https://doi.org/10.53656/isct-2025.17

Transport Equipment

PREDICTION OF TRAIN DERAILMENT LOCATIONS USING ARTIFICIAL NEURAL NETWORKS

Milan Milosavljević[1], Biljana Milosavljević[2], Marko Bursać[1], Sandra Kasalica[1]
[1]Academy of Technical and Art Applied Studies Belgrade, Belgrade, Serbia,
[2]Serbian Railway Infrastructure JSC, Belgrade, Serbia

Abstract. Train derailments are undesired events in which at least one wheel of a railway vehicle comes off the tracks. Derailments can lead to severe consequences, affecting not only the regularity and continuity of railway traffic but also the environment. Therefore, predicting the potential location of a derailment is a crucial first step in prevention and increasing safety levels.

For this reason, it is essential to determine the causes of derailments and identify the locations where they tend to occur. The aim of this paper is to predict the location of train derailments using artificial neural networks. The input variables for the neural network are: year of the derailment, month of the derailment, derailment location, station, operational department, cause, and type of operation at the time of the derailment. The output variable of the neural network is the railway track number where a derailment is expected to occur.

The model presented in this paper was developed using artificial neural networks to forecast the railway line on which a derailment might happen, implemented through the Weka software tool. One dataset was used to train the network, while another dataset was used to test the model against real-world data from the accident and incident database.

A comparison between the model's results and statistical data on past derailments confirmed that the developed model provides reliable predictions. These results are of great significance, as identifying potential derailment locations enables railway infrastructure managers to take preventive actions, thereby reducing the number of derailments and directly enhancing the overall safety of railway traffic.

Keywords: railway; train derailments; safety; neural networks; Weka

1. Introduction

Train derailments represent undesirable events in which at least one wheel of a railway vehicle comes off the rails. These events may occur during train operations or the movement of shunting formations within stations during tasks such as forming new trains or redistributing railcars. Derailments can happen at any part of the railway infrastructure, though the most vulnerable areas include switch zones, station tracks, and open railway lines (especially in curves).

These unplanned and unwanted events can lead to very serious consequences, affecting not only the regularity and flow of rail traffic but also the surrounding environment. Therefore, predicting the potential location of derailments is the first step in prevention and improving safety levels. A derailment can directly cause an interruption of rail traffic, train delays, or the closure of a section of the railway line. Since dangerous substances of various types are frequently transported by rail, a potential derailment could result in severe consequences for the environment and the safety and health of people.

For this reason, it is essential to determine the causes of derailments and the locations where they tend to occur. The aim of this paper is to identify the potential locations of derailments based on a specific set of statistical data provided by the Serbian Railway Infrastructure JSC.

2. Literature review

In railway transportation, neural networks have been used to predict the number of sleeper cars [1], based on seven input parameters to determine how many railcars an operator has available.

Another study that demonstrates the use of artificial neural networks in traffic safety is by [2]. In that work, artificial neural networks and decision tree techniques were applied to identify the most common causes of road accidents in Nigeria.

The use of the Apriori algorithm and EM clustering in the Weka software tool to determine the causes of traffic accidents in the Alghat region of Saudi Arabia was the subject of a study by [3]. The aim of the research was to demonstrate the implementation of Weka in data processing techniques.

The study by [4] presented the possibilities of applying the Weka software tool for data mining tasks in Big Data analysis of sensor data in road traffic.

The connection between artificial neural networks and railway traffic safety was presented in the study by [5]. This paper examines the operational safety of urban railway subsystems using artificial neural networks. The authors compared 31 urban railway subsystems worldwide through two models.

3. Accidents and incidents in railway traffic

Traffic safety is a scientific discipline that deals with the study of harmful consequences of traffic and methods for their reduction. All events that have negative consequences on the operation of railway traffic are called accidents and incidents. An accident is an unwanted or unplanned event that has significant consequences, while an incident is an event that also negatively impacts the safety of train traffic but with lesser consequences.

All accidents and incidents in railway traffic in Serbia are categorized into two groups [6]:

Accidents during railway traffic operations, which include: collision of a train with a railway vehicle, collision of a train with an obstacle within the clearance profile, derailment of a vehicle or shunting composition, accident at a level crossing, accident-causing casualties caused by moving railway vehicles, fire and explosion and other accidents.

Incidents during railway traffic operations, which include: avoided collision of a train with a railway vehicle, avoided collision of a train with an obstacle within the clearance profile, passing a train past a signal that prohibits movement, broken rail, broken wheel or axle, track deformation, train separation and other incidents.

