

# Predicting the Probability of Cargo Theft for Individual Cases in Railway Transport

Augustyn LORENC, Małgorzata KUŹNAR, Tone LERHER, Maciej SZKODA

**Abstract:** In the heavy industry, the value of cargo transported by rail is very high. Due to high value, poor security and volume of rail transport, the theft cases are often. The main problem of securing rail transport is predicting the location of a high probability of risk. Because of this, the aim of the presented research was to predict the highest probability of rail cargo theft for areas. It is important to prevent theft cases by better securing the railway lines. To solve that problem the authors' model was developed. The model uses information about past transport cases for the learning process of Artificial Neural Networks (ANN) and Machine Learning (ML). The ANN predicted the probability for 94.7% of the cases of theft and the Machine Learning identified 100% of the cases. This method can be used to develop a support system for securing the rail infrastructure.

**Keywords:** artificial neural network; cargo theft; drone monitoring; machine learning; rail transport security; security support system; supply chain disruption

## 1 INTRODUCTION

Thefts from wagons expose railways to huge losses. Therefore, appropriate cargo security is of key importance for the entire transport chain [1-3]. The carrier is responsible for the cargo from the moment it is accepted for transport until its delivery to the recipient. Any damage must be paid for; it is often compensated for as part of the transport fee [4]. Stolen cargoes cause losses in excess of EUR 1.4 million annually in Poland [5].

The main products stolen in rail transport in Poland are coal and fuels. In the United Kingdom, one of the commodities stolen in transport is metal. The British government is tightening regulations which apply to scrap-metal dealers to prevent metal theft, control urban areas, industry, other offences and road accessibility [6]. In the United States of America, the government is working hard to protect the rail network against all forms of attack [7]. The regulations and the possibility of securing cargo are described *inter alia* by [8-10]. However, there are hardly any papers on how to predict theft in transport or use applicable systems. A method to predict the risk of theft in road transport similar to the one presented here was described in [11]. Rail carriers do not insure cargoes due to the gigantic costs of such insurance and the volumes of cargoes transported in a year.

The cargo transported by the rail could be secured by using closed wagons, and using door wiring, but it is a method just for slowing down the thief, and it never fully secures it. The better solution could use the padlock, but it will be problematic (the key or code should be transported to the final place, and the process of securing and opening the wagon will be complicated. So the technical security of wagons just slows down the thieves, and does not scare them away. Because of this, the better way is to use the drones as a solution to scare, take the pictures as evidence, and follow the theft. So that solution helps to avoid the theft or catch the thief. But with regards to the cost of that solution, it could not be used to monitor all transports. Because of that the method to identify the riskiest cases should be developed.

The aim of this research was to analyse the possibilities of using archive information about transport, transport theft, current transported type of cargo, value and to predict the possibility of theft. In the paper, the ANN and LM

method and the correlation of the input data with the results were examined. To achieve the goals the analysis of archive theft cases was done. In the next chapter authors present the theft cases that occurred in Poland, and the importance of the problem.

The ANN and ML methods were considered in the research, because of the complexity of the problem, high uncertainty, variety of transport processes and a number of data. The literature study proves that ANN and ML method can be successfully used for similar cases. The ANN is a very powerful tool with a wide range of application, like: transportation [12-14], vehicle routing [15], railway [16], electricity [17], hazardous risk [18] etc. In a different area of research, but also with the complexity of the problem, variety or uncertainty, this method helps to achieve good results.

## 2 IMPORTANCE OF THE PROBLEM

According to the information provided by PKP Cargo, the highest number of thefts occurs in Silesia - about 65÷70% of the total figure for Poland. In 2014, PKP Cargo recorded 1637 incidents related to thefts on the tracks. The value of the stolen goods was estimated at EUR 1.5 million. The area of Silesia is of key importance for rail transport in Poland because 60 percent of domestic and 70 percent of export consignments are sent from there [19].

Thieves steal goods from wagons not only during longer stops of trains, e.g. on railway sidings. Often a short stop before a semaphore or even slower train speed is enough. They try to open wagons and snatch the cargo after the train departs. Cargoes are protected by the Railway Protection Guards (SOK).

The following measures are applied to make cargoes more difficult to steal: painting with lime, wiring of wagons, monitoring of wagons with drones.

To protect consignments against theft, PKP Cargo uses two types of drones reaching a speed of 60 km/h. The technique enables sending the image in real time to the headquarters of the Threat Prevention Team. PKP Cargo uses Phantom III and Eagle drones. Phantom III drones are equipped with a 4K camera. Their task is to perform patrol flights of PKP Cargo trains and thus to prevent train stoppages by thieves. There are many ways of stopping trains - tipping coal, damaging wagons or placing blocks

on tracks. When a train stops, the drone records everything that is happening around. The task of Eagle drones is to patrol large areas from a high altitude, with two cameras. Thieves have something to fear because one camera provides very good quality images of faces even from several hundred meters. The other camera - a thermal imaging one - offers the possibility of night patrols. The operator is able to recognize a human figure from a distance of up to two kilometres.

With the use of drone-based surveillance, the number of theft cases decreased by 55%. After 10 months of using drones, the number of thefts in 2015 dropped to 1044, and the losses were estimated at EUR 0.75 million [19]. The share of PKP Cargo in the rail freight market in Poland is 44% [20].

According to a rail transport safety report produced by the Supreme Audit Office (NIK) in 2018, the number of offences and incidents decreased by approximately 33% [21].

