Big Data Analytics and Anomaly Prediction in the Cold Chain to Supply Chain Resilience

Augustyn Lorenc

Assistant Professor
Cracow University of Technology
Poland

Michał Czuba

PhD student
Cracow University of Technology
Faculty of Civil Engineering
Poland

Jakub Szarata

R&D Director
SKK S.A
Poland

The purpose of the research was to develop a prediction method to prevent disruption related to temperature anomaly in the cold chain supply. The analysed data covers the period of the entire working cycle of the thermal container. In the research, automatic Big Data analysis and mathematical modelling were used to identify the disruption. Artificial Neural Network (ANN) was used to predict possible temperature-related disruption in transport. The provided research proves that it is possible to prevent over 82% of disruptions in the cold chain. The ANN enables analyses of the temperature curve and prediction of the disruption before it occurs. The research is limited to coolbox transportation of food under -20^oC, but the method could also be used for Full Transport Load (FTL) in refrigerated transport. The research is based on real data, and the developed method helps to reduce the waste in the cold chain, improve transport quality and supply chain resilience. The presented method enables not only to avoid cold chain breaks but also to reduce product damage as well as improve the transport process. It could be used by cargo forwarders, Third-Party Logistics (3PL) companies to reduce costs and waste. The literature review confirms that there is no similar method to prevent disruption in the transport chain. The use of the Internet of Things (IoT) sensors for collecting data connected with Big Data analysis and ANN enables chain resilience provision.

Keywords: Disruption in the cold chain; predict and prevent disruption; provide chain resilience; IoT for food transport monitoring; ANN prediction model; Big Data analytics;temperature anomaly

1. INTRODUCTION

Refrigerated transport of food products (Cold Chain) is one of the most important types of transport considering that it is primarily meant to transport food products necessary for life and proper functioning of every human being. Representatives of the Transport-Forwarding-Logistics (TFL) sector consider the impossibility of analysing the risk of damage to goods and the insufficient possibility of monitoring their conditions environment during distribution as one of the key problems preventing full optimisation of the supply chain. Each year the food industry, i.e. production companies and stores, bear the costs resulting from the deterioration of products before their expiry date. Very often these are not only financial losses but also image-related losses because large networks ordering such products have very strict quality standards. On the contrary, the manufacturer implementing the orders of a major customer has no control over the entire logistics chain, as part of the transport may be carried out by the ordering party or external companies –Third-Party Logistics (3PL) [1].

Both temperature monitoring of goods during the transport as well as management of information about changes in the ambient conditions inside the refrigerated vehicle are important parts of the cold chain.

Moreover, changes in the conditions that may occur during transport and irregularities in the structure of the body may cause significant reduction in the quality of transported products and the associated risks to consumers. This threat results in the cases of shortening the expiration date of products and their deterioration before the date guaranteed by the producer. Additionally, and the loss of key properties such as consistency, colour and taste appears resulting in the loss of brand image. It is assumed that about half of transported food products require transport at a controlled temperature, and food losses caused by abnormal transport conditions reach over 30%, which in turn generates unnecessary costs for the enterprise [2].

From the perspective of a person managing a cold supply chain, the most important thing is to deliver the goods within the set time frame and at the lowest possible cost and maintaining at the same time full quality of the goods. To be able to manage efficiently, they need reliable data that can form the basis for taking corrective actions - continuous improvement process [3].

In this article the authors focus mainly on transporting cold supply chain products using thermal insulation containers - coolboxes. This is one of the most popular ways of delivering products, especially for differentiated goods or multi orders.

Control of the circulation process of thermally insulated coolboxes is most often carried out in two areas: quality and implementation of process indicators (e.g. preparation of coolboxes, maximum delivery time, the maximum time of coolbox returning to the warehouse, etc.). Process indicators are set individually by the entity responsible for logistics. Measuring process indicators is possible based on data available in process registration systems such as Warehouse Management System (WMS), Transportation Management System (TMS), Proof of Delivery (PoD), etc. By combining data from different systems, the logistic cycle can be traced and the maintenance of the assumed indicators checked.