In addition to categorizing accidents and incidents by type, they are also classified by causes, which may be:

– Accidents and incidents caused by negligence of the infrastructure manager/ railway carrier:

– Human error during duty performance;

– Technical causes;

– Accidents and incidents caused by force majeure;

– Accidents and incidents caused by negligence of passengers and third parties.

Regarding the structure of the number of accidents and incidents, in the last four years, a total of 1,844 accidents and incidents have occurred on the railways of Serbia, averaging 461 annually, as shown in Table 1.

Table 1. Number of Accidents and Incidents in the Period from 2020 to 2023

Year	2020	2021	2022	2023	Total
Accident	208	231	227	212	878
Incident	193	287	246	240	966
Total	401	518	473	452	1844

A closer look at the structure of the number of accidents, which are more serious and therefore require detailed analysis due to the severity of their consequences, shows that derailments have the largest share among all accident categories, accounting for about 40% of the total number of accidents, or about 20% of the total number of accidents and incidents. The next category of accidents is those at level crossings, which make up about 20% of the total number of accidents annually. The number of derailments and their percentage share are shown in Table 2.

Table 2. Number of Derailments and Percentage Share in the Period from 2020 to 2023

Year	2020	2021	2022	2023	Total
Derailment	84	89	90	83	346
% Derailments in the total number of accidents	40	39	40	39	39
% Derailments in the total number of accidents and incidents	21	17	19	18	19

Derailments include cases when at least one wheel of a train or shunting composition slips off the rails. This category of accident can be both a cause and a consequence of an undesired event.

Namely, a train derailment can occur due to some technical defect in an element of the railway infrastructure or due to personal negligence. However, a derailment can also result from another accident, in which case the derailment is considered a consequence of that accident or incident.

The damage caused as a result of derailments, i.e., the direct costs incurred, amount to 165,741,448.38 dinars, which averages to 33,094,289.676 dinars over the four-year period. The number of hours of traffic disruption amounts to 4,214 hours, or 176 days, meaning that on average, due to derailments, some part of the railway or an entire railway line was out of service and traffic was prevented for approximately 44 days per year.

Considering the consequences derailments have on traffic flow, secondary delays of other trains, line closures, and negative impacts on infrastructure conditions – especially tracks and switches – it is very important to identify potential locations, i.e., railway sections, where derailments may occur based on available data. This will enable preventive actions, reconstruction of certain tracks or track sections to prevent such events.

For this reason, and based on the available data, a neural network model has been developed to predict the railway lines on which derailments are likely to occur.

4. Artificial neural networks

The architecture of artificial neural networks is based on a simplified model of the brain. Information processing is carried out by a large number of neurons (processing elements).

Neural networks consist of three types of nodes, which can be input, output, and hidden, depending on the type of neural network. All signals originating outside the neural network are received by the network through input nodes. On the other hand, output nodes send signals outside the network. In multilayer neural networks, besides these two types of nodes, there is also a hidden layer of neurons located between the input and output layers, where computation takes place. The structure of a neural network refers to the number of nodes and their connectivity. An artificial neural network is fully defined when its structure, activation function, and training method are specified.

Networks are trained by presenting them with a set of input values (usually along with a set of output values) in order to adjust the weights of the network branches. A large number of training algorithms have been developed, each with its own advantages and disadvantages. Characteristics related to modeling indicate the class of nonlinear functions that the network can accurately reproduce. The chosen network structure can affect the convergence speed during training [7].

5. Application of neural networks to the problem of determining the track where wheel slip will occur

Artificial neural networks are most commonly used for the following four types of problems [7]: classification, prediction, recognition and optimization.

In this paper, an artificial neural network model was developed that deals with prediction, specifically the model will provide an answer to the question of on which railway line train derailments can be expected based on given input parameters.

5.1. Data Collection

For the purpose of forecasting, a two-layer neural network was used, in which there is full connectivity between the input and hidden layers. All data for this paper were obtained from the internal database on statistical monitoring of accidents and incidents on the railway network in Serbia of the company "Infrastructure of Serbian Railways" a.d. The database used covers the last four years, from 2020 to 2023 [8].

Table 3. Dataset Used for Training the Neural Network

No.	Year	Month	Derailment location	Operational department	Derailment cause	Derailment location	Type of train operation	Track number
1	2020	1	Niš	Niš	Technical cause	Open track	Train running	128
2	2020	1	Ruma	Ruma	Human error	Point switch	Train running	101
3	2020	1	Velika Plana	Beograd	Technical cause	Station area	Train running	102
4	2020	1	Markovac	Beograd	Technical cause	Open track	Shunting operations	311
5	2020	1	Bačka Topola	Subotica	Technical cause	Point switch	Shunting operations	105
6	2020	1	Lapovo ranžirna	Niš	Technical cause	Point switch	Shunting operations	124

7	2020	1	Lapovo ranžirna	Niš	Technical cause	Station area	Train running	124
…	…	…	…	…	…	…	…	…
297	2023	9	Niš	Niš	Technical cause	Point switch	Shunting operations	102
298	2023	9	Subotica teretna	Subotica	Human error	Point switch	Shunting operations	105

The network consists of 322 instances, of which 298 instances were used for training and 24 for testing the network. The first set of instances used for training the neural network includes data on wheel slips from January 2020 to October 2023, while the last three months of 2023 (October, November, and December) were used for testing the network. This test set is completely independent of the training set and does not influence the training process, comprising about 7.5% of all data. The input layer has 7 nodes (attributes), and the output layer has one node. The training dataset is presented in Table 3.

5.2. Database Preparation

The input data for the neural network include the year and month when the derailment occurred, the location of occurrence, the operational department where the derailment happened, the cause of the derailment, the derailment location, and the type of train operation, while the output data is the track on which the derailment occurred.

The analysis covers the last four years from January to December, 2020 through 2023. The location of occurrence is an input attribute defined as the station or official place within whose area the derailment happened. The operational department covers a specific part of the railway network supervised and managed by a central office. Ten operational departments were used in this work: Belgrade, Novi Sad, Subotica, Zrenjanin, Pančevo, Niš, Zaječar, Kraljevo, Požega, and Ruma.

The cause of the derailment is an attribute relates to whether the derailment was caused by a technical problem, force majeure, passenger or third-party negligence, or human error of a railway employee.

The remaining two attributes are the type of train operation, which defines whether the derailment occurred during the operation of a train or shunting operations, and the derailment location, i.e., whether the derailment occurred on an open track, point swich or station area.

In addition to these data, the database contains more detailed information about the causes, damage caused by the derailment, traffic interruptions, number of slipped cars, but since these data could negatively affect the final result, they were removed through vertical selection.

Furthermore, horizontal data selection was performed to remove data entries with incorrect or abnormally distant values compared to other values of the same attribute.

The output attribute of the model is the railway line on which the derailment will occur. The track where the slip derailment is defined according to [9], which divides all tracks into five categories: 101 for main lines; 201 for regional lines; 301 for local lines; 401 for shunting lines and 501 for museum-tourist lines.

5.3. Model Development Using the Weka Software Tool

The problem was addressed using the Weka software tool. Weka is a data analysis and machine learning software used for exploration, model training, evaluation, and application of machine learning algorithms. It was developed at the University of Waikato in New Zealand, from which its name derives ("Weka" is an acronym for "Waikato Environment for Knowledge Analysis") [10].

Weka enables various data mining tasks such as data preparation, classification, regression analysis, clustering, association rule learning, relevant attribute selection, and data visualization.

The software supports different data formats like ARFF (Attribute-Relation File Format), CSV (Comma-Separated Values) and binary. It is also possible to import data from URLs or SQL databases.

After loading data into the software, certain filters can be applied to add or remove attributes, discretize, normalize, sample, etc.

In this paper, the data transformation method used involved normalization of attributes (min-max normalization) to improve algorithm accuracy and scale all attribute values to a range between 0 and 1. Besides attribute normalization, discretization was performed, which is the process of transforming continuous numerical attributes into discrete ones. This means continuous attribute values are divided into discrete ranges or intervals. This process is useful because it enables machine learning algorithms to work more efficiently with numerical data. After automatic discretization, continuous attributes are replaced by new attributes representing discrete intervals. This facilitates data analysis and improves machine learning algorithm performance.

Thanks to these capabilities, it is very easy to transform data from Excel tables into an ARFF file that is then loaded for use in the software. Once loaded, the database can be modified (corrected) within the Preprocess window, where editing can be done manually or automatically.

In this paper, three different algorithms were used: Multilayer Perceptron, Naive Bayes, and RandomTree. A set of 298 instances was used to train the network, followed by a set of 24 completely independent instances. Cross-validation and percentage split of the available dataset into training and testing sets were applied automatically as configured in the software (Figure 1).

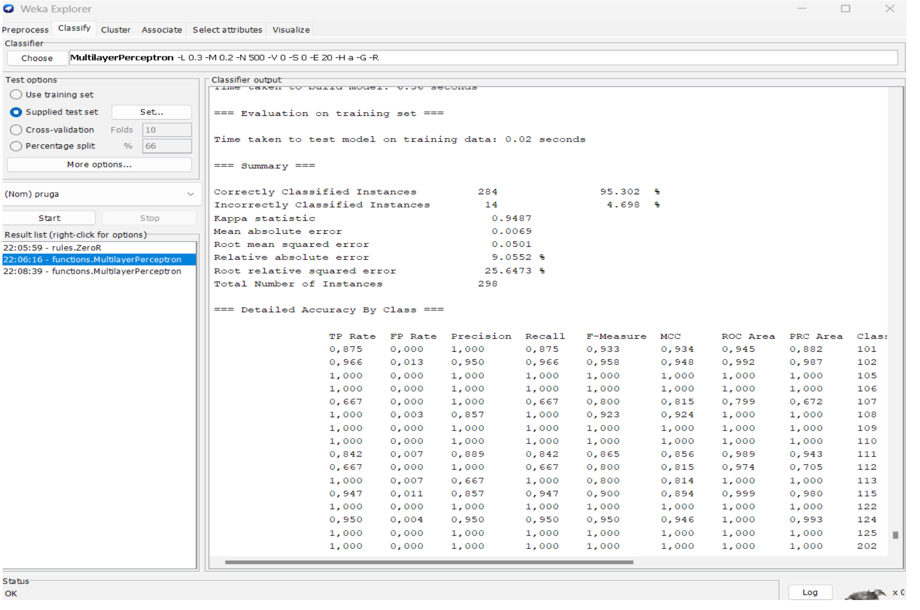

Figure 1. Window for Training and Testing the Network in the Weka Software Tool

6. Results and discussion

During the training phase using the Multilayer Perceptron algorithm, the artificial neural network predicted the track on which the wheel slip occurred with an accuracy of 95.3%, meaning that in 14 instances the classification was incorrect. When the dataset for testing the network was used, the correct prediction of the track where the slip would occur was 95.8%, with a mismatch occurring in only one case.

Table 4. Output Results of Three Prediction Models Applied to the Training Dataset

	Multilayer Percepcion	**RandomTree**	**Naive Bayes**
Correctly Classified Instances (%)	95,302	97,9866	71,8121
Incorrectly Classified Instances (%)	4,698	2,0134	28,1879
Kappa statistic	0,9487	0,978	0,686
Mean absolute error	0,069	0,0017	0,0352
Root mean squared error	0,0501	0,029	0,1284
Relative absolute error (%)	9,0552	2,1913	45,9124
Root relative squared error (%)	25,6473	14,8267	65,7395

When the model is tested using different algorithm variations, slightly different results are obtained. Table 4 shows the performance of three models based on three different algorithms applied to the training dataset. It is noticeable from the table that the model using the Naive Bayes algorithm performs the worst, exhibiting the lowest performance, while the best model is the one using the RandomTree algorithm, although the Multilayer Perceptron algorithm also achieved over 95% accuracy in classifying instances.

Table 5. Output Results of Three Prediction Models Applied to the Testing Dataset

	Multilayer Percepcion	**RandomTree**	**Naive Bayes**
Correctly Classified Instances (%)	95,8333	95,8333	91,6667
Incorrectly Classified Instances (%)	4,1667	4,1667	8,3333
Kappa statistic	0,9512	0,9512	0,9022
Mean absolute error	0,0105	0,0035	0,0262
Root mean squared error	0,0472	0,0417	0,0957
Relative absolute error (%)	13,8938	4,5977	34,6435
Root relative squared error (%)	24,6398	21,7571	49,9713

Table 5 shows the performance of three models using three different algorithms applied to the dataset for testing the network. In this case, the classification accuracy is the same for the models using the Multilayer Perceptron and RandomTree algorithms, while it is the lowest for the model with the Naive Bayes algorithm. This means that for the first two models, the result did not match in one out of 24 cases, whereas for the last model, there was a discrepancy in two cases.

Conclusion

The model results demonstrate high reliability, especially after testing, where the classification accuracy exceeded 90%. The best results were achieved by the model using the RandomTree algorithm for both training and testing datasets, with an accuracy of about 98%, meaning 23 out of 24 cases matched the real situation on the ground.

By applying the relief attribute method, the attributes were ranked according to their importance relative to the output variable – the track on which the slip may potentially occur. The ranking showed that the operational department has the greatest influence, while the year and month of the slip have the least. For this reason, excluding these two attributes should be considered in future model development.

The considered model can be expanded by adding several more input attributes and by including a larger database spanning a longer time period. Such an artificial neural network model can serve as an auxiliary tool for infrastructure managers when making decisions about the reconstruction or overhaul of certain tracks or track sections, all aimed at safer, more regular, and more reliable railway traffic.

The application of the model also lies in its universality and potential use in other types of accidents and incidents, as well as in identifying the causes of accidents or incidents.

REFERENCES

[1] MACURA, D., ŠELMIĆ, M., MILETIĆ, M., DIMITRIJEVIĆ, B., Neural network based model for assessment of the number of sleeping car in railway transport, SYM-OP-IS 2014, pp. 593 – 597, 2014, ISBN 978-86-7395-325-0.

[2] OLUTAYO, A., ELUDIRE, A., Traffic Accident Analysis Using Decision Trees and Neural Networks, I.J. Information Technology and Computer Science, Vol 02, pp. 22 – 28, 2014, ISSN 2074-9007. doi: 10.5815/ijitcs.2014.02.03.

[3] NAFIE ALI, F., HAMED, A., Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents, Journal of Information and Telecommunication, pp. 1 – 15, 2018, ISSN 2475-1839. doi: 10.1080/24751839.2018.1437142.

[4] JANKOVIĆ, S., MLADENOVIĆ, D., Applications Of Software Tools In Big Data Analysis Of Sensor Data In Traffic, XXXVIII Symposium on New Technologies in Postal and Telecommunications Traffic – PosTel 2020, Belgrade, 1 and 2. december, pp. 239 – 248, 2020, ISBN 978-86-7395-430-1.

[5] AWAD, F., GRAHAM, D., SINGH, R., AITBIHIOUALI, L., Predicting urban rail transit safety via artificial neural networks, Safety Science, Volume 167, 106282, pp. 1 – 18, 2023, ISSN 0925-7535, doi: 10.1016/j.ssci.2023.106282.

[6] RULEBOOK ON REPORTING, Investigation, Recording, Statistical Monitoring, and Publication of Data on Accidents and Incidents, Official Gazette of the Republic of Serbia, No. 32/21 of April 2021.

[7] TEODOROVIĆ, D., ŠELMIĆ, M., Računarska inteligencija u saobraćaju, Saobraćajni fakultet, Beograd, pp. 157 – 172, 2019, ISBN 978–86–7395–295–6.

[8] STATISTICAL DATA of the number of accidents and incidents on the railways of Serbian Railways for the period 2020–2023, Internal data, Infrastructure of Serbian Railways JSC, 2023.

[9] DECREE on the Categorization of Railway Lines Belonging to the Public Railway Infrastructure, Official Gazette of the Republic of Serbia, Nos. 92 of 29.6. 2020; 6 of 29.1.2021; 33 11.3.2022; 63 of 28.7.2023.

[10] SUMAM, S., PUTHIYIDAM, J., Evaluating Students Performance by Artificial Neural Network using WEKA, International Journal of Computer Applications (0975 – 8887), Volume 119, No. 23, pp. 36 – 39, 2015, ISSN 0975-8887. DOI: 10.5120/21190-4311.

✉ **Milan Milosavljević, Teaching assistant, MSc.**
ORCID iD: 0000-0003-2226-1765
Department School of Railroad Transport
Academy of Technical and Art Applied Studies Belgrade
14, Zdravka Čelara, Belgrade, Serbia
E-mail: milan.milosavljevic@vzs.edu.rs

✉ **Biljana Milosavljević, MSc.**
Serbian Railway Infrastructure JSC
6, Nemanjina, Belgarde, Serbia
E-mail: biljana.milosavljevic@srbrail.rs

✉ **Marko Bursać, teaching assistant, MSc.**
ORCID iD: 0000-0002-3453-2019
Department School of Railroad Transport
Academy of Technical and Art Applied Studies Belgrade
14, Zdravka Čelara, Belgrade, Serbia
E-mail: marko.bursac@vzs.edu.rs

✉ **Prof. Dr. Sandra Kasalica**
ORCID iD: 009-0002-0239-9687
Department School of Railroad Transport
Academy of Technical and Art Applied Studies Belgrade,
14, Zdravka Čelara, Belgrade, Serbia
E-mail: sandra.kasalica@vzs.edu.rs