Attacks on trains to steal cargo are usually scheduled – thieves know when and what train will be carrying valuable commodities. Therefore, their actions are usually well planned. Thefts are frequent when trains stop (due to blockades on the tracks and during planned stopovers). In addition, thefts occur also while the train is in motion – in such a case, the cargo is tipped from the wagons onto the track. For example, in 2014, PKP Cargo S.A. recorded theft of over 8000 tonnes of cargo [22] which, in relation to the transported 109.59 million tonnes [23] accounted for 0.75% of all transported goods.



Figure 1 Barricade on the track forcing the train to stop; wagons are opened to attempt to steal the cargo [19]

Tab. 1 presents the volume of cargoes transported by rail in Poland in 2017 and 2018. The grey rows mark the most frequent thefts. An analysis of quarterly fluctuations for particular product groups in a year indicated that they do not exceed 2.95%, therefore in rail transport, seasonality can be considered marginal.

Other groups presented in Tab. 1, according to the NST classification (percentages according to transport work) included:

- food, beverages and tobacco products - 0.49%
- textiles and clothing, leather and leather products - 0.01%
- wood, wood and cork products, straw, paper and paper products, printing products and recordings - 1.06%
- products made from other non-metallic raw materials - 1.47%
- machinery, electrical and electronic equipment - 0.07%
- transport equipment - 0.55%
- furniture, other finished goods - 0.16%
- secondary raw materials, municipal waste - 1.58%
- letters, parcels and courier parcels - less than 0.001%
- empty packaging - 0.75%

- cargoes transported, other non-tradeable cargoes - less than 0.001%
- mixed goods, excluding food - 33%
- unidentifiable goods - 8.51%
- other goods - 0.62%

Table 1 Rail transport in Poland in 2018

Cargo type	Share by transport work				Avg. dist. of 1 tonne km
	mln.t	%	mln. tkm	%	
Hard coal, lignite, oil and natural gas	66	39	10673	27	160
Metal ores, mining and quarrying products	46	27	11796	29	251
Briquettes, refined petroleum products	18	10	6051	15	330
Chemicals, chemical products	6	3	2170	5	333
Metals, readymade metal products	7	4	1796	4	252
Other groups according to NST classification	21.85	13	6910	17	314

Delays in deliveries are important from the viewpoint of efficient operation of the transport chain. Unfortunately, rail transport cannot be said to be highly punctual. According to the data provided by the Railway Transport Office (UTK), in 2018 the punctuality rate for freight deliveries was only 38% in domestic traffic and 28% in international traffic which, with an average delay time of 586 minutes in domestic traffic and 438 minutes in international traffic, is a major rail problem. In addition, due to the low speed of trains and numerous stops, there is an increased risk of cargo theft. Train delays are shown in Tab. 2.

Table 2 Train delays [24]

Delays	Number of delayed trains	Proportion of delayed trains
up to 4 minutes 59 sec	832	0.70%
5 minutes - 59 minutes 59 sec	9097	7.64%
60 minutes - 119 minutes 59 sec	8029	6.74%
over 120 minutes	57073	47.92%

### 3 METHODOLOGY

The industry meetings help to identify the problem of locating the riskiest areas to monitor them by the drones. Many company representatives said that it was not possible because of the costs to monitor each transport. And because of that, the main problem is to identify the places where the theft probability is the highest. Based on this, authors developed a method of predicting the theft probability with regards, inter alia, to the type of wagon, type of cargo, area, route, method of securing the wagon, organizing the transport process and more. There is a real demand for this research and theft prediction method.

To be able to predict when and where thefts of cargoes in rail transport may occur, an analytical model using artificial neural networks was proposed. The main factors affecting the occurrence of thefts were also identified. The input was based on statistical data presented in the UTK and PKP Cargo S.A. reports. The model which was developed could generate 10,000 transport cases. The data showed 0.73% theft cases which, in comparison with 0.75% of real cases, is a confirmation of the model's good

representation of reality. The input data were compiled in the following way:

- $P1$  - type of cargo. According to the data [3], thefts of coal concerned 65÷70% of all cases; normal distribution was used to generate the data with an interval of 1÷5:
  - Aggregates - marked 1
  - Metals - marked 2
  - Coal - marked 3
  - Fuels - marked 4
  - Other goods - marked 5

$$P1 = Z \left[ \text{Min} \left( 4; \text{Max} \left( 1; F^{-1}(R()); 3; 5/3 \right) \right); 0 \right] \quad (1)$$

- $P2$  - Type of wagons (2), based on data with the largest market share in 2017 [11], [17]:
  - Wagons of the normal type E - marked as 1, wagons of the special type F - marked as 2, number of used wagons type E and F: 58657 pcs
  - tank cars type Z - marked as 3, number of wagons operated: 7142 pieces
  - wagons with sliding roof type T - marked as 4, wagons with sliding walls S - marked as 5, a number of used wagons type T and S, approx.: 3253 pcs

$$P2 = \begin{cases} Rz(1; 2) & \text{gdy } P1 = 1 \text{ for } P1 = 3 \\ Rz(4; 5) & \text{for } P1 = 2 \\ 3 & \text{for } P1 = 4 \\ Rz(1; 5) & \text{for } P1 = 5 \\ 0 & \text{for rest cases} \end{cases} \quad (2)$$

- $P3$  - Train speed (3). According to UTK data, the average speed of freight trains in 2017 was 25.1 km / h [18], so for generating data, the normal distribution was used with the interval 20÷30 km / h

$$P3 = Z \left[ \text{Min} \left( 30; \text{Max} \left( 10; F^{-1}(R()); 25.1; 4 \right) \right); 0 \right] \quad (3)$$

- $P4$  - Method of securing the wagons (4):
  - No protection - marked as 0,
  - Indication of the load with lime - marked as 1,
  - Wagon wiring - marked as 2,
  - Increased SOK control - marked as 3,
  - Drones monitoring - marked as 4.

$$P4 = Rz(0; 4) \quad (4)$$

- $P5$  - Travel time at night (5), assumed as a percentage of the total route

$$P5 = Rz(1; 100) \quad (5)$$

- $P6$  - Region (6), was established in accordance with the administrative division of Poland:
  - Lower Silesia- marked as 2,
  - Kuyavian-Pomeranian - marked as 7,
  - Lublin Voivodeship - marked as 8,
  - Lubusz Voivodeship - marked as 9,
  - Lodz Voivodeship - marked as 10,

- Lesser Poland - marked as 3,
- Masovian - marked as 11,
- Opole Voivodeship - marked as 6,
- Subcarpathian - marked as 12,
- Podlasie - marked as 13,
- Pomeranian - marked as 5,
- Silesian - marked as 1,
- Swietokrzyskie Voivodeship - marked as 14,
- Warmia-Masurian - marked as 15,
- Greater Poland - marked as 4,
- West Pomeranian - marked as 16.

$$P6 = Rz(1; 16) \quad (6)$$

- $P7$  - Distance (7), base on the UTK data [25] the average distance of freight transport was 228.5 km in 2017, therefore a normal distribution in the interval was used to generate data. 100÷600 km

$$P7 = Z \left[ \text{Min} \left( 600; \text{Max} \left( 100; F^{-1}(R()); 228.5; 190 \right) \right); 0 \right] \quad (7)$$

- $P8$  - Number of planned stops (8), the generated data was assumed in the range 1÷10,

$$P8 = Rz(1; 10) \quad (8)$$

- $P9$  - Number of unplanned stops and delays (9), train delays were based on data from table 2 [24], the following delay intervals have been assumed:
  - up to 4 minutes 59 sec - marked as 1,
  - about 5 minutes - 59 minutes 59 sec - marked as 2,
  - about 60 minutes - 119 minutes 59 sec - marked as 3,
  - over 120 minutes - marked as 4,
  - arrival in accordance with the plan - marked as 5.

$$P9 = Z \left[ \text{Min} \left( 5; \text{Max} \left( 0; F^{-1}(R()); 3.4; 0.7 \right) \right); 0 \right] \quad (9)$$

### 3.1 Data Preparation for the Artificial Neural Network

For every  $P1 \div P9$  parameter, the appropriate factor to determine theft probability was estimated. This was done using Eqs. (10) - (18):

$$W1 = \begin{cases} 0.7 & \text{for } P1 = 3 \\ 0.4 & \text{for } P1 \in \{2; 4\} \\ 0.2 & \text{for } P1 = 1 \\ 0.1 & \text{for } P1 = 5 \\ 0 & \text{for rest cases} \end{cases} \quad (10)$$

$$W2 = \begin{cases} 0.7 & \text{for } P2 = 1 \\ 0.2 & \text{for } P2 = 3 \\ 0.4 & \text{for } P2 = 2 \\ 0.2 & \text{for } P2 \in \{4; 5\} \\ 0 & \text{for rest cases} \end{cases} \quad (11)$$

$$W3 = Z \left( 1 - \frac{P3}{30}; 2 \right) \quad (12)$$

$$W4 = \begin{cases} 0.7 & \text{for } P4 = 0 \\ 0.7 & \text{for } P4 = 1 \\ 0.5 & \text{for } P4 = 2 \\ 0.4 & \text{for } P4 = 3 \\ 0.3 & \text{for } P4 = 4 \\ 0 & \text{for rest cases} \end{cases} \quad (13)$$

$$W5 = Z\left(\frac{P5}{250}; 2\right) \quad (14)$$

$$W6 = \begin{cases} 0.7 & \text{for } P6 = 1 \\ 0.6 & \text{for } P6 = 2 \\ 0.5 & \text{for } 3 \leq P6 \leq 6 \\ 0.05 & \text{for rest cases} \end{cases} \quad (15)$$

$$W7 = Z\left(\frac{P7}{6000}; 2\right) \quad (16)$$

$$W8 = \frac{P8}{20} \quad (17)$$

$$W9 = \begin{cases} 0.05 & \text{for } P9 = 1 \\ 0.1 & \text{for } P9 = 2 \\ 0.2 & \text{for } P9 = 3 \\ 0.3 & \text{for } P9 = 4 \\ 0 & \text{for rest cases} \end{cases} \quad (18)$$

Based on the  $W1 \div W9$  factors, the target value was calculated. To calibrate the model for reproducing real statistical data to the maximum, the additional factor  $C_f$  was estimated. This factor was used to sort a number of risky cases by the highest risk probability. The  $C_f$  factor was calculated using the above formulas (19, 20).

$$count(realRiskyCases) = count\left(\frac{1}{9} \sum_{i=1}^9 w_i \geq C_f\right) \quad (19)$$

$$target = \begin{cases} 0 & \text{if } \frac{1}{9} \sum_{i=1}^9 w_i < C_f \\ 1 & \text{if } \frac{1}{9} \sum_{i=1}^9 w_i \geq C_f \end{cases} \quad (20)$$

For the model presented, the value 0,01065 for  $C_f$  was determined.

For the ANN learning process, the following set of teaching data was compiled:

- 70% of the collection - teaching data,
- 15% of the collection - validation data,
- 15% of the set - test data.

During data preparation, the following Matlab functions were used:

- Mapminmax - normalizing the input values to the range of -1, 1 (acceleration of calculations),
- Removeconstantrows - removal of input vectors consisting of the same values.