In the case of the quality control process the requirements must be met according to ATP convention. When transporting food products in thermally insulated coolboxes, the most common way is to check samples of transport orders using temperature recorders. A temperature recorder is a small, autonomous device capable of measuring the ambient temperature at specific measuring intervals. The measurement result is saved in the device memory. At the beginning of the process, the temperature recorders are placed in randomly selected coolboxes assigned to transport orders. During the process, the device registers the temperature inside the coolbox. After the process is completed, the device is removed from the coolbox to read for the saved measurements to be read. Having the remaining data from the process (the assumed and the actual time of delivery), the fulfilment of the quality indicator for the transported goods is assessed [4].

In the case of data download methods, USB interface is the most common, which allows direct connection of the device to a computer. It is also the cheapest option. Devices with wireless communication interfaces are also chosen increasingly. Downloading data is possible using a telephone or a special data collector. There are also devices with a built-in interface that uses a GSM network to transmit data online. However, these devices are more expensive and usually larger.

There are no regulations or recommendations regarding the amount of control. The number of control samples is as high as possible. However, due to the high labour intensity of the process and mass quantities of transport orders, it is usually well below 1%.

The authors' contacts with companies have helped to identify the real problems with transportation quality in the cold chain. Every product must be transported and stored in the defined temperature and humidity conditions to ensure the good quality of products in the final destination (markets). The law requires just the temperature conditions for general groups of products, the humidity measure is not required by the law. The 3PL companies comply with legal requirements but also use their own good practice rules. For example, for specific products, the temperature is lower than required by the law, and the measurement interval is reduced to ten or fifteen minutes.

A trend in cold chain transportation is to use coolboxes instead of FTL (Full Transport Load). This new transport solution enables companies to use standard road transport which is cheaper than that refrigerated transport. What is more, in the case of using coolboxes the increasing temperature of products during too long awaiting for unloading products in markets of small shops is not a problem. Coolboxes enable the low temperature, so the products are still in proper condition.

Companies identify a few problems resulting from using coolboxes in the cold chain. The main problems mentioned by transport managers and dispatchers include:

− leaky coolboxes, during service the seal between the coolbox and the door becomes damaged,

− too high temperature of coolboxes while products are being loaded coolboxes should be cooler before products and cooling cartridge are packed, but sometimes the coolbox is not cooled adequately. In these cases, during transport, the temperature inside the coolbox rises rapidly which may lead to product defrosting on long distances.

− too low refrigerant in the coolbox, which causes not reaching the required temperature inside the coolbox, so the transport condition is not met.

− too much refrigerant in the coolbox, which does not cause deterioration of products' quality temperature inside the coolbox is lower than required, but the company bears the higher and unnecessary cost of the higher quantity of refrigerating medium.

− unclosed coolbox, which is infrequent but the effects are serious − all products in the coolbox could be defrosted.

The high temperature of the products put into the coolbox, sometimes the temperature of the product's surface is proper but the temperature inside the product is still high. This problem results from the improper storing temperature in the warehouse. The high temperature of products inside the coolbox accelerates the consumption of refrigerant, and the required temperature cannot be reached.

These problems directly disrupt the cold supply chain and, as a result, create a risk of damage to the goods. To help solve the real problems in the cold chain companies, the authors developed a method to identify the high probability of risk for the above-mentioned problems. The research aimed to analyse the temperature curve and refer to the gradient or change rate to the causes. Next, a mathematical model was developed to enable automatic identification of all cases. However, sometimes the transport curve is hard to relate to the cause, so to solve this problem an ANN tool was used. Finally, the method helps to identify risky cases before they occur. Owing to this method the expected risky cases can be avoided. As a result, managers will get precise guidance on the process areas in which adjustments should be made.

The paper was limited to the temperature, disregarding humidity because, in the company that provided the data the transported products are limited to frozen ones, humidity is not measured. For fresh products research could be extended to the humidity analysis and the correlation of humidity with temperature.

2. LITERATURE REVIEW

In this paper the term "Big Data" can be defined as a set of a large amount of data, most often derived from

many sources. The "BigData" term is used for the data set characterised by volume, velocity, variety and value [5,6]. The concept of a big data set is essential and denotes a situation when the set cannot be used within trivial methods. The data used for this research are structured but they are hard to analyse. It is possible to visualise the temperature inside each coolbox, but finding disruptions in the data is not simple. Connecting the data from sensors with TMS, CSV files results in data set variety.

The current usage of this term is associated with the use of predictive analytics and advanced analytical methods, e.g. fuzzy logic or artificial neural networks.

Temperature control in the cold supply chain is a difficult and complex issue in logistics. The cold chain is used in many areas. It not only affects the industry and financial losses of companies but also people's health, e.g. medicaments or food transport. The state-of-the-art technology allows research of processes on a larger scale and faster analysis of a big amount of data. This is due to the rapid development of IoT (Internet of Things) solutions with a complex system and growth of artificial intelligence algorithms.

To analyse the cold chain in terms of temperature overruns, data from the WMS system and a centralised wireless temperature monitoring system can be used. The issue of temperature prediction and analysis was raised by many researchers.

Temperature prediction is important in many fields of science. For example, Corchado and Fyfe present an unsupervised neural method for temperature forecasting of oceanographic water masses. The main purpose of the work is a comparison of the obtained results (Negative Feedback Artificial Neural Network) with a Linear Regression and an ARIMA model [7].

Temperature forecasting is important for applications in industry and the environment. Abdel-Aal proposes alternative abductive networks associated with modern machine learning techniques. The abductive network models were developed covering forecasting both extreme and normal temperatures. The paper presents next-day and next-hour models. The performance of the presented method is more favourable than the neural network model based on the same data at that time [8].

A lot of methods have been applied to predict temperature data and weather. Li-Wer Lee, Li-Hui Wang, Shyi-Ming Chen present a method of temperature prediction based on fuzzy logical relationships and genetic algorithms. The method is based on two-factor fuzzy logical relationships built on historical data. Besides, the authors use the same method to TAIFEX forecasting. Their method has a higher forecasting accuracy rate than the existing methods. The same authors proposed a new modified method using genetic simulated annealing techniques, was intended to deal with the mutation operations of the generic algorithm [9,10].

Temperature forecasting methods can also help reduce damage to plants and animals caused by extreme temperature. Smith, Hoogenboom and McClendon presented a project that explores the application of artificial neural networks for automated year-round temperature prediction. The paper proposes a prediction model at the time ranging from 1 – 12 h. Ward-like artificial neural networks were used for this research. This model was implemented on the Georgia Automated Environmental Monitoring Network. The described methods were more accurate than those developed heretofore. In 2010 Hoogenboon, McClendon and R. Chevalier presented a comparison of methods of air temperature prediction. The methods tested were Support vector with reduced training sets and artificial neural networks. The results were comparable with the previous models based on ANN. The SVR models were more accurate in winter-only and year-round evaluation data for all twelve of the prediction horizons [11, 12].

Hu lsmann and Brenner performed an analysis of the cold supply chain in their book "Causes and effects of cold chain ruptures: performance of fragment versus integrated cold chains". They found over 100 reports related to temperature disorders, including wrong temperature management during storage and land transport. In water transport, the cooling period of coolboxes until reaching the set temperature can by up to two weeks. However, in air transport temperature fluctuations are a big problem [13].

The quality problem of fruits transport can be associated with improper conditions of transport and storage. R. Jedermann, U. Praeger, M. Geyer and W. Lang studied the subject of remote quality monitoring in the banana chain. One of the tested environmental factors was the temperature. The measurements were taken by a set of twenty wireless sensors and additional data loggers (iButton). The authors presented models for green life and temperature prediction inside a loaded coolbox. They also suggested a method of cooling improvement by proper packing [14].

In 2014 R. Jadermann, M. Nicometo, I. Uysal and W. Lang in their paper "Reducing food losses by intelligent food logistics" highlighted an important issue related to the food transport and storage. The remaining shelf life of food may be predicted by biological models calculating the influence of temperature and other factors. Their combined research indicates that the correct temperature management is a basic requirement in food logistics but, practically, is often overlