Received: 1 November 2020 | Revised: 29 May 2021 | Accepted: 2 June 2021

IET Intelligent Transport Systems

DOI: 10.1049/itr2.12095

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

An estimation framework to quantify railway disruption parameters

Bhagya Shrithi Grandhi[1] | Emmanouil Chaniotakis[2] | Stephan Thomann[3] | Felix Laube[3] | Constantinos Antoniou[4]

[1] Institute of Transportation and Urban Engineering, Technical University of Braunschweig, Braunschweig, Germany

[2] MaaSLab, UCL Energy Institute, University College London (UCL), London, England

[3] Traffic Management Systems, Emch and Berger Group, Bern, Switzerland

[4] Chair of Transportation Systems Engineering, Technical University of Munich (TUM), Munich, Germany

Correspondence
Bhagya Shrithi Grandhi, Institute of Transportation and Urban Engineering, Technical University of Braunschweig , Hermann-Blenk-strasse 42, 38108 Braunschweig, Germany.
Email: s.grandhi@tu-braunschweig.de

Abstract

Railway network operations form complex systems. Any disruption adversely impacts the operations, causing long delays. Many studies investigate the effect of a railway incident; however, a holistic quantification is lacking. This study aims to present an estimation framework for flexible traffic management systems, which can help reduce network delays and enable dispatchers to make better-informed decisions. An incident's impact on the network is estimated by creating a sequence of models, which predict two key variables. Firstly, the incident duration is predicted, which is used to predict the second variable: total delay caused by the incident. Various influencing attributes are examined, such as weather, network and railway-related attributes. Their relationship with the response variables is studied in order to understand the incident's impact. Using incident data from the Danish Railways, machine learning models are estimated. The results show that neural networks outperform other competing models for total delay modelling, resulting in improved prediction by the estimation framework, thus giving higher accuracy than the stand-alone models in the study.

1 | INTRODUCTION

Real-time operations of a railway system are unavoidably subject to incidents and disruptions, resulting in delays and influences the timetables. As the duration of an incident increases, the impact on the network varies, leading to potential cumulative delays. Recovering from a disrupted situation requires timetable changes, and in some cases, changes in the rolling stock and crew rosters as well [1].

Currently, there is a rising need to quantify the impacts of disruptions [2]. Incidents are critical and need to trigger an immediate response, as a small delay could quickly propagate into the network, disrupting many connections and causing more delays, compromising customer service [3]. In railway traffic management systems (RTMS), any response to the occurrence of an incident aims at the fastest possible recovery of the system in failure [4]. Most of the literature is based on traditional railway systems that are usually microscopic. Microscopic approaches discuss aspects of delay mitigation and prediction in detail for a specific element such as a train, a line, or a type of incident [1].

Nevertheless, macroscopic approaches are lately receiving a growing interest. Macroscopic approaches consider the railway network on a higher level in the form of nodes and edges. Having a more holistic view on a network level in terms of cumulative delays caused by an incident can help railway dispatchers to come up with more informed responses to an incident, such as a quick rescheduling or a better allocation of resources [5].

However, to the best of the author's knowledge, little has been done on investigating network-wide aspects of disruptions in terms of examining the relationship between duration and total delay caused by an incident [6] and the prediction of total delay caused by an incident. This study plans to address this gap by investigating specifically to railway disruption prediction. It contributes to predicting macroscopic parameters that capture and quantify incidents impact in terms of two response variables: the total delay caused by an incident and incident duration and their correlation. For clarity and consistency, the term “total delay” in this paper refers to the total delay caused over the network by one incident, and “duration” is how long the same incident lasts.

Studying such macroscopic variables could act as a key part of the scheduler system in the railway dispatching environment such as railway dynamic traffic management system (RDTMS) [7] or dynamic timetables based on service intentions [8]. Rescheduling requires rapid decision making in response to an incident on the network. Conventional traffic management systems are partially manual and partially automated. Many organisations are stepping into the direction of automated dynamic rescheduling. One such attempt is being done in Denmark railways. In such dynamic systems, an automated scheduler is an essential component which recursively checks all the information related to the incident and accordingly reschedules.

Frameworks like the one presented in this paper can provide helpful information to be used as the basis for rescheduling using the empirical data. Knowing the predicted duration of an incident and the total delay caused by it, operators can better shift and rearrange operations around the affected part of the network. Models discussed in this research could help in training the scheduler upon detection of an incident, by providing a forecasted duration of the incident and the total delay caused by it, which could help create more informed production plans; something currently taking place mostly based on personal experience and knowledge of the traffic dispatcher’s (TD) [1].

A broad spectrum of independent variables are investigated, such as railway, weather and network-related characteristics as inputs. Besides prediction, these models also help in understanding the intricate inter-dependencies between the incident impact and the various other attributes. This helps in preparing for response action and reducing the propagation of delays into the network [9]. The main idea for the estimating framework is that, when an incident occurs, initially, the possible duration of the incident is predicted, based on data. Then the predicted duration is used for the estimation of the total delay caused by the incident.

Analysis is performed using incident data from the Danish Railway, over one and a half years. The raw data included various data fields such as date of occurrence, incident duration, number of trains impacted and the delays caused to each of them, stations in the incident area and many more. Such an extensive data-set on delays and incidents allows the performance of useful analysis and provides information about data related aspects, using supervised machine learning algorithms for modelling.

This paper is structured as follows. Section 2 discusses an overview of models for the prediction of duration, delays in railway systems and interdependencies in delays. Also, a summary of researches related to influencing factors like weather and centrality measures is presented. Section 3 presents all the variables and a descriptive analysis of the data, discussing various trends and relationships observed in the variables. Additionally, it describes the data preparation steps for modelling. Section 4, presents the types and details of the implemented models. Also, the corresponding results are presented and evaluated. Additionally, the estimation framework is also examined in this section. The last section 5 presents all the conclusions deduced from the study, further discussing the limitations and future work.

2 | LITERATURE REVIEW

Many studies have been conducted on railway disruption management, in terms of prediction of delay parameters and impact assessment. The reader is referred to [1], which presents a comprehensive review in railway disruption management. With regards to duration prediction in railway, the pertinent literature is found to be limited. One practical application of the duration prediction has been a part of the French railways (SNCF - Société Nationale des Chemins de Fer francais) traffic management system EXCALIBUR, and it is being used in real-time operations [10]. EXCALIBUR system is built on statistical models based on the historical incident database. The variables related to the incidents are categorised into static and dynamic based on their function. Median model, generalised linear models(GLM) and regression trees are implemented. It is found that GLM and regression trees performed well among the three implemented models [10].

Zilko.et.al [11] uses a Copula Bayesian model to handle the uncertainties where the copula represents the conditional probabilities between variables. This research uses different influencing factors like time, location, weather, contract type and cause. Here the weather is considered a binary variable where 0 is for temperature below 25 and 1 °C otherwise. The location is considered directly in terms of map coordinates.

Extending [11] for disruption length modelling Ghaemi et.al [7] creates a framework for three models: length prediction model, short-turning model and passenger assignment. Finally, Bayesian models have also been evaluated for duration prediction and found in [8, 12, 13]. In contrast, numerous studies for incident duration prediction are found in road transport context whose concepts could be implemented in railway context, mentioned in [14].

Regarding delay prediction, [15] defined a simple regression model used to analyse the dependence of the departure delays of a pair of interconnected trains on their arrival delays. The regression model explains the variation of both delays and how the delays are carried forward. To improve the model, the authors use a least trimmed squares approach to find the outliers in the data. [16, 17] also use regression-based models for finding delays. Some studies have investigated patterns and causality of the delays for understanding the spread.

Kono.et.al [18] use association rules to identify the causality of delays. They consider different types of primary delays to find out the secondary delays and establishes a relationship. Conte [19] presents a stochastic graphical modelling approach called the Tri-Graph. Linear regression and optimisation are used for identifying the inter-dependencies of delays. Finally, pattern mining is used extensively to find underlying patterns in delays [20, 21].

The use of machine learning techniques for delay prediction is also present in the pertinent literature. Yaghini et.al [22] introduces a methodology to predict railway passenger train delay using neural networks (NN) and eventually compare the results with decision trees and multinomial logistic regression. The NN models use different architectures, namely quick, dynamic and multiple, to implement the models. The dynamic architecture changes the NN network by adding units and then training the network until the best network with the desired accuracy is achieved. In contrast, the multiple architectures are training multiple networks in parallel to identify the NN with the highest performance. The study also uses data discretisation by converting the continuous variables into intervals and categorical variables. An accuracy range of 90–93% is achieved among all the models. The NN models with multiple architectures achieve the highest accuracy.

In a similar context, Nilsson and Henning [23] use NN for the prediction of delays on one of the high traffic line in Sweden. Weather data is combined with the departure schedules of the trains between two stations along the mentioned line. In addition to NN, Decision Trees with and without the Ada Boost model are built. The NN model outperforms the other two. Additional studies worth mentioning on the use of ANN are [22, 24–27]. Additionally, ANN was found to outperform the multiple regression models in the following studies [15–17, 28]. Wang.et.al [29] consider weather as a variable for the prediction using gradient boosting regression tree models. They use 3 months dataset of weather, train delays, and train schedule records to understand the patterns of train delays and predict long-term train delay time. The study is based on data from China. They use a density-based clustering algorithm (DBSCAN) to identify the time interval threshold to determine whether the delay of a train at a given station propagates to the following train. Gradient-boosted regression trees model (GBRT) was used in this study to build the prediction model for train delays.

Deep learning methods are found to be used extensively in literature. Wen.et.al [30] developed a long short term memory model (LSTM) to predict the delay time while considering the interactions and delay propagation. Actual running data from a section of lines in the Netherlands was used for the study. They consider seven independent (input) variables like arrival and departure times at current and previous station and running times to name a few. They compare the results to neural networks and random forest models that were also built. It was found that the LSTM model gave the best performance. Similarly, Huang et al. [31] present a study using deep learning methods for a case study of train delay prediction of four railway lines in the Chinese Railways. For this purpose, they develop a deep learning (DL) approach that combines 3-dimensional convolutional neural networks (3D CNN), LSTM recurrent neural network, and fully-connected neural network (FCNN) architectures and names it as CLF-Net. Using performance metrics of root mean square error and mean absolute error, the proposed CLF net models outperform the conventional deep learning model. Oneto .et.al [32] present a big data-based train delay predicting system using a supervised learning framework of extreme learning machines. They use shallow and deep extreme learning machines for predicting train delays for large scale railway networks by treating it as a time series forecast problem where every train movement represents an event in time. From January 2016 to June 2016, 6 months of TM records from the entire Italian railway network data were used as validation and compared with the in-place system used for predicting delays. However, Huang.et.al [33] address the prediction problem from another direction. They propose a hybrid model machine learning model of support vector regression and Kalman filter to improve the train running time prediction accuracy during railway disruptions.

Using Bayesian networks, Corman and Kecman [34] use a stochastic train delay prediction model, which predicts the probability distribution P of the random variable that describes an event. The time-dependent random variable is described based on the train delay over time and space. A basic linear regression is carried out on the response and the explanatory variables for selecting the predictors. Based on goodness of fit (R^2) values from these linear models, correlations between the response and the predictor is studied.

Lessan.et.al [35] use three different Bayesian networks, namely, heuristic hill-climbing, primitive linear and hybrid structure, for a high-speed line. The first basis for performance evaluation was the comparison between the predicted and the actual values and then using scatter plots and distributions plots at each station to understand how good the predictions are compared with the original; with 56% accuracy, increased to 80% after the application of discretisation.

Huang et al. [36] use Bayesian networks for predicting the three components, namely, primary delay, number of impacted trains and the total delay times. They use real-time data from the high-speed railway lines of China based on a 5-fold cross-validation method. The BN model predicts both the spatial and temporal propagation of interruptions on train operations. Eventually, four structures are compared: BN structures, integrating expert knowledge, the inter-dependencies learned from real-world data, and real-time prediction and operational requirements [36]. The BN model was the best and outperformed all the conventional predictive models.

Ulak et al. [37] also estimate a Bayesian network model and metrics for quantifying delay dependencies between transit stops. The data originated from a real-time transit information app that is used for the Long Island Rail Road.

The studies mentioned above use various variables for their models apart from the visible railway-related variables like weather, location, cause, passenger behaviour and many more. Railway operations are prone to be affected by various external parameters, like weather and location [11, 23 29]. Xia.et.al [38] analyses the effects of weather on railway operations in the Scandinavian region which faces one of the harshest climates. Standard linear regression models are implemented and levels of a wind gust, precipitation, and snow. The temperature difference between the maximum and minimum values within a day and the measure of proper temperature are the variables included in the study. They are found to have a strong positive correlation with railway infrastructure disruptions.

Zakeri and Olsson [39] measure the punctuality of railway services under extreme weather by using Pearson's coefficient of correlation and regression analysis in the Norwegian region. The results show the highly significant relationship between punctuality and the sum of snow depth leading to the deduction that snow depth best explains the variations in punctuality. A reference incident used for this study was the accident that happened on the "Great Belt Bridge" in January 2019 in Denmark. Its geographical position made the authors pursue the importance of the location of an incident in terms of centrality indicators. Network centrality is a measure of the importance of any node in a heterogeneous network compared to other nodes. There are various metrics to measure the centrality, which can be used to characterise complex networks.

Some of the most widely used metrics are the degree, closeness and betweenness centrality. degree centrality is considered the most straightforward centrality measure, which can be defined as the number of connections or nodes attached to a particular node. The closeness centrality of a node measures its average farness (inverse distance) to all other nodes. Nodes with a high closeness score have the shortest distances to all other nodes. Betweenness centrality measures the extent to which a vertex lies on paths between other vertices. Vertices with high betweenness may have considerable influence within a network under their control over information passing between others [40].

Barthelemy [41] also uses the betweenness centrality(BC) in large networks and studies the critical nodes and the impact of removing such nodes on the neighbouring ones. Sybil [42] also uses BC in evaluating 28 metro systems around the world, focusing on the global trends in those areas, concluding that considering centrality in the planning process can be valuable for better distribution of passenger flows. Finally, Erath [43] deduces that closeness and betweenness centrality are essential in the network analysis of complex road and railways networks.

The studies discussed above show the importance of predicting delays and incident duration and the investigation of a diverse set of factors that influence them. However, establishing the relationship between the duration of an incident and the total delay caused by it is rarely seen. Based on the literature review, the base data available for the study and the possible influencing variables were selected based on railway expert advice. The variables are explained in detail in the next section. Also, various model types were selected based on the frequency of use and performance in the literature. At the start of the modelling process, a large variety of models were tested. Finally, based on these initial modelling results and from the literature review, linear regressions, generalised linear models, gradient boosting models and neural networks were selected for further modelling processes.

3 | DESCRIPTIVE ANALYSIS

Railway incident data spanning from September 2017 to February 2019 were extracted from the Danish Railway database. It contained information of incidents all over the national network, as well as the Copenhagen commuter lines. It is worth noting that data was manually entered by the traffic operators and no automatic data acquisition was available. The railway-related variables used from the base data were incident duration, total delay by the incident, number of trains impacted, severity, type of the incident, date, time and incident location. The whole country was divided into areas based on the region, lines and traffic operators, the associated variable is named area code. Using these classifications, the average headway and the track-type were assigned to each area, which was also obtained from the Danish Railway database and from railway expert supervision. For temporal variables the data was divided into different variables like day of the week, month and hour of the day to find any possible cohort. All three variables were considered as categorical variables. Initially hour was considered checked as both categorical and continuous. After several modelling attempts using "hour" as a categorical variable improved the model. It was found to be important variables in those models.

It should be noted that traffic operators use standard notation, set by the Danish Railway, to assign the levels for severity and type code to an incident based on the following criteria and definitions. Each level of type code indicates a type of cause of the incident. Each level of severity signifies the kind of severity of the incident in terms of requirement of manual assistance or not.

The different types of type codes:-

- 1 - Errors in planning and dispatching
- 2 - Error/mistakes by drivers
- 3 - Incidents in freight trains
- 4 - Material errors in engines and wagons
- 5 - Trains with inbuilt engines
- 6 - Passenger or train guard related incidents
- 7 - Signalling and interlocking errors
- 8 - Train operating companies
- 9 - Accidents due to weather or unexpected external influences

The different levels of severity:-

- Severity 1 - Incidents causing major impact, requires manual assistance for solving the problem.
- Severity 2 - Incidents causing impact less than the level one, also requires manual assistance.
- Severity 3 - Incidents causing much less impact and can be fixed without manual assistance.
- Severity 4 - Incidents that do not require manual assistance at all and the impact is negligible.

Severity and type code are recorded, when the cause of an incident is known as it helps assess the required intervention. According to Danish Railway experts(Jacob Nielsen, personal interview, May 2019)., the severity and type codes of the incidents are recorded immediately, based on preset rules, if the cause of an incident can be detected in their systems (Technical problems like signalling or interlocking problems). In other cases, information from the drivers helps determine the initial

TABLE 1 Complete overview of variables

No.	Variable	Type	Levels/values	Units
1	Total delay	Continuous	Numeric	Hours
2	Duration	Continuous	Numeric	Hours
3	No. of trains impacted	Continuous	Numeric	Count
4	Headway	Continuous	Numeric	Hours
5	Track-type	Factor	3 levels	-SingleTrack(1), -DoubleTrack(2), -Junction(0)
6	Hour	Factor	24 levels	Number
7	Day of the week	Factor	7 levels	7 days of the week
8	Month	Factor	12 levels	12 Months
9	Severity	Factor	4 levels	1 to 4, 1 being most severe
10	Temperature	Continuous	Numeric	Degrees celsius
11	Betweenness centrality	Continuous	Numeric	Number
12	Closeness centrality	Continuous	Numeric	Number
13	Degree centrality	Continuous	Numeric	Number
14	Type codes	Factor	9 Levels	1 to 9 each type of problem
15	Area codes	Factor	23 Levels	1 to 22 depicting each region

severity and type code of the incident, when they happen due to external factors, weather, or causes undetectable by their systems. Hence, the two-variable are essential and must be known for the prediction of duration and further total delay.

Additionally, weather and network attributes were considered as external factors. Weather data in terms of temperature was obtained from the darksky API [44]. Network data was defined as centrality measures, namely degree, betweenness and closeness centrality. The data for the centrality indicators were collected from Open Street Maps using python OSMNX.

A total of 13 input influencing parameters were considered for the study and two response parameters which have been shown in the Table 1.

Understanding the data and identifying possible irregularities and errors was the starting point for the analysis. Around 20,000 incident entries were recorded in the Danish Railways, spanning from September 2017 to February 2019. As the focus of this paper is on operational modelling aspects, duration values higher than 750 h and total delay higher than 120 h were removed. The 750 h threshold was chosen as it amounts to 1 month. Similarly, the 120 threshold was chosen as it was found to be the maximum total delay for an incident duration of 750 h. One reason for considering these values was to understand the variations between duration, total delay and impacted trains. Finally, cleaning the data was done based on its correctness, availability of all the essential data fields, the availability of duration, and the total delay for the respective entry.

A descriptive analysis was carried to understand the variables and their relationships. The duration and the total delay were studied. Figure 1(a, b) provide an overview of the whole dataset for the two variables. It is evident from the figures that the data for both variables vary significantly. However, the densest regions were found at the lower values of the variables.

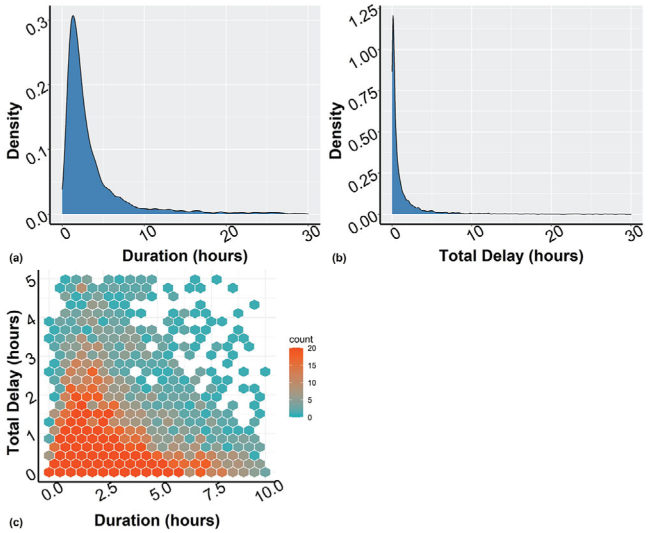

FIGURE 1 Total delay and duration relationships. (a) Density graph for duration. (b) Density graph for total delay. (c) Count plot between total delay < 5 h and duration< 10h

Figure 1(c) is focused on some of the denser regions showing the variation between duration and the total delay when values were restricted to less than 5 and 10 h, respectively. 93% of the incident entries had total delay values less than 5 h; about 4% had a value between 5 and 10 h, and only 1% of entries had a delay value greater than 20 h. In the duration data, 80% of the incidents had values less than 5 h, and about 11% of the incidents had values in between 5–10 h, 3.3% of values between 10–15 h and about 7% values greater than 20 h.

Furthermore, it is worth mentioning that some incidents with high total delay were of low duration. After analysing a sample of these incidents it was found that they occur at critical

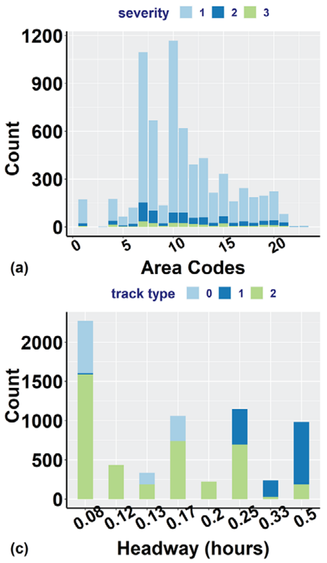

FIGURE 2 Incident frequency w.r.t. different parameters. (a) Incident frequency per area and severity. (b) Incident frequency per Month and Type. (c) Incident frequency based on headway

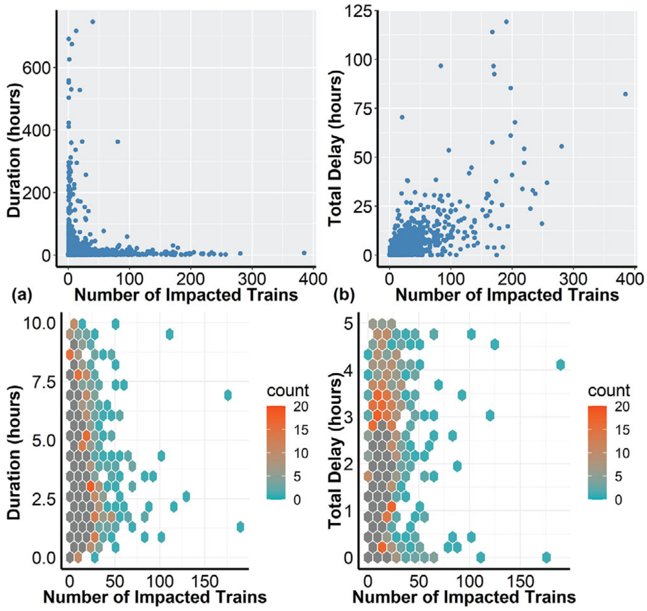

FIGURE 3 Number of Impacted Trains and relationship with other variables. (a) Impacted trains versus duration, (b) impacted trains versus total delay, (c) impacted trains versus duration < 10 h , (d) impacted trains Vs total delay < 5 h

junctions or lines with high traffic. In such areas, the headways between trains could be small, or these are critical areas prone to quick delay propagation. In such cases, though an incident is resolved quickly, it will still cause a high total delay as the impacted trains might be more.

In contrast, incidents with high duration and low total delay are also found in the data, requiring significant corrective measures. In such cases, though there are initial delays, overtime total delay reduces as new/rearranged operations run on changed schedules, leading to no additional delay, even though the duration increases.

Upon investigating other variables, Figure 2(a) in terms of areas where the incidents occurred and the severity highest number of incidents were found in area 11 and area 7. Area 11 represents the area near the borders with Germany in the region Padborg. Area 7 represents the S Bane network (commuter lines) and the central railway network of the capital city of Copenhagen. The Copenhagen region experiences the highest traffic and ridership in the country. It was observed that incidents of severity type 1 amounted to up to 86%, indicating that most of the incidents were delay causing incidents. The high traffic could attribute to high delay propagation since the head-ways between the trains were small.

When classified based on the type of the incident in Figure 2(b), most of the incidents belonged to the category of 7(85%) and 9(13%), which were incidents related to interlocking/signalling problems and external influences like weather, respectively. Additionally, an increased number of incidents was observed during winter months, making it the reason for considering the weather-related attributes.

Figure 2(c) shows the relation between the incidents frequency and the headway. 61% of the incidents happened on double tracks, 22% on single-track and the remaining on junctions(17%). Most incidents in the lower headways (≤10 min) take place on double-tracks. It implies that there are more chances of incidents happening or subsequent delay propagation with shorter headway. All the incidents on the junctions were with smaller headways. The majority of the incidents on single tracks had longer headways (10 min), implying that most of the lines with higher headway have single line tracks as the frequency of trains is lower.

Critical to this evaluation is the number of impacted trains, which determines the actual impact of an incident. Figure 3(a) illustrates the relationship between total delay and the number of impacted trains. A possible direct relation is observed; as the total delay increases, the number of Impacted Trains is also increasing. It can be seen that the majority of the incidents have a total delay that is lower than 25 h. They amount to 10% of the total impacted trains belonging to incidents with total delay >24 h. Also, the majority of the incidents have impacted trains <100 trains. 98% of the incidents have impacted trains <100.

Calculation of impacted trains requires the duration of an incident. Hence, it can not be used in the prediction of duration. However, understanding the relationship between the two variables could give valuable information. Figure 3(b) shows an "L" shaped curve; i.e. a lower number of impacted trains is observed with higher duration and vice versa. The possible explanation for such a relationship could be that the duration increases over time, even though the incident is not resolved. However, the operation in the affected area could have been handled already, resulting in no further impacted trains. Similarly, it also explains the case where the duration of an incident is high, but the total delay is low.

However, there are some uniform values to the lower part of the vertex of the "L" curve. The majority of incidents (61%) had duration < 5 h and impacted trains < 10. Only 6% of the

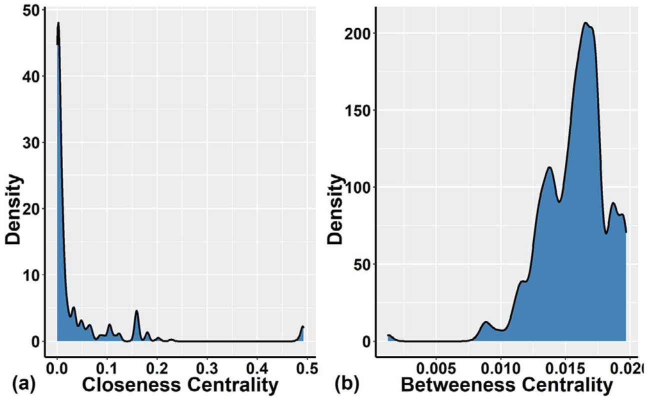

FIGURE 4 Centrality indicators. (a) Density graph for closeness centrality. (b) Density graph for betweenness centrality

incidents had duration ≥ 5 with impacted trains ≥10. However, the number of impacted trains amounted to 23% of the total impacted trains, of which 3% belonged to incidents longer than a day. Figure 3(c,d) provide a detailed view of the impacted trains when the duration is less than 10 h, and total delay is <5 h.

Another set of essential variables considered was some centrality measures (betweeness and closeness centrality, in particular). In practical terms, the higher the betweenness centrality of a node, the more central it is, when considering how many shortest paths pass through the node. Whereas in the case of closeness centrality, the higher the value the less central were the nodes (in terms of closeness to other nodes). Figure 5 shows how the closeness centrality varies over the whole network of Denmark.

In Figure 4(b) the higher betweenness centrality, the longer is the duration and the total delay, indicating that when critical nodes get affected, they cause further delay. In the case of closeness centrality, the smaller the value, the longer the duration explaining the longer total delays in the central regions. In a central location, the delay spreads quickly, adding up to a more significant number. The majority of the incidents occurred at nodes with medium values of degree centrality. Thus, these nodes were connected to enough neighbouring nodes to cause quick propagation of delay into the network.

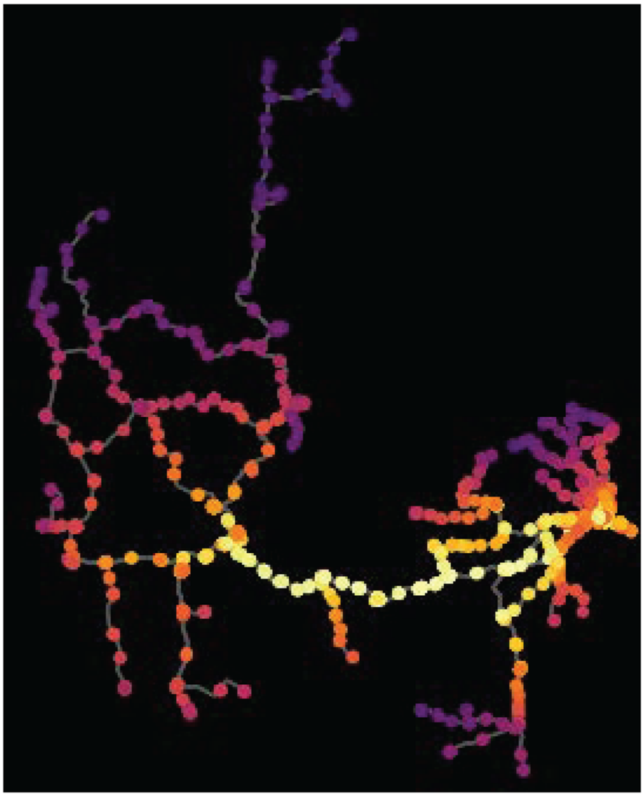

FIGURE 5 Closeness centrality variation over the network

TABLE 2 Multicollinearity check

Variable	GVIF	DF
Day of the week	1.128	6
Month	4.186	11
Track-type	2.417	2
Severity	1.253	2
Area codes	1.259	21
Hour	1.442	23
Duration	1.136	1
No. of impacted trains	1.056	1
Closeness centrality	1.335	1
Degree centrality	1.039	1
Betweenness centrality	1.145	1
Temperature	3.754	1
Headway	2.283	1
Type of problem	1.127	8

3.1 | Data preprocessing

Correlation analysis was also performed on all the variables. Since the variables were a mix of categorical and numerical types, the Goodmankruskal concept was used for finding correlations [45]. The results showed a few high correlation associations, but those associations were ruled out after further cross-checking for multicollinearity by computing variance inflation factor (VIF). The smallest possible value for VIF is 1, which indicates the complete absence of collinearity. As a rule of thumb, a VIF value that exceeds 5 or 10 indicates the presence of collinearity [46]. Table 2 presents the results where GVIF stands for generalized variable variance-inflation factor, and DF stands for degrees of freedom. From Table 2, the GVIF values for all the variable resulted in less than 5.

The next step was to check for outliers in the data. Since the data included wide-ranged values, the use of quantile outliers identification was deemed inappropriate. Cook's distance was used for finding the outliers. Cook's distance was used in regression analysis to find the influential outliers, negatively affecting the regression model. It measures the change in a regression model when an outlier is removed. A cut off distance of $\frac{4}{n}$ was used where n stands for the number of observations. Points lying above the cut off distance were considered as outliers [47].

The final step was normalisation. It was an essential step in predictive modelling as it improves the training process by making it more accurate and faster. Normalisation transforms the data into a standard format. Among the various normalisation methods, min–max normalization was used for numerical values. Binary membership coding (One Hot Encoding) was used for categorical values, as it corrects for bias due to the numbering order method [48]. After all the data processing steps, the final dataset was reduced to 6693 entries with 88 column variables. Finally, The near-zero variance method was used for finding the variables which could be considered removed from the dataset while modelling [49].

Finally, the data was divided into training and test dataset for further predictive modelling. The split percentage used for this study was 75–25%.

4 | MODELLING AND RESULTS

The main idea for modelling was to find the best models with high prediction accuracy for the duration of an incident and the total delay caused by it. An iterative approach was used wherein the hyper-parameters of the model types, significant variables and removal of of non zero variables were changed every iteration until the best model setup was found. The different models implemented were linear regression models, extreme gradient boosting machines, generalized linear models (GLM) and neural networks (NN). These model types were chosen based on the literature review and initial ensemble model results.

Among all the variables considered for modelling, the number of impacted trains and the total delay was not considered an input in all the duration models. It was essential to avoid unobserved variables in the modelling process, as such, determining the number of trains before knowing the duration of an incident was highly unlikely. Additionally, severity and type of incident(type code) were considered essential variables for initiating the framework to avoid uninformed assumptions on unobserved variables. Finally, all the variables were used in the total delay model estimation. For the evaluation of model performance RMSE (root mean squared error) and R^2 were used.

4.1 | Linear regression model

In the literature, many studies were found to use linear regression for understanding the dependencies between response variables and predictors, the behaviour of the data, to check the properties of the data, data linearity, homoscedasticity and, the significant predictors [15–17]. Linear Regression attempt to model the relationship between two variables by fitting a linear equation to observed data. One is considered as an explanatory variable, and the other is considered as a dependent variable. A linear regression line has an equation of the form $Y = a + bX$ where X is the explanatory variable and Y is the dependent variable. This form with one explanatory variable is called simple regression. Whereas, when multiple explanatory variables were used, then the model is called multiple regression.

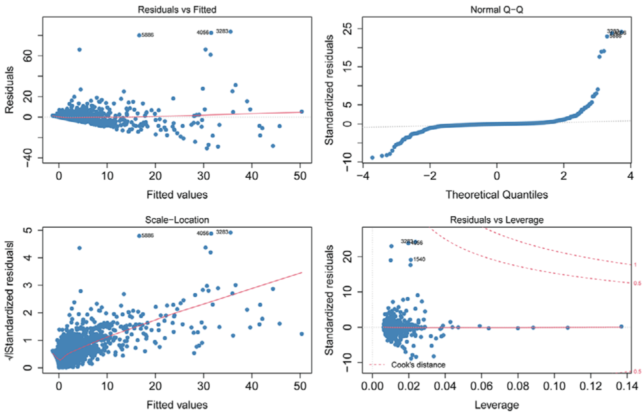

FIGURE 6 Diagnostics plot for linear model for total delay

The linear regression model used and discussed here was with multiple variables. In terms of total delay, the goodness of fit has a satisfactory value. However, diagnostic plots in Figure 6 suggest linear regression is not capturing the true relationship in the data. In the scale-location graph in Figure 6, the regression line is not horizontal. The points were concentrated at one end of the line indicating heteroscedasticity in the data, pointing to implementing other non-linear methods.

4.2 | Generalised linear models

Generalized linear models (GLM) models were also found to be implemented in the literature for duration prediction [10]. The output from the linear regressions suggested the lack of normal distribution for the residuals, prompting the implementation of the GLM, which unifies different distribution types under one framework. Unlike linear regression, the response variable is assumed to follow an exponential family distribution with mean μ_i, which was assumed to be some (often nonlinear) function of $x_i^T \beta$. It fits a wide range of linear, logistic and multinomial, Poisson and Cox regression models. GLM model was implemented using glmnet where the alpha value lies between 0 and 1, and lambda values range from 0 to infinity.

Model estimation for the duration models suggests poor fit, performing relatively low accuracy values. All the variables were considered without the removal of the influential points for this model. The λ and α values for the best model were 3.064 and 0.1, respectively. The best model for the duration had an RMSE of 0.041 with a test set RMSE value of 0.036. Similarly, the best model for the total delay had an RMSE of 0.027 for the training set and an RMSE of 0.035 for the test set. This gap between test and train RMSE indicates proneness to over-fitting for the model.

4.3 | Gradient boosting machines

Gradient boosting machine was implemented for its flexibility and the ability to yield high levels of accuracy, as indicated in the pertinent literature. Boosting was a method which improves the slow learners into a strong learner. In boosting, each new decision tree fit a modified version of the original data set. GBM models were gradient descent-based formulations, which works based on the negative gradient of the loss function [50]. Loss functions are selected based on the type of response variable, i.e. categorical or continuous. In this implementation, the loss function uses a learning rate that controls the model's run rate.

Among the various packages available for implementing GBM models, one of the most popular packages, XGBOOST (extreme gradient boosting), was used. For implementing this model, data preparation was required. The categorical variables in the data was converted into numerical, using the method discussed earlier. However, in decision tree implementations, normalisation was not implemented on continuous data. Also, boost takes the data in matrix format. The important hyper-parameters that were iteratively varied for the GBM models for finding the best model were the number of (decision) trees, depth of rows, learning rate and sub-sampling.

The hyper-parameters varied were maximum depth with values of 3,6,9, learning rate at 0.3 and the gamma value at 0, respectively. The best model for the total delay model was found at 100th iteration, a depth of 3, a learning rate of 0.3, a R^2 value of 0.599 and an RMSE of 0.029 on the training set were observed. Upon prediction for the test data, R^2 value was 0.447 and RMSE of 0.058 was observed. Variable importance for this model was also calculated, showing that impacted trains was the most crucial variable. Closeness centrality and the duration of the incident were the following essential variables. The duration was also an essential feature in the other models indicating it affected the total delay caused by an incident prompting in estimating the framework for the two response variables.

Similarly, the best duration model was found with a max depth of 8 with maximum iterations of 300 and a learning rate of 0.01. An RMSE of 0.039 and a R^2 of 0.112 was observed on the training set. Upon prediction on the test data, values of RMSE 0.038 and R^2 0.058 was observed.

4.4 | Neural networks

NN are popular machine learning techniques that simulate the mechanism of learning in biological organisms. NN is a series of algorithms that try to find the latent relationships in a dataset by mimicking how the human brain works. It refers to a system of neurons, organic or artificial. NN adapts to the changing input as the network generates the best possible result without changing the output criteria [51].

From the literature, NN was found to be widely used and yielding high levels of accuracy. As mentioned in [22], a mix of multiple and dynamic architectures was used here. Additionally, the components changed for each iteration to find the best model were variables, hyperparameters, data with and without outliers, number of neurons, increment of neurons, layers, removal/non-removal of the near-zero variables (nzv) and the training control options. The changes were done one after the other. The learning rate was varied from 0.0001 to 0.1. The number of training epochs to perform was set to 200. Feed-forward and standard back-propagation models were implemented. The activation function used for the model was rectified linear unit (ReLU) activation function. This function was used for its range from 0 to ∞ and converted all the negative values. Also, since the problem was a regression problem, the ReLU function helps in getting a quantified output.

Numerous model runs were performed to find the best model for both the total delay and the duration by using the methodology mentioned in the previous section. A double-layered model was chosen for the prediction of both the response variable. Because the RMSE results from the single-layered results reached a value of 0.067 (50.256). Variables were selected based on the results from the linear regression and the non zero variables.

Best model for the total delay model was a two-layer model with 28 neurons in the first and 19 neurons in the second layer with a learning rate of 0.05. Upon prediction of values on the test data, R^2 was 0.608, which was slightly more than that of the best train R^2 value of 0.596. The variables considered for the model were duration, number of impacted trains, headway, closeness centrality, betweenness centrality, degree centrality, temperature, month, track-type, day of the week, area codes, hour and severity. variable importance was also calculated for knowing which variable was impacting the response variable the most. Table 3 shows the variable importance values. The number of impacted trains was the most important the total delay the most. The following impacting variable was the duration proving that duration was an important variable in

TABLE 3 Variable importance for duration from neural network models

No.	Variable	Importance
1	Severity 1	100.000
2	Type code 9- weather & external problems	3.983
3	Temperature	3.332
4	Type code 7- signalling & interlocking problems	3.2920
5	Severity 2	3.185
6	Month 8 (August)	2.756
7	Headway (June)	2.692
8	Hour 13	2.511
9	Closeness centrality	2.215
10	Day of week (Tuesday)	1.591
11	Area code 10	1.357
12	Month 7 (July)	1.261
13	Betweenness centrality	1.101
14	Month 12 (December)	0.933

TABLE 4 Variable importance for total delay from neural network models

No.	Variable	Importance
1	No. of trains impacted	100.000
2	Incident duration	1.354
3	Severity 1	0.774
4	Severity 2	0.397
5	Temperature	0.385
6	Track-type 1 -single track	0.365
7	Severity 3	0.359
8	Track-type 0 - double track	0.316
9	Month 12 (December)	0.242
10	Month 1 (January)	0.236
11	Day of the week	0.211
12	4th hour of the day	0.188
13	Area code 10 - Copenhagen region	0.186
14	Headway	0.181

predicting total delay and justifying our initial consideration of the relation between the two variables.

In case of incident duration, a similar process was followed. Poor results were observed from these models. While numerous models were tested, the best model was with the highest value of R^2 value of 0.150 and least RMSE of 0.036. The best model was a two layered model with 25 neurons on the first layer and 33 neurons on the second. Variable importance was also calculated for the duration model, from results in Table 4, with severity to be the most important variable.

The following attribute was area code 10, the area of regional lines in the Copenhagen area. The third most important variable was closeness centrality. The more central the incident, the duration could be affected accordingly; the less central the node is, the higher the duration, and the more central the node is, the lower the duration. The variables used in this model were number of impacted trains, headway, closeness centrality, betweenness centrality, degree centrality, temperature, month, track-type, day of the week, area codes and hour.

TABLE 5 Consolidated models results

		Duration			
No.	**Model type**	**Train R^2**	**Test R^2**	**Train RMSE**	**Test RMSE**
1	Linear (normal)	0.129	0.114	0.043	0.037
2	GLM	0.133	0.060	0.041	0.036
3	XGBOOST	0.112	0.058	0.039	0.038
4	Neural network	0.150	0.172	0.043	0.073
		Total delay			
5	Linear (normal)	0.529	0.609	0.028	0.034
6	GLM	0.591	0.468	0.027	0.035
7	XGBOOST	0.599	0.447	0.029	0.058
8	Neural network	0.596	0.608	0.028	0.025
9	Estimation framework	-	**0.720**		0.038

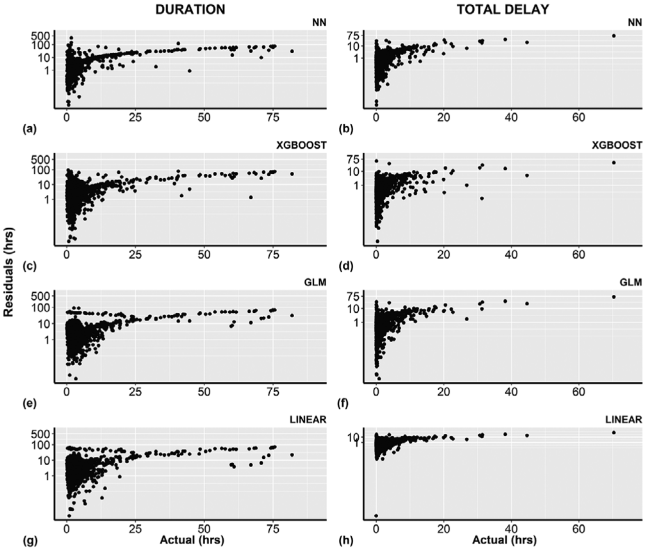

FIGURE 7 Semi-log graph between residuals versus actual values plots with log scale on y-axis. (a) Residuals versus actual values for the duration with all the models. (b) Residuals versus actual values for the total delay with all the models

4.5 | Evaluation

Table 5 presents the model metrics results RMSE and R^2 values of test and training data are presented. The presented RMSE results were normalised. Hence, the values are extremely low. Based on the results, it is observed that duration models seem to be weaker models as their RMSE are higher and low R^2. However, total delay models seem to perform better with lower RMSE and higher R^2.

To further assess the model performances, Figure 7 shows the log scale graphs with log on the y axis between residuals and actual values from all models for both the variables based on test data. It shows how good are the predictions at a given actual value of the variable. Higher residual points are observed in duration models than in the total delay model, indicating weaker models for the duration. All the models seem to work well for lower actual values where most residuals lie below 10 h. However, a plateau is observed in all the duration models for higher actual values, with higher residuals indicating that the models are weaker for predicting higher duration values.

The GLM and linear models show very similar performance. In these models, a distinct line of points for lower actual values is observed with higher residuals. They are indicating a mixed behaviour of weaker and stronger prediction behaviour.

NN and XGBOOST perform slightly better, showing no such distinct line of points for lower actual values. However, they show a slight dip in residuals for lower actual values. Few extreme residual points are observed in the NN model.

XGBOOST model seems to work the best for duration prediction among all the models. It has the lowest RMSE and the lowest difference between the test and the train RMSE as shown in the Table 5. Further, from Figure 7 XGBOOST model has fewer high residual points than the NN model on both lower and higher actual values.

In the case of total delay models, all models performed rather well based on RMSE values (Table 5). Their RMSE values are close to each other, indicating consistent performance from all the models. From Figure 7 it is evident that all the models have very low residual values for lower actual values (<1 h residual). Also, the residuals do not increase as the actual values increase. In the linear model majority of the residuals lie below 10 h. Though the graph looks promising for prediction, it could also point to underestimating values by the model.

Further, the models seem to perform relatively well on higher actual values, with most residuals lying lower the 10 h, indicating more robust performance on higher values by all the models. In NN, XGBOOST and GLM models, few higher residuals are observed at lower actual values. Further, the residuals seem to be slightly increasing as the actual values increases. Upon further investigation, the linear and GLM models are found to have a significant underestimation of values.

Based on the Table 5 and the Figure 7 it can be deduced that in predicting total delay, NN performs better than other models with the least difference between training and test RMSE (less over-fitting) and the slightly better ability to predict higher values. All the other models found higher differences between the test and the RMSE, indicating possible over-fitting/under-fitting models. However, to avoid any bias and maintain consistency in the framework, similar model types were considered for both variables.

4.6 | Estimation framework

Though different and separate models for the duration and the total delay have been created, best use of such a framework would be as combination of the two stand-alone models together. From Table 4, it can be seen that duration was an important variable in predicting the total delay of an incident. Thus, prompting the use of the duration model, together with the total delay model. The idea was to use the predicted duration values from the selected duration model as an input in the selected total delay model. Using the predicted duration values as input; the total delay values were predicted. Further, the predicted values were checked if predicted duration improves the total performance and the prediction itself. This builds up the framework for total delay prediction.

For the framework process, randomly selected 1000 incidents from the main data was used for the duration prediction using the best model. The results obtained from the framework can be seen in Table 5 in entry 9. It is evident that there is an decrease of almost 11.8% in the prediction accuracy between entries 8 and 9. This decrease could be attributed to the poor performance of duration models.

5 | CONCLUSION

This study aims at showcasing an estimation framework using a series of predictive models, which estimate two main disruption variables: duration of an incident and the total delay caused by it. These variables help quantify the impact of an incident on the network, contributing to the operations management and specifically dispatching in a Railway Traffic Management System. It enables better decisions making related to response operations after an incident. Another significant contribution, as mentioned earlier, could be in the area of the RDTMS system; upon incident/deviation detection, based on information from the discussed framework, the scheduler can be initiated. Further, the framework helps quantify the incident impact in terms of parameters, and the two parameters chosen explain a great deal apart from the prediction itself. The entire incident situation could be understood much better using the explained attribute relationships and the predicted values.

Different input parameters were utilised for the study collected from various sources (Table 1). All the inputs were considered as input in the total delay prediction model. However, the number of impacted trains was omitted as input from duration models as it could be calculated only after knowing the duration prediction with the help of railway schedules. Further, the total delay was also omitted as input to avoid any cross-referencing.

One of the study's main objectives was to understand the behaviour of the response variables about the attributes considered. The first part of the study was to determine the appropriate attributes for predicting the response variables. The results from the feature importance tables of the final models suggest that weather impacts both the response variables. Table 3 and Table 4 show that Temperature and the months December, January and February are considered essential for prediction. Type Code 9 (weather and external problems) was among the most critical variables in duration prediction, signifying the importance of weather. These observations make it evident that weather plays a vital role in such studies.

The duration model was also influenced by closeness and betweenness centrality. However, they had a lesser effect on total delay prediction. As defined in the literature review, the closeness and the betweenness centralities give the locational importance of a node in the network. It helps define the impact the current node causes to the nodes around, which could help define the duration of an incident. Further, it could help determine the number of impacted trains that significantly influence the total delay prediction (Table 4). In the case of the total delay model, the duration is the following best attribute.

Further, duration prediction could be initiated only after knowing the severity and type of incident. Consequently, in the feature importance Table 3, both the attributes were significant in predicting duration. In the case of total delay prediction, again, severity plays a crucial role. Since severity determined the need for assistance, it directly influences the duration, which was evident in the feature importance. Similarly, with the requirement of assistance, the duration increases, possibly increasing delay in most cases.

Another essential deduction was the prominence of the variable headway in both duration and delay models. The possible relation could be that a network with heavy traffic and less headway (high frequency) upon incident occurrence could result in network blockage, delayed arrival of assistance and higher delay propagation and delay accumulation per incident.

Further, track type was an essential variable in total delay but not in duration prediction. However, it does explain that the infrastructure type influences the total delay; incident occurrence on a single track system causes faster delay propagation. A good example is constructing the second main-line in Munich commuter lines [52]. A single line could cause increased delay propagation resulting in a higher cumulative delay. A double-track provides the possibility of diverting traffic to the operational line.

In the model implementation, it was evident that the total delay models had a better fit and performance than duration models. The poor performance from duration models could be attributed to the fact that there could be errors in the data itself unbeknownst to the authors since the data was entered manually. In general, all models perform well in predicting low values of duration and total delay. However, the lack of data for incidents with large values of duration and total delay makes it difficult for models to predict larger values correctly, making them weaker for such incidents. Nonetheless, collectively, the higher predictive power of the models and the estimating framework points out that the methodology discussed in this paper is a step in the right direction towards predicting network delay.

Finally, the estimation framework helps predict the total delay caused by an incident with slightly higher accuracy. The lower performance compared to the stand-alone models could be attributed to weaker duration models.

Regarding limitations and future work, the lack of clarity regarding the accuracy of the data, as it was manually entered data from the source, poses a threat to the correctness of the models estimated, prompting, at the same time, the need for automation in reporting of incidents. Additionally, more incident data and other attributes which are not considered in this study might help changed the performance of the current models. The low predictive power of the estimated duration models showcases that there is a lot of room for improvement in understanding the underlying factors that might affect the duration and the fact that there is a need to increase the accuracy of the reporting methods. When implementing the framework, it relies on few unobserved attributes that may not be readily available, which could delay the initiation of the model. Here lies a limitation of this framework for the practical implementation of the model. Obtaining the unobserved values requires advanced automated detection systems which are currently not available. Nonetheless, the model helps understand the factors that affect delays and their connection to incidents. This valuable information could help lay the critical foundation for creating a robust detection framework and its usage.

Additionally, the exact locations of the incidents could help improve the study even more, i.e. improved incident detection could help create more realistic models. Data from a more extensive network or a network with more traffic could be helpful in this study as the number of scenarios involved also change, making the models better at learning as there are more patterns to learn. One drawback in terms of improving the model itself, data discretisation is discussed in the literature review in [22] could be used to the data for improved results. It could be applied based on the area sections considered. It seems like a good approach as the categorical attributes are high, and the analysis could be carried based on each parameter level in a single model. Additionally, the discussed estimation framework could provide good support for dynamic railway traffic management systems.

Extensions of this work could include using different machine learning methods and an external validation process where system performance can be tested accordingly. Besides, an impact evaluation function that combines duration, network delay and other factors (i.e. capacity loss, emissions, influence on human, car damage) is considered a somewhat helpful direction.

ACKNOWLEDGEMENT

This research has been partly supported by the DAAD Project number 57474280 Verkehr-SuTra: Technologies for Sustainable Transportation, within the Programme: A New Passage to India - Deutsch-Indische Hochschulkooperationen ab 2019. The authors also appreciate the support of Banedanmark and Emch+Berger AG for making this project happen. The authors also acknowledge support by the German Research Foundation and the Open Access Publication Funds of Technische Universität Braunschweig.

ORCID

Bhagya Shrithi Grandhi https://orcid.org/0000-0001-7059-4404
Emmanouil Chaniotakis https://orcid.org/0000-0002-4523-9838
Constantinos Antoniou https://orcid.org/0000-0003-0203-9542

REFERENCES

1. Cacchiani, V., et al.: An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B: Methodological 63, 15–37 (2014)
2. Bešinović, N.: Resilience in railway transport systems: a literature review and research agenda. Transport Reviews 40(4), 457-478 (2020)
3. Nyström, B.: Aspects of Improving Punctuality: From Data to Decision in Railway Maintenance. Ph.D. Thesis, Luleå University of Technology, (2008)
4. Ghaemi, N., Cats, O., Goverde, R.M.P.: Railway disruption management challenges and possible solution directions. Public Transport 9(1–2), 343–364 (2017)
5. Placido, A., Cadarso, L., D'Acierno, L.: Benefits of a combined micro-macro approach for managing rail systems in case of disruptions. Transportation Research Procedia 3, 195–204 (2014)
6. Burr, T.: Reducing Passenger Rail Delaysby Better Management of Incidents. National Audit Office (2008), https://www.nao.org.uk/wp-content/uploads/2008/03/0708308.pdf. Accessed 12 May 2019
7. Ghaemi, N., et al.: Impact of railway disruption predictions and rescheduling on passenger delays. J. Rail Transp. Plann. Manage. 8(2), 103–122 (2018)

8. Boyles, S. et al.: Naive Bayesian classifier for incident duration prediction. Paper presented at Transportation Research Board 86th Annual Meeting, Transport Research Board, Washington DC, 21–25 Jan 2007
9. Jespersen-Groth, J. et al.: Disruption management in passenger railway transportation. In: Robust and Online Large-Scale Optimization, pp. 399–421. Springer, Berlin, Heidelberg, (2009)
10. Chandesris, M.: Dynamic and real-Time Prediction of Duration of Incident. Railway Research, The Global Reference for Rail Innovation, SNCF, (2006). Accessed 10 May 2019
11. Zilko, A.A., Kurowicka, D., Goverde, R.M.P.: Modeling railway disruption lengths with copula Bayesian networks. Transp. Res. Part C: Emerging Technol. 68, 350–368 (2016)
12. Pettet, G. et al.: Incident analysis and prediction using clustering and Bayesian network. In: 2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, pp. 1–8. IEEE, Piscataway, NJ (2018)
13. Li, D., Cheng, L.: Incident duration prediction based on bayesian network. Journal of Beijing Institute of Technology (English Edition) 19(SUPPL. 2), 119–123 (2010)
14. Li, R., Pereira, F.C., Ben-Akiva, M.E.: Overview of traffic incident duration analysis and prediction. European Transport Research Review 10(2), 22 (2018)
15. Goverde, R.M.P.: Punctuality of Railway Operations and Timetable Stability Analysis. Netherlands TRAIL Research School, Delft (2005)
16. Gorman, M.F.: Statistical estimation of railroad congestion delay. Transportation Research Part E: Logistics and Transportation Review 45(3), 446–456 (2009)
17. Seriani, S., Fujiyama, T., De Ana Rodriguez, G.: Boarding and alighting matrix on behaviour and interaction at the platform train interface. Paper presented at RRUKA Annual Conference 2016, SPARK, London, 3 Nov 2016.
18. Kono, A., Yakubi, H., Tomii, N.: Identifying the cause of delays in urban railways using datamining technique. In: Asian Conference on Railway Infrastructure and Transportation, pp. 227–230. Korean Society for Railway, Seoul (2016)
19. Conte, C.: Identifying dependencies among delays. Doctoral thesis, George August Univeristy of Goettingen, (2008)
20. Cerreto, F., et al.: Application of data clustering to railway delay pattern recognition. J. Adv. Transp. 2018, 1–18 (2018)
21. Cule, B. et al.: Mining train delays. In: Advances in Intelligent Data Analysis X, pp. 113–124. Springer, Berlin, Heidelberg (2011)
22. Yaghini, M., Khoshraftar, M.M., Seyedabadi, M.: Railway passenger train delay prediction via neural network model. J. Adv. Transp. 47(3), 355–368 (2013)
23. Robert, N., Kim, H.: Predictions of train delays using machine learning. Examensarbete Inom Datateknik, KTH Stockholm (2018), www.diva-portal.org/smash/get/diva2:1217917/FULLTEXT01.pdf. Accessed 10 May 2019
24. Malavasi, G., Ricci, S.: Simulation of stochastic elements in railway systems using self-learning processes. Eur. J. Oper. Res. 131, 262–272 (2001)
25. Peters, J. et al.: Prediction of delays in public transportation using neural networks. In: International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, vol. 2, pp. 92–97. IEEE, Piscataway, NJ (2005)
26. Pongnumkul, S. et al.: Improving arrival time prediction of thailand's passenger trains using historical travel times. In: 2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp. 307–312. IEEE, Piscataway, NJ (2014)
27. Marković, N., et al.: Analyzing passenger train arrival delays with support vector regression. Transp. Res. Part C: Emerging Technol. 56, 251–262 (2015)
28. Flier, H. et al.: Mining Railway Delay Dependencies in Large-Scale Real-World Delay Data. vol. 5868 LNCS of Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg (2009)
29. Wang, P., Zhang, Q.-p.: Train delay analysis and prediction based on big data fusion. Transportation Safety and Environment 1(1), 79–88 (2019)
30. Wen, C., et al.: A predictive model of train delays on a railway line. J. Forecast. 39(3), 470–488 (2019)
31. Huang, P., et al.: A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems. Information Sciences 516, 234–253 (2020)
32. Oneto, L., et al.: Train delay prediction systems: A big data analytics perspective. Big Data Res. 11, 54–64 (2017)
33. Huang, P., et al.: A hybrid model to improve the train running time prediction ability during high-speed railway disruptions. Safety Sci. 122, 104510 (2019)
34. Corman, F., Kecman, P.: Stochastic prediction of train delays in real-time using bayesian networks. Transp. Res. Part C: Emerging Technol. 95, 599–615 (2018)
35. Lessan, J., Fu, L., Wen, C.: A hybrid Bayesian network model for predicting delays in train operations. Computers and Industrial Engineering 127, 1214–1222 (2019)
36. Huang, P., et al.: A Bayesian network model to predict the effects of interruptions on train operations. Transp. Res. Part C Emerging Technol. 114, 338–358 (2020)
37. Ulak, M.B., Yazici, A., Zhang, Y.: Analyzing network-wide patterns of rail transit delays using Bayesian network learning. Transp. Res. Part C-Emerging Technol. 119, 102749 (2020)
38. Xia, Y., et al.: Railway infrastructure disturbances and train operator performance: The role of weather. Transp. Res. Part D: Transport and Environment 18, 97–102 (2013)
39. Zakeri, G., Olsson, N.: Investigating the effect of weather on punctuality of Norwegian railways: a case study of the Nordland line. Journal of Modern Transportation 26(4), 255–267 (2018)
40. Rodrigues, F.A.: Network Centrality: An Introduction. Springer, Cham (2019)
41. Barthélemy, M.: Betweenness centrality in large complex networks. Eur. Phys. J. B 38(2), 163–168 (2004)
42. Derrible, S.: Network centrality of metro systems. Plos One 7(7), e40575 (2012)
43. Erath, A., Löchl, M., Axhausen, K.W.: Graph-theoretical analysis of the swiss road and railway networks over time. Networks and Spatial Economics 9(3), 379–400 (2009)
44. Adam, G.: Dark sky api. (2021) https://darksky.net/dev. Accessed 6 May 2021
45. GoodmanKruskal: Association Analysis for Categorical Variables version 0.0.3 from CRAN. https://rdrr.io/cran/GoodmanKruskal/. Accessed 9 June 2021
46. James, G. et al.: An Introduction to Statistical Learning. Springer, Berlin, Heidelberg (2013)
47. Türkan, S., et al.: Outlier detection by regression diagnostics based on robust parameter estimates. Hacettepe Journal of Mathematics and Statistics 41(1), 147–155 (2012)
48. Kuhn, M.: 3 Pre-Processing | The caret Package. https://topepo.github.io/caret/pre-processing.html. Accessed 9 June 2021
49. nearZeroVar function, RDocumentation. https://www.rdocumentation.org/packages/caret/versions/6.0-88/topics/nearZeroVar. Accessed 9 June 2021
50. Natekin, A., Knoll, A.: Gradient boosting machines, a tutorial. Front. Neurorobot. 7, 21 (2013)
51. Aggarwal, C.C.: In: An Introduction to Neural Networks, pp. 1–52. Springer, Cham (2018)
52. Scheller, A.: The second S-Bahn trunk line in munich/die zweite s-bahn-stammstrecke münchen. Geomech. Tunnelling 8(2), 115–128 (2015)

How to cite this article: Grandhi, B.S., et al.: An estimation framework to quantify railway disruption parameters. IET Intell. Transp. Syst. 2021;1–13. https://doi.org/10.1049/itr2.12095