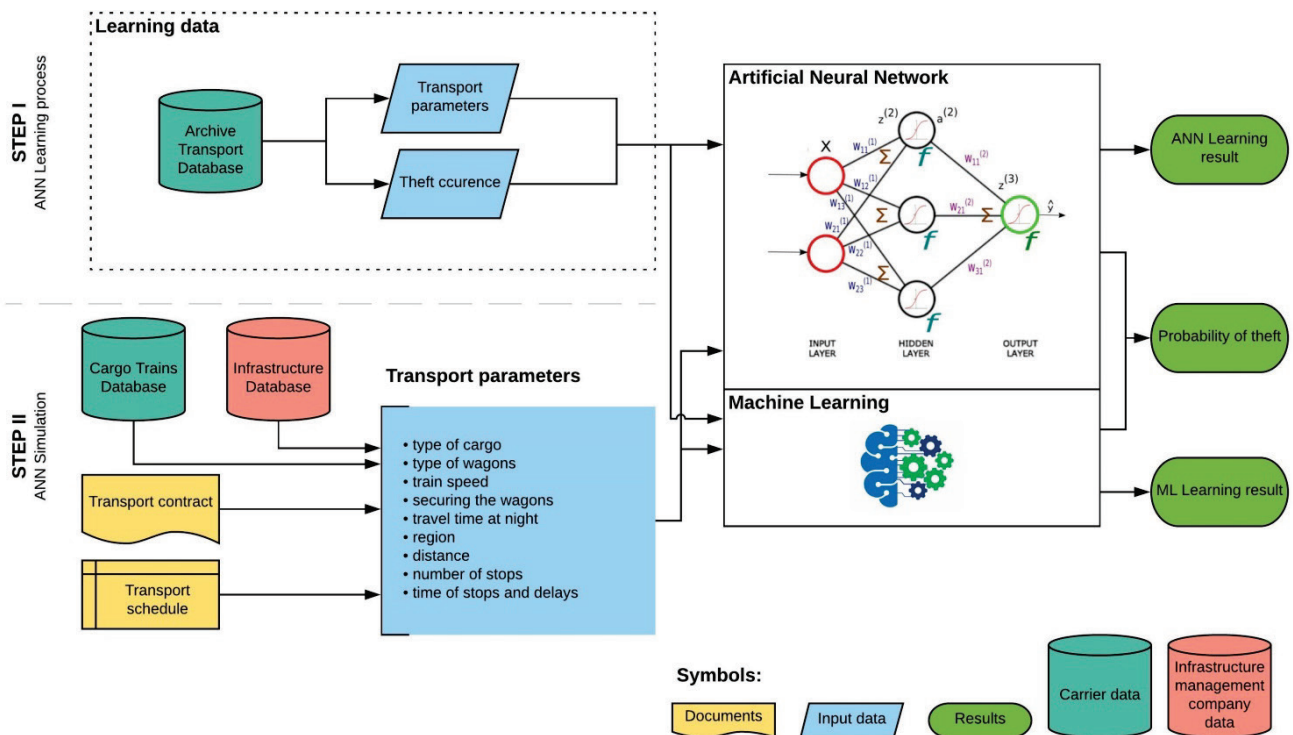


Figure 2 Method of theft probability prediction

The authors decided to develop a new method to predict theft cases in rail transport in the supply chain. One of the good tools which can be used for this is the Artificial Neural Network (ANN) and Machine Learning (ML). In this chapter, the authors show their method to predict

disruptions in road transport by using the Matlab ANN Toolbox and the Classification Learner in Matlab. This method consists of two steps: ANN/ML learning, Simulation of theft probability.

Fig. 2 shows the method, which was developed to estimate the probability of the occurrence of the theft risk and its potential foundation. The Matlab R2015 software was used to model the Artificial Neural Networks and Machine Learning.

As a first step, it was necessary to create a database containing data such as transport parameters and cases of theft. The transport parameters may include the cargo type, type of wagons, train speed, method of securing the wagons, travel time at night, region, distance, number of planned stops and the time of stops and delays. Based on this information, the probability of transported cargo theft was determined.

Table 3 Example of data set used for training ANN and ML

Input									Target
Type of cargo	Type of wagons	Train speed	Method of securing the wagon	Travel time at night / %	Region	Distance	Number of stops	Time of stops and delays	
P1	P2	P3	P4	P5	P6	P7	P8	P9	
3	1	28	1	14	4	205	2	3	0
3	1	18	3	58	1	100	10	4	1

In the simulation, a feed-forward network was used. The research applied a different ANN structure. For Machine Learning, a different algorithm was used. Tab. 4 presents some of the analysed variants. In the transport, parameters such as cargo type (P1), type of wagons (P2), train speed (P3), method of securing the wagons (P4), travel time at night as percentage of total route time (P5), region (P6), distance (P7), number of planned stops (P8) and the time of stops and delays (P9) were used as the

learning input data for ANN and ML. For the learning target, theft cases were applied and calculated using the formula (*target*). Every case of transport was represented as a vector. The examples of data used in the training are presented in Tab. 3.

In Tab. 4 is presented basic statistical information for training dataset. The value in each column was described at the beginning of the Methodology chapter.

Table 4 Statistical data for training dataset

Variable	Type of cargo	Type of wagons	Train speed	Method of securing the wagon	Travel time at night / %	Region	Distance	Number of stops	Time of stops and delays
	P1	P2	P3	P4	P5	P6	P7	P8	P9
Median	3	2	25	2	50	8	230	6	3
Min	1	1	10	0	1	1	100	1	1
Max	4	5	30	4	100	16	600	10	5

In real data, theft occurs in just 0.75% of the total number of cases. Because the overall accuracy could be high and the identification of theft cases could be poor, in order to make sure that the method generates a good return, the accuracy for theft cases must be considered.

The analysed variants presented in Tab. 5 differ in their overall accuracy and the accuracy for theft cases. It could be seen that for No. 1 (Tab. 5) the ANN method with 3 hidden layers of 9 neurons each, the overall accuracy is very high (99.3%) but the theft accuracy is 0%. The best results were achieved by using a simple ANN with 9 neurons in two hidden layers and the Machine Learning by Boosted Trees. It gave a 100% accuracy for theft cases.

Table 5 Analysed variant for the learning process

No.	Method	Parameters	Overall Accuracy	Accuracy for Theft cases
1	ANN No. 1	3 layers of 9 neurons each	99.3%	0.0%
2	ANN No. 2	1 layer of 9 neurons	98.6%	94.7%
3	ANN No. 3	2 layers of 9 neurons each	87.8%	98.4%
4	ML, Weighted KNN	Number of Neighbours: 10; Distance Metric: Euclidean; Distance Weight: Squared Inverse	99.4%	16.7%
5	ML, Weighted KNN	Number of Neighbours: 10; Distance Metric: Cosine; Distance Weight: Squared Inverse	99.4%	33.3%
6	ML, Fine KNN	Number of Neighbours: 1; Distance Metric: Correlation; Distance Weight: Squared Inverse	99.2%	55.6%
7	ML, Fine KNN	Number of Neighbours: 1; Distance Metric: Euclidean; Distance Weight: Equal	99.2%	38.9%
8	ML, Cubic SVM	Kernel Function: Cubic; Box Constraint Level: 1; Kernel Scale Mode: Auto	100.0%	77.8%
9	ML, Cubic SVM	Kernel Function: Quadratic; Box Constraint Level: 100; Kernel Scale Mode: Auto	100.0%	66.7%
10	ML, Boosted Trees	Ensemble method: RUS Boost; Learner Type: Decision Tree; Number of Learners: 200;	87.8%	100.0%
11	ML, Boosted Trees	Ensemble method: Bag; Learner Type: Decision Tree; Number of Learners: 200;	99.7%	66.8%

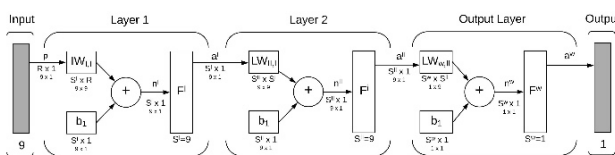


Figure 3 Structure of artificial neural network for case No. 3

The selection of the Artificial Neural Network structure was based on the method of subsequent

approximations. The artificial neural network which was created consists of one hidden layer of 9 neurons and one output layer.

$$a^w = f^3 \left[ LW_{3,2} f^2 \left( LW_{2,1} f^1 (IW_{1,1} p + b_1) \right) \right] + b_3 = y \quad (21)$$

where: *R* - number of elements in input vector; *S* - number of neurons in layer; *a* - neuron output vector; *IW* - initial

weights;  $LW$  - layer weights;  $p$  - vector of  $R$  input elements;  $f$  - sigmoid transfer function;  $b$  - bias value;  $y$  - output vector of ANN.

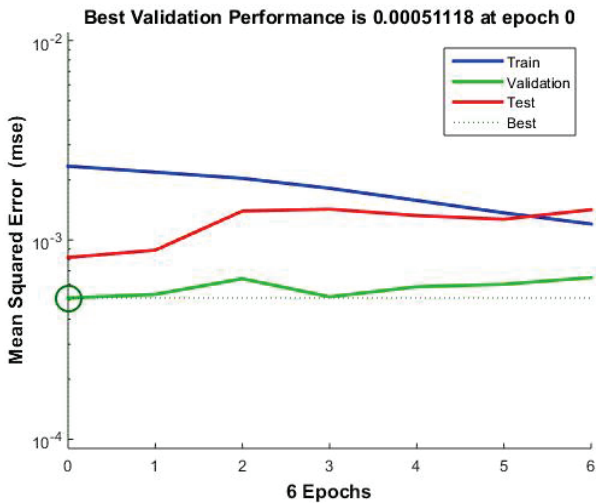


Figure 4 MSE error during the training of artificial neural network for case No. 3

The Levenberg-Marquardt backpropagation method was used as the learning algorithm. As an activation function, the tangensoidal function was used. The biggest neural network examined in the research gives the worst results. The network structure was selected on the basis of the mean square error (MSE) which was about 0.00051118 (Fig. 4).

The results achieved in the training, the validation and testing of neural network processes are presented in Fig. 5.

For Machine Learning, the Boosted Trees method was used. For its configuration, the RUS Boost ensemble method with 200 number of learners was used. Fig. 6 presents the ROC Curve for the ensemble method.

Fig. 6 indicates that the Machine Learning method generates high quality results.

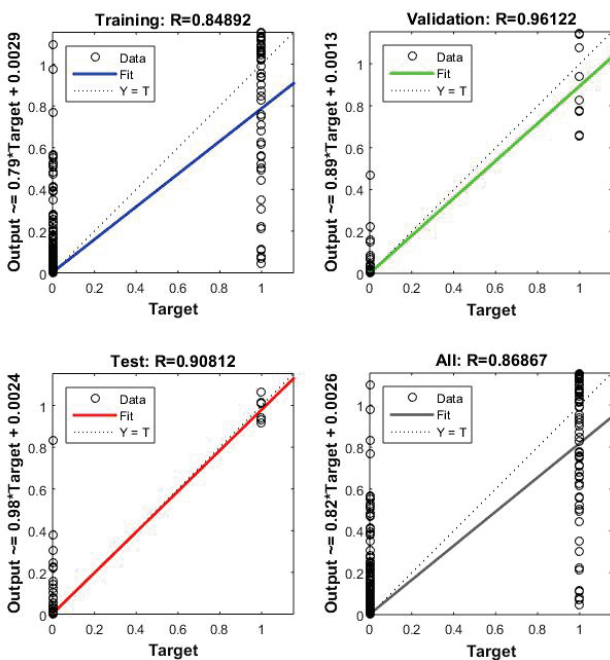


Figure 5 Regression plot for training, validation and testing of ANN for case No. 3

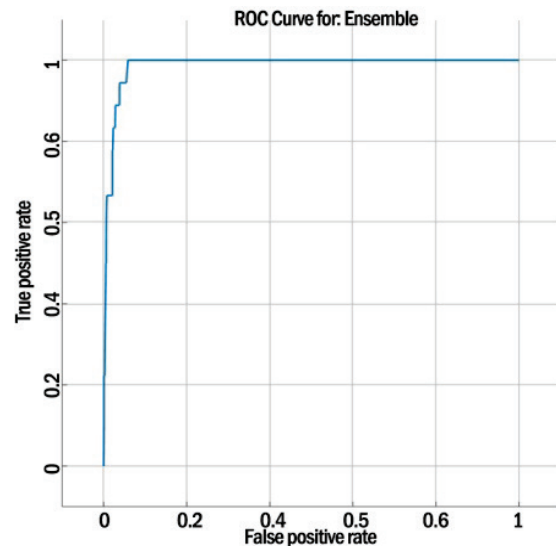


Figure 6 ROC Curve for ensemble method for case No. 10

#### 4 RESULTS

The methodology presented in the previous chapter was used to predict the theft probability for each rail cargo transport. The simulation used 10 000 sample transport cases characterized by the parameters as described ( $P1-P9$ ). In that sample, 75 theft cases occurred. The ANN predicted the probability for 94.7% of the cases of theft and the Machine Learning identified 100% of the cases. The method also predicted theft in cases where they do not occur in reality. That area is potentially risky and the railway infrastructure should be monitored there. That will reduce the number of theft cases.

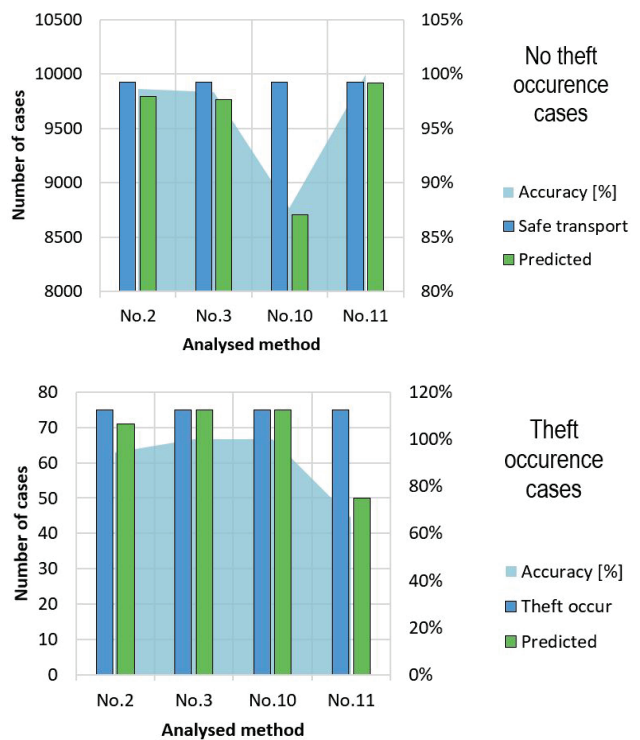


Figure 7 Results of theft prediction by the best-used method

In the cases of cargo transport under analysis, it was always coal that was stolen from standard  $E$  type wagons. To predict thefts from other cargo types, the input data set

must be bigger and they must represent more cases of theft. With the absence of such information in the reports, such cases cannot be simulated.

Theft cargo cases were detected for six Voivodships (P6); in those regions, the biggest number of incidents occurred in real data (1 - Silesian: 24 cases, 2 - Lower Silesian: 15 cases, 3 - Lesser Poland: 7 cases, 4 - Greater Poland: 14 cases, 5 - Pomeranian: 7 cases, 6 - Opole: 8 cases). The cases detected in correlation with the number of stops and distances are presented in Fig. 8.

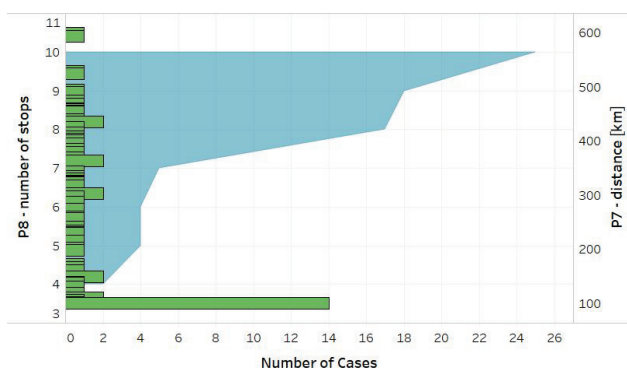


Figure 8 Predicted theft cases by the number of stops and distance

The correlation of theft cases to the train speed and travel time at night is presented in Fig. 9.

