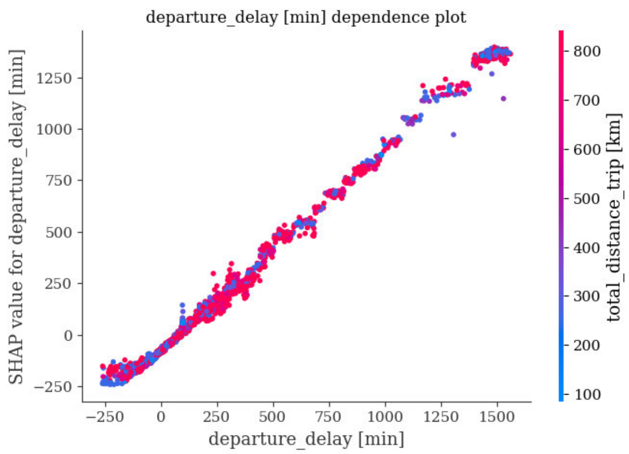


FIGURE 8. Feature dependence plot of departure delay time vs total distance trip.

findings are consistent with previous research on passenger trains, which found that departure delay time is a significant predictor of arrival delay time [7], [8]. This study, however, is the first to show the same correlation in freight rail operations. FIGURE 7 and FIGURE 8 show how important it is to consider the weight per length of the train and the total distance of the trip as variables in predicting arrival delay time. These findings can be used to inform future operational interventions, such as optimizing routes to reduce distance and weight per length of the train, to improve overall freight railway reliability.

It is also worth noting that this study is based on data from a single freight rail company; thus, it would be advantageous to expand this research by including data from other freight rail companies to develop a more comprehensive study. This would allow for more reliable conclusions and a more comprehensive understanding of freight rail operations' behavior.

C. DISCUSSION

This study is the first of its kind to use gradient boosting models to predict arrival delay times in freight rail operations in the short-term. The resulting model is highly efficient and can handle large-scale datasets with high-dimensional features. LightGBM has been shown to outperform other popular machine learning algorithms such as random forest and XGBoost in various benchmarks and real-world applications, making it a popular choice for predictive modeling tasks [56], [57]. Other studies in the field have used different ML models such as neural networks to address other problems in freight rail operations [41], [69].

Previous research has found that train length has an impact on both passenger and freight train punctuality [70], [71]. However, Van Der Kooij et al., discovered that enforcing temporary speed restrictions on longer and heavier passenger trains to safeguard the use of infrastructure could produce significant network delays [72].

This study created a short-term predictive data-driven model to predict the arrival delay time of a train that has already departed from the previous control station and examined the features associated with arrival delay time. This study makes the following significant contributions:

- The development of a consistent short-term predictive data-driven model, which discovered that the lightGBM model surpasses other data-driven models in predicting arrival delay time in freight rail operations.
- The impact of the features associated with arrival delay time was examined, and it was discovered that the departure delay time, the distance of the trip, and the train composition are critical in predicting the arrival delay time in freight rail operations.
- The possibility of CFL implementing the short-term prediction model developed in this study as a STDSS that can be accessed through a simple web service to predict arrival delay times and assess future operational interventions to reduce disruptions and the resulting delays in freight operations.

The findings of this study are useful for the National Railway Company of Luxembourg and other freight rail operations because they can use the predictive model to anticipate delays in the short-term and take proactive measures to reduce disruptions and their consequences. In addition, the analysis of the characteristics associated with arrival delay time provides insights for future research and optimization of freight rail operations.

Some of the authors of this paper previously published a study [31] in which they used the same dataset to build a long-term prediction model using a binary classification approach to identify the rail operating features associated with intermodal freight rail operation delays, allowing them to predict whether a train will be delayed or not in the long run based on its composition. Although both the previous and current studies are concerned with developing predictive models for train delay times in intermodal freight rail operations in Luxembourg, there are significant differences between the two. In the previous study, a binary classification approach was developed for long-term predictions, whereas in this new study, a regression approach was developed for short-term predictions, allowing for real-time decision making. Furthermore, the SHAP method is used in this new study to identify the relationships between input features and delay times, allowing for a more thorough analysis of the causes of delay and the expected impact on real-time operations. Furthermore, the Luxembourg National Railway Company can use the short-term prediction model developed in this study as a decision-support system, providing more support to reduce disruptions and subsequent operational delays, for example, using a simple web service.

VI. CONCLUSION

This study presents a comprehensive approach to predicting freight rail arrival delay times, as well as investigating the underlying causes of delays and their expected impact on operations. The goal is to predict operational delays in real time and to create a STDSS that will assist decision-makers in future operational interventions to reduce disruptions and the resulting delays in freight operations. This will improve railway reliability in the freight transport sector in the long run.

Previous studies have developed models that predict the occurrence of disruptions or delay times in railway operations, but most of them have focused on passenger trains. Freight train research has primarily focused on examining the impact of network delays rather than train delays and has been unable to predict short-term delay times once the train has departed from the previous control station.

In this study, we used regression algorithms to train five data-driven models and analyzed predefined evaluation metrics (R2, RMSE, MAPE and MAE). For the examined dataset, which included railway operations carried out between Luxembourg and nine stations in Belgium, France, Germany, Poland, and Italy over a 17-month period, the lightGBM model stood out as the best data-driven model to predict arrival delay times in freight rail operations.

The lightGBM model has demonstrated that departure delay time, trip distance, and train composition are variables with a significant impact on the prediction of arrival delay times in railway operations. Our findings show that longer trains, longer distances, and heavier trains all have a direct relationship with arrival delay times in general. These findings may pave the way for future research into optimizing the routes of these freight trains' operations to reduce distances, resulting in not only shorter operating times, but also shorter arrival delay times.

However, it is important to note that analyzing the behavior of freight rail operations using only data from one company is insufficient when compared to multiple freight rail companies. As a result, future phases of this study could include data from other companies operating in the region to develop a broader study at the continental level, which could include data from other sources, such as historical climatic data in railway operations. Furthermore, future research studies may be geared toward the creation of a workflow capable of automating all the processes required, from data extraction to the construction and implementation of the models developed in this study. These models enable practitioners to predict arrival delay times in freight rail operations in real time, thereby supporting decisions to reduce the impact of these delays on the overall system operation.

REFERENCES

[1] V. Cacchiani, A. Caprara, and P. Toth, "Scheduling extra freight trains on railway networks," *Transp. Res. B, Methodol.*, vol. 44, no. 2, pp. 215–231, Feb. 2010, doi: 10.1016/j.trb.2009.07.007.

[2] P. Mcmahon, T. Zhang, and R. Dwight, "Requirements for big data adoption for railway asset management," *IEEE Access*, vol. 8, pp. 15543–15564, 2020, doi: 10.1109/ACCESS.2020.2967436.

[3] Q. Li, J.-C. Sibel, M. Berbineau, I. Dayoub, F. Gallee, and H. Bonneville, "Physical layer enhancement for next-generation railway communication systems," *IEEE Access*, vol. 10, pp. 83152–83175, 2022, doi: 10.1109/ACCESS.2022.3192971.

[4] N. Zhao, C. Roberts, S. Hillmansen, and G. Nicholson, "A multiple train trajectory optimization to minimize energy consumption and delay," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2363–2372, Oct. 2015, doi: 10.1109/TITS.2014.2388356.

[5] M. S. Artan and I. Sahin, "Exploring patterns of train delay evolution and timetable robustness," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11205–11214, Aug. 2022, doi: 10.1109/TITS.2021.3101530.

[6] F. Bigi, T. Bosi, J. Pineda-Jaramillo, F. Viti, and A. D'Ariano. (2023). *Addressing the Impact of Maintenance in Shunting Operations Through Shunt-In Policies for Freight Trains Operations*. [Online]. Available: http://hdl.handle.net/10993/53485

[7] N. Bešinović, R. M. P. Goverde, E. Quaglietta, and R. Roberti, "An integrated micro–macro approach to robust railway timetabling," *Transp. Res. B, Methodol.*, vol. 87, pp. 14–32, May 2016, doi: 10.1016/j.trb.2016.02.004.

[8] F. Corman and L. Meng, "A review of online dynamic models and algorithms for railway traffic management," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1274–1284, Jun. 2015, doi: 10.1109/TITS.2014.2358392.

[9] P. Bhavsar, I. Safro, N. Bouaynaya, R. Polikar, and D. Dera, "Machine learning in transportation data analytics," in *Data Analytics for Intelligent Transportation Systems*, M. Chowdhury, A. Apon, and K. Dey, Eds. Amsterdam, The Netherlands: Elsevier, 2017, pp. 283–307, doi: 10.1016/B978-0-12-809715-1.00012-2.

[10] J. D. Pineda-Jaramillo, R. Insa, and P. Martínez, "Modeling the energy consumption of trains by applying neural networks," *Proc. Inst. Mech. Engineers, F, J. Rail Rapid Transit*, vol. 232, no. 3, pp. 816–823, Mar. 2018, doi: 10.1177/0954409717694522.

[11] S. Milinković, M. Marković, S. Vesković, M. Ivić, and N. Pavlović, "A fuzzy Petri net model to estimate train delays," *Simul. Model. Pract. Theory*, vol. 33, pp. 144–157, Apr. 2013, doi: 10.1016/j.simpat.2012.12.005.

[12] V. De Martinis and F. Corman, "Data-driven perspectives for energy efficient operations in railway systems: Current practices and future opportunities," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 679–697, Oct. 2018, doi: 10.1016/j.trc.2018.08.008.

[13] J. Pineda-Jaramillo, P. Martínez-Fernández, I. Villalba-Sanchis, P. Salvador-Zuriaga, and R. Insa-Franco, "Predicting the traction power of metropolitan railway lines using different machine learning models," *Int. J. Rail Transp.*, vol. 9, no. 5, pp. 461–478, Sep. 2021, doi: 10.1080/23248378.2020.1829513.

[14] H. Alawad, S. Kaewunruen, and M. An, "A deep learning approach towards railway safety risk assessment," *IEEE Access*, vol. 8, pp. 102811–102832, 2020, doi: 10.1109/ACCESS.2020.2997946.

[15] P. Kecman and R. M. P. Goverde, "Predictive modelling of running and dwell times in railway traffic," *Public Transp.*, vol. 7, no. 3, pp. 295–319, Dec. 2015, doi: 10.1007/s12469-015-0106-7.

[16] D. Li, W. Daamen, and R. M. P. Goverde, "Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station," *J. Adv. Transp.*, vol. 50, no. 5, pp. 877–896, Aug. 2016, doi: 10.1002/atr.1380.

[17] P. Huang, C. Wen, Q. Peng, C. Jiang, Y. Yang, and Z. Fu, "Modeling the influence of disturbances in high-speed railway systems," *J. Adv. Transp.*, vol. 2019, pp. 1–13, Mar. 2019, doi: 10.1155/2019/8639589.

[18] C. Wen, Z. Li, J. Lessan, L. Fu, P. Huang, and C. Jiang, "Statistical investigation on train primary delay based on real records: Evidence from Wuhan–Guangzhou HSR," *Int. J. Rail Transp.*, vol. 5, no. 3, pp. 170–189, Jul. 2017, doi: 10.1080/23248378.2017.1307144.

[19] P. Huang, T. Spanninger, and F. Corman, "Enhancing the understanding of train delays with delay evolution pattern discovery: A clustering and Bayesian network approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 9, pp. 15367–15381, Sep. 2022, doi: 10.1109/TITS.2022.3140386.

[20] J. Barta, A. E. Rizzoli, M. Salani, and L. M. Gambardella, "Statistical modelling of delays in a rail freight transportation network," in *Proc. Title, Winter Simul. Conf. (WSC)*, Dec. 2012, pp. 1–12, doi: 10.1109/WSC.2012.6465188.

[21] F. Corman, A. D'Ariano, and I. A. Hansen, "Evaluating disturbance robustness of railway schedules," *J. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 106–120, Jan. 2014, doi: 10.1080/15472450.2013.801714.

[22] F. Corman and P. Kecman, "Stochastic prediction of train delays in real-time using Bayesian networks," *Transp. Res. C, Emerg. Technol.*, vol. 95, pp. 599–615, Oct. 2018, doi: 10.1016/j.trc.2018.08.003.

[23] A. Balster, O. Hansen, H. Friedrich, and A. Ludwig, "An ETA prediction model for intermodal transport networks based on machine learning," *Bus. Inf. Syst. Eng.*, vol. 62, no. 5, pp. 403–416, Oct. 2020, doi: 10.1007/s12599-020-00653-0.

[24] I. A. Hansen, R. M. P. Goverde, and D. J. van der Meer, "Online train delay recognition and running time prediction," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst.* Lisbon, Portugal: IEEE, Sep. 2010, pp. 1783–1788, doi: 10.1109/ITSC.2010.5625081.

[25] J. Luo, Q. Peng, C. Wen, W. Wen, and P. Huang, "Data-driven decision support for rail traffic control: A predictive approach," *Exp. Syst. Appl.*, vol. 207, Nov. 2022, Art. no. 118050, doi: 10.1016/j.eswa.2022.118050.

[26] J. Peters, B. Emig, M. Jung, and S. Schmidt, "Prediction of delays in public transportation using neural networks," in *Proc. Int. Conf. Comput. Intell. Model., Control Autom. Int. Conf. Intell. Agents, Web Technol. Internet Commerce (CIMCA-IAWTIC)*, 2005, pp. 92–97, doi: 10.1109/CIMCA.2005.1631451.

[27] S. Pongnumkul, T. Pechprasarn, N. Kunaseth, and K. Chaipah, "Improving arrival time prediction of Thailand's passenger trains using historical travel times," in *Proc. 11th Int. Joint Conf. Comput. Sci. Softw. Eng. (JCSSE)*, May 2014, pp. 307–312, doi: 10.1109/JCSSE.2014.6841886.

[28] L. Oneto, E. Fumeo, G. Clerico, R. Canepa, F. Papa, C. Dambra, N. Mazzino, and D. Anguita, "Dynamic delay predictions for large-scale railway networks: Deep and shallow extreme learning machines tuned via thresholdout," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 10, pp. 2754–2767, Oct. 2017, doi: 10.1109/TSMC.2017.2693209.

[29] W. Barbour, J. C. M. Mori, S. Kuppa, and D. B. Work, "Prediction of arrival times of freight traffic on US railroads using support vector regression," *Transp. Res. C, Emerg. Technol.*, vol. 93, pp. 211–227, Aug. 2018, doi: 10.1016/j.trc.2018.05.019.

[30] N. Minbashi, H. Sipilä, C.-W. Palmqvist, M. Bohlin, and B. Kordnejad, "Machine learning-assisted macro simulation for yard arrival prediction," *J. Rail Transp. Planning Manage.*, vol. 25, Mar. 2023, Art. no. 100368, doi: 10.1016/j.jrtpm.2022.100368.

[31] J. Pineda-Jaramillo and F. Viti, "Identifying the rail operating features associated to intermodal freight rail operation delays," *Transp. Res. C, Emerg. Technol.*, vol. 147, Feb. 2023, Art. no. 103993, doi: 10.1016/j.trc.2022.103993.

[32] N. Minbashi, M. Bohlin, C.-W. Palmqvist, and B. Kordnejad, "The application of tree-based algorithms on classifying shunting yard departure status," *J. Adv. Transp.*, vol. 2021, pp. 1–10, Sep. 2021, doi: 10.1155/2021/3538462.

[33] P. Huang, J. Lessan, C. Wen, Q. Peng, L. Fu, L. Li, and X. Xu, "A Bayesian network model to predict the effects of interruptions on train operations," *Transp. Res. C, Emerg. Technol.*, vol. 114, pp. 338–358, May 2020, doi: 10.1016/j.trc.2020.02.021.

[34] P. Huang, C. Wen, L. Fu, Q. Peng, and Y. Tang, "A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems," *Inf. Sci.*, vol. 516, pp. 234–253, Apr. 2020, doi: 10.1016/j.ins.2019.12.053.

[35] Y. Yang, P. Huang, Q. Peng, J. Li, and C. Wen, "Statistical delay distribution analysis on high-speed railway trains," *J. Modern Transp.*, vol. 27, no. 3, pp. 188–197, Sep. 2019, doi: 10.1007/s40534-019-0188-z.

[36] J. Lessan, L. Fu, and C. Wen, "A hybrid Bayesian network model for predicting delays in train operations," *Comput. Ind. Eng.*, vol. 127, pp. 1214–1222, Jan. 2019, doi: 10.1016/j.cie.2018.03.017.

[37] M. A. Nabian, N. Alemazkoor, and H. Meidani, "Predicting near-term train schedule performance and delay using bi-level random forests," *Transp. Res. Record, J. Transp. Res. Board*, vol. 2673, no. 5, pp. 564–573, May 2019, doi: 10.1177/0361198119840339.

[38] R. Nair, T. L. Hoang, M. Laumanns, B. Chen, R. Cogill, J. Szabó, and T. Walter, "An ensemble prediction model for train delays," *Transp. Res. C, Emerg. Technol.*, vol. 104, pp. 196–209, Jul. 2019, doi: 10.1016/j.trc.2019.04.026.

[39] J. Lessan, L. Fu, C. Wen, P. Huang, and C. Jiang, "Stochastic model of train running time and arrival delay: A case study of Wuhan–Guangzhou high-speed rail," *Transp. Res. Record, J. Transp. Res. Board*, vol. 2672, no. 10, pp. 215–223, Dec. 2018, doi: 10.1177/0361198118780830.

[40] W. Barbour, C. Samal, S. Kuppa, A. Dubey, and D. B. Work, "On the data-driven prediction of arrival times for freight trains on U.S. railroads," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2289–2296, doi: 10.1109/ITSC.2018.8569406.

[41] L. Oneto, E. Fumeo, G. Clerico, R. Canepa, F. Papa, C. Dambra, N. Mazzino, and D. Anguita, "Train delay prediction systems: A big data analytics perspective," *Big Data Res.*, vol. 11, pp. 54–64, Mar. 2018, doi: 10.1016/j.bdr.2017.05.002.

[42] F. Cerreto, B. F. Nielsen, O. A. Nielsen, and S. S. Harrod, "Application of data clustering to railway delay pattern recognition," *J. Adv. Transp.*, vol. 2018, pp. 1–18, Apr. 2018, doi: 10.1155/2018/6164534.

[43] İ. Şahin, "Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances," *J. Rail Transp. Planning Manage.*, vol. 7, no. 3, pp. 101–113, Dec. 2017, doi: 10.1016/j.jrtpm.2017.08.006.

[44] F. Cerreto, O. A. Nielsen, S. Harrod, and B. F. Nielsen, "Causal analysis of railway running delays," in *Proc. 11th World Congr. Railway Res. (WCRR)*, Milan, Italy, 2016. [Online]. Available: https://orbit.dtu.dk/en/publications/causal-analysis-of-railway-running-delays

[45] A. A. Zilko, D. Kurowicka, and R. M. P. Goverde, "Modeling railway disruption lengths with copula Bayesian networks," *Transp. Res. C, Emerg. Technol.*, vol. 68, pp. 350–368, Jul. 2016, doi: 10.1016/j.trc.2016.04.018.

[46] J. Pineda-Jaramillo, W. McDonald, W. Zheng, and F. Viti, "Identifying the major causes associated to rail intermodal operation disruptions using causal machine learning," in *Proc. 101st Transp. Res. Board*, Washington, DC, USA, 2022. [Online]. Available: http://hdl.handle.net/10993/53922

[47] D. Bollegala, "Dynamic feature scaling for online learning of binary classifiers," *Knowl.-Based Syst.*, vol. 129, pp. 97–105, Aug. 2017, doi: 10.1016/j.knosys.2017.05.010.

[48] R. Kang, J. Wang, J. Chen, J. Zhou, Y. Pang, and J. Cheng, "Analysis of failure features of high-speed automatic train protection system," *IEEE Access*, vol. 9, pp. 128734–128746, 2021, doi: 10.1109/ACCESS.2021.3113381.

[49] P. Gaur, K. McCreadie, R. B. Pachori, H. Wang, and G. Prasad, "An automatic subject specific channel selection method for enhancing motor imagery classification in EEG-BCI using correlation," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102574, doi: 10.1016/j.bspc.2021.102574.

[50] S. Dündar and İ. Şahin, "Train re-scheduling with genetic algorithms and artificial neural networks for single-track railways," *Transp. Res. C, Emerg. Technol.*, vol. 27, pp. 1–15, Feb. 2013, doi: 10.1016/j.trc.2012.11.001.

[51] C. Wen, W. Mou, P. Huang, and Z. Li, "A predictive model of train delays on a railway line," *J. Forecasting*, vol. 39, no. 3, pp. 470–488, Apr. 2020, doi: 10.1002/for.2639.

[52] P. Huang, C. Wen, L. Fu, J. Lessan, C. Jiang, Q. Peng, and X. Xu, "Modeling train operation as sequences: A study of delay prediction with operation and weather data," *Transp. Res. E, Logistics Transp. Rev.*, vol. 141, Sep. 2020, Art. no. 102022, doi: 10.1016/j.tre.2020.102022.

[53] N. Marković, S. Milinković, K. S. Tikhonov, and P. Schonfeld, "Analyzing passenger train arrival delays with support vector regression," *Transp. Res. C, Emerg. Technol.*, vol. 56, pp. 251–262, Jul. 2015, doi: 10.1016/j.trc.2015.04.004.

[54] S. Suthaharan, *Machine Learning Models and Algorithms for Big Data Classification*, vol. 36. Boston, MA, USA: Springer, 2016, doi: 10.1007/978-1-4899-7641-3.

[55] K. S. Kim, H. H. Choi, C. S. Moon, and C. W. Mun, "Comparison of K-nearest neighbor, quadratic discriminant and linear discriminant analysis in classification of electromyogram signals based on the wrist-motion directions," *Current Appl. Phys.*, vol. 11, no. 3, pp. 740–745, May 2011, doi: 10.1016/j.cap.2010.11.051.

[56] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2017, pp. 3147–3155.

[57] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A comprehensive evaluation of ensemble learning for stock-market prediction," *J. Big Data*, vol. 7, no. 1, p. 20, Dec. 2020, doi: 10.1186/s40537-020-00299-5.

[58] M. Fayaz and D. Kim, "A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings," *Electronics*, vol. 7, no. 10, p. 222, Sep. 2018, doi: 10.3390/electronics7100222.

[59] M. F. Gorman, "Statistical estimation of railroad congestion delay," *Transp. Res. E, Logistics Transp. Rev.*, vol. 45, no. 3, pp. 446–456, May 2009, doi: 10.1016/j.tre.2008.08.004.

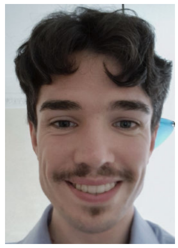
[60] M. S. Devi, R. M. Mathew, and R. Suguna, "Regressor fitting of feature importance for customer segment prediction with ensembling schemes using machine learning," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 952–956, Aug. 2019, doi: 10.35940/ijeat.F8255.088619.

[61] J. Xu, A. Wang, N. Schmidt, M. Adams, and M. Hatzopoulou, "A gradient boost approach for predicting near-road ultrafine particle concentrations using detailed traffic characterization," *Environ. Pollut.*, vol. 265, Oct. 2020, Art. no. 114777, doi: 10.1016/j.envpol.2020.114777.

[62] I. Ivanoska, L. Pastorino, and M. Zanin, "Assessing identifiability in airport delay propagation roles through deep learning classification," *IEEE Access*, vol. 10, pp. 28520–28534, 2022, doi: 10.1109/ACCESS.2022.3158313.
[63] T. Kuflik, E. Minkov, S. Nocera, S. Grant-Müller, A. Gal-Tzur, and I. Shoor, "Automating a framework to extract and analyse transport related social media content: The potential and the challenges," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 275–291, Apr. 2017, doi: 10.1016/j.trc.2017.02.003.
[64] A. N. Richter and T. M. Khoshgoftaar, "Learning curve estimation with large imbalanced datasets," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 763–768, doi: 10.1109/ICMLA.2019.00135.
[65] D. Husmeier, "Introduction to learning Bayesian networks from data," in *Probabilistic Modeling in Bioinformatics and Medical Informatics*, London, U.K.: Springer-Verlag, 2005, pp. 17–57, doi: 10.1007/1-84628-119-9_2.
[66] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Dec. 2017, pp. 4768–4777. [Online]. Available: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions
[67] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2012.
[68] M. Ali. (2020). *PyCaret: An Open Source, Low-Code Machine Learning Library in Python*. [Online]. Available: https://www.pycaret.org
[69] C. Wen, P. Huang, Z. Li, J. Lessan, L. Fu, C. Jiang, and X. Xu, "Train dispatching management with data- driven approaches: A comprehensive review and appraisal," *IEEE Access*, vol. 7, pp. 114547–114571, 2019, doi: 10.1109/ACCESS.2019.2935106.
[70] A. Økland and N. O. E. Olsson, "Punctuality development and delay explanation factors on Norwegian railways in the period 2005–2014," *Public Transp.*, vol. 13, no. 1, pp. 127–161, Mar. 2021, doi: 10.1007/s12469-020-00236-y.
[71] N. O. E. Olsson and H. Haugland, "Influencing factors on train punctuality—Results from some Norwegian studies," *Transp. Policy*, vol. 11, no. 4, pp. 387–397, Oct. 2004, doi: 10.1016/j.tranpol.2004.07.001.
[72] R. B. K. van der Kooij, A. D. Landmark, A. A. Seim, and N. O. E. Olsson, "The effect of temporary speed restrictions, analyzed by using real train traffic data," *Transp. Res. Proc.*, vol. 22, pp. 580–587, 2017, doi: 10.1016/j.trpro.2017.03.047.

JUAN PINEDA-JARAMILLO received the Ph.D. degree in engineering from the Technical University of Valencia, Spain. He is currently with the MobiLab Transport Research Group, University of Luxembourg, on the prediction of disruptions and optimization in rail intermodal operations as part of an FNR-funded project between the University of Luxembourg and CFL Multimodal. He is also an experienced data scientist and a researcher with passion in building and deploying successful algorithms and predictive models in different areas within transport planning, traffic safety, and railway engineering. His research interests include data science and transportation analytics.

FEDERICO BIGI is currently pursuing the Ph.D. degree with the Transportation Department, University of Luxembourg. His research interest includes the optimization of the shunting movement for freight trains management to agent-based modeling and simulation.



TOMMASO BOSI is currently pursuing the Ph.D. degree in computer science and automation with Roma Tre University, Italy. He is one of the Ph.D. Student Representative with the Department of Civil, Computer Science and Aeronautical Technologies Engineering. He is also the Co-Founder of Aless Don Milani, a foundation that develops national and European projects in the field of sustainability innovations. His research interests include the application of operations research and big data in transportation systems, and aiming to spread the ideals of sustainable mobility.

FRANCESCO VITI received the M.Sc. degree in civil engineering from the University of Naples "Federico II" and the Ph.D. degree in transport planning from the Delft University of Technology. He is currently the Head of the MobiLab Transport Research Group, Department of Engineering, and an Associate Member of the MIT SCALE Center for Logistics and Supply Chain Management (LCL) and the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg. He is the author of more than 300 journals and conference papers. He is an Associate Editor of the *Journal of Intelligent Transportation Systems*, *Transportation Research—C: Emerging Technologies*, /Journal of Advanced Transportation, and the *Journal of Big Data Analytics in Transportation*.

ANDREA D'ARIANO received the B.S. and M.S. degrees in computer science, automation and management engineering from Roma Tre University and the Ph.D. degree from the Department of Transport and Planning, Delft University of Technology, in April 2008, under the supervision of Prof. I. A. Hansen. Currently, he is an Associate Professor with the Department of Civil, Computer Science and Aeronautical Technologies Engineering, Roma Tre University. His research interest includes the study of scheduling problems with application to public transportation and logistics. He is an Associate Editor of well-known international journals, such as *Transportation Research—B: Methodological*, *Transportation Research—C: Emerging Technologies*, and *Transportation Research—E: Logistics and Transportation Review*, and conferences, such as IEEE ITSC.