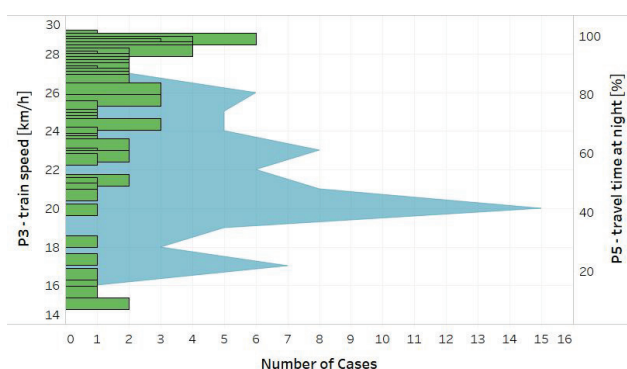


Figure 9 Predicted theft cases by train speed and travel time at night

The average transport speed was about 21 km/h, with a minimum of 15 km/h and a maximum of 29 km/h. Most cases occurred at night - about 76% of the route was during the night. The average distance was 276.4 km. 25 thefts were done during scheduled stops, and 80% occurred at eight or more stops. 45 thefts happened during unscheduled stops longer than 120 minutes, 27 cases at stops between 60÷120 minutes, and 3 cases at stops between 5÷60 minutes.

## 5 CONCLUSIONS

Cargo thefts are never accidental. Thieves plan to steal cargo from rail wagons taking many factors into consideration beforehand. They steal coal and similar commodities, which are easy to sell in the black market. Also it is easy to drop that cargo down from stopped wagons and take it directly to pick-up vehicles or dump it from moving wagons on the trackway and load it later onto the vehicles. Thieves plan their actions to happen at night and in unsecured areas to reduce the risk of being detected by railway security guards. There is no chance to monitor

all wagons and infrastructure during transport so railway security guards do it randomly. The presented method based on ANN predicted the probability for 94.7% of the cases of theft and the Machine Learning identified 100% of the cases. There is no similar method or system, so the presented approach and research are innovative.

The method also predicted theft in cases where they do not occur in reality. That area is potentially risky and the railway infrastructure should be monitored there. That will reduce the number of theft cases. The method presented here can be used to develop a system for identifying high-risk places, the transport and the time of potential thefts. This type of system is needed by rail cargo forwarders and railway carriers, the infrastructure managers and transport planners. The implementation of such a system will help schedule infrastructure patrols.

In future research, the presented method will be tested in a railway carrier company to improve the prediction method. And after confirmation of the high quality of the developed model, work will be started to raise funds for the development of the system based on the presented method.

## List of Abbreviations

$Z(x; n)$  - rounds the number  $x$  to  $n$  decimal digits

$\text{Min}(a_1; a_2; \dots a_n)$  - selects the smallest number from the string of numbers  $a_1; a_2; \dots a_n$

$\text{Max}(a_1; a_2; \dots a_n)$  - selects the largest number from the sequence of numbers  $a_1; a_2; \dots a_n$

$R()$  - argument less function, gives a real random number with a uniform (rectangular) distribution from the interval  $<0; 1)$ . The random number generator of the spreadsheet is used to receive it by MS Excel in regards to the algorithm Mersenne Twister (MT19937)

$Rz(x_1; x_2)$  - function giving a random integer from the range  $<x_1; x_2>$

$F^{-1}(p; \mu; \sigma)$  - inverse distribution of normal distribution, where:  $p$  - probability,  $\mu$  - expected value,  $\sigma$  - standard deviation

## 6 REFERENCES

- [1] Lorenc, A., Michnej, M., & Szkoda, M. (2016). Information system aiding the logistics processes of loading and securing in railway transport. *International Journal of Shipping and Transport Logistics*, 8(5). <https://doi.org/10.1504/IJSTL.2016.078680>
- [2] Straka, M., Khouri, L., & Feliks, J. (2018). Design of large-scale logistics systems using computer simulation hierarchic structure. *International Journal of Simulation Modelling*, 17(1), 105-118. [https://doi.org/10.2507/IJSIMM17\(1\)422](https://doi.org/10.2507/IJSIMM17(1)422)
- [3] Yang, F., Yang, M., Xia, Q., & Liang, N. A. (2017). Cooperation between two logistics service providers with different distribution ranges. *International Journal of Shipping and Transport Logistics*, 9(2), 86. <https://doi.org/10.1504/IJSTL.2017.082524>
- [4] Teixeira Peixoto, J. P., de Queiroz, A. C. de F., Silveira, de C., Vannucchi, F., & Massimo, V. (2018). A Brief Review on Palletized Products Transportation Technology. *FME Transactions*, 46, 101. <https://doi.org/10.5937/fmet1801101T>
- [5] Straty PKP (2012). Retrieved from: <https://www.wprost.pl/kraj/27386/Straty-PKP.html>.
- [6] Ashby, M. P. J. & Bowers, K. J. (2015). Concentrations of railway metal theft and the locations of scrap-metal dealers.

- Applied Geography*, 63, 283-291.  
<https://doi.org/10.1016/j.apgeog.2015.07.005>
- [7] Hartong, M., Goel, R., & Wijesekera, (2008). Security and the US rail infrastructure. *International Journal of Critical Infrastructure Protection*, 1, 15-28.  
<https://doi.org/10.1016/j.ijcip.2008.08.006>
- [8] Burges, D. (2013). *Cargo Theft, Loss Prevention, and Supply Chain Security*. Butterworth-Heinemann.  
<https://doi.org/10.1016/B978-0-12-416007-1.00013-3>
- [9] Fischer, R. J., Halibožek, E. P. & Walters, D. C. (2019). *Introduction to Security (Tenth Edition)*. Butterworth-Heinemann.  
<https://doi.org/10.1016/B978-0-12-805310-2.00027-5>
- [10] Tyska, L. A. (2012). *Handbook of Loss Prevention and Crime Prevention (Fifth Edition)*. Butterworth-Heinemann.
- [11] Lorenc A. & Kuźnar, M. (2018). An Intelligent System to Predict Risk and Costs of Cargo Thefts in Road Transport.
- [12] Sremac, S., Tanackov, I., Kopic, M., & Radović, D. (2018). ANFIS model for determining the economic order quantity. *Decision Making: Applications in Management and Engineering*, 1(2), 81-92.  
<https://doi.org/10.31181/dmame1802079s>
- [13] Ćirović, G., Pamučar, D., & Božanić, D. (2014). Green logistic vehicle routing problem: Routing light delivery vehicles in urban areas using a neuro-fuzzy model. *Expert Systems with Applications*, 9, 4245-4258.  
<https://doi.org/10.1016/j.eswa.2014.01.005>
- [14] Pamučar, D., Vasin, L., Atanasković, P., & Miličić, M. (2016). Planning the City Logistics Terminal Location by Applying the Green p-Median Model and Type-2 Neurofuzzy Network. *Computational Intelligence and Neuroscience*, 2016, 1-15.  
<https://doi.org/10.1155/2016/6972818>
- [15] Pamucar, D. & Ćirović, G. (2018). Vehicle route selection with an adaptive neuro fuzzy inference system in uncertainty conditions. *Decision Making: Applications in Management and Engineering*, 1(1), 13-37.  
<https://doi.org/10.31181/dmame180113p>
- [16] Ćirović, G. & Pamučar, D. (2013). Decision support model for prioritizing railway level crossings for safety improvements: Application of the adaptive neuro-fuzzy system. *Expert Systems with Applications*, 40(6), 2208-2223.  
<https://doi.org/10.1016/j.eswa.2012.10.041>
- [17] Stojčić, M., Stjepanović, A., & Stjepanović, Đ. (2019). ANFIS model for the prediction of generated electricity of photovoltaic modules. *Decision Making: Applications in Management and Engineering*, 2(1), 35-48.  
<https://doi.org/10.31181/dmame1901035s>
- [18] Pamučar, D., Ljubojević, S., Kostadinović, D., & Đorović, B. (2016). Cost and risk aggregation in multi-objective route planning for hazardous materials transportation-A neuro-fuzzy and artificial bee colony approach. *Expert Systems with Applications*, 65(C), 1-15.  
<https://doi.org/10.1016/j.eswa.2016.08.024>
- [19] Czmiel, M. (2016). PKP walczy ze złodziejami za pomocą dronów. Kradzieże spadły o połowę. Retrieved from: <http://wroclaw.wyborcza.pl/wroclaw/1,35771,19501369,pkp-walczy-ze-zlodziejami-za-pomoca-dronow-kradzieze-spadly.html>.
- [20] UTK, Udział przewoźników wg przewiezionej masy towarów. Retrieved from: <https://www.utk.gov.pl/pl/raporty-i-analazy/analazy-i-monitoring/statystyka-przewozow-to/14488,Dane-eksploatacyjne-w-2018-r.html>.
- [21] NIK (2018). Bezpieczeństwo przewozów kolejowych.
- [22] PKP CARGO S. A. (2014). PKP Cargo wykorzystuje drony przeciw kradzieżom ładunków. 2014. Retrieved from: <https://kurierkolejowy.eu/aktualnosci/22108/pkp-cargo-wykorzystuje-drony-przeciw-kradziezom-ladunkow.html>.
- [23] UTK (2014). Przewozy towarowe w 2014 r.
- [24] UTK (2018). Udział pociągów opóźnionych. Retrieved from: <https://www.utk.gov.pl/pl/raporty-i-analazy/analazy-i-monitoring/statystyka-przewozow-to/14379,Punktalnosc-przewozow-towarowych-w-2018-r.html>.
- [25] UTK (2018). Rok 2017 w przewozach pasażerskich i towarowych. Podsumowanie Prezesa UTK.

**Contact information:**

**Augustyn LORENC**, PhD  
 (Corresponding author)  
 Cracow University of Technology,  
 Institute of Rail Vehicles,  
 Al. Jana Pawła II 37, 31-86 Cracow, Poland  
 E-mail: [alorenc@pk.edu.pl](mailto:alorenc@pk.edu.pl)

**Małgorzata KUŹNAR**, MSc  
 Cracow University of Technology,  
 Institute of Rail Vehicles,  
 Al. Jana Pawła II 37, 31-86 Cracow, Poland  
 E-mail: [malgorzata.kuznar@mech.pk.edu.pl](mailto:malgorzata.kuznar@mech.pk.edu.pl)

**Tone LERHER**, PhD, Prof.  
 University of Maribor,  
 Faculty of Logistics,  
 Mariborska cesta 7, 3000 Celje, Slovenia  
 E-mail: [tone.lerher@um.si](mailto:tone.lerher@um.si)

**Maciej SZKODA**, PhD, Prof.  
 Cracow University of Technology,  
 Institute of Rail Vehicles,  
 Al. Jana Pawła II 37, 31-86 Cracow, Poland  
 E-mail: [maciej.szroda@mech.pk.edu.pl](mailto:maciej.szroda@mech.pk.edu.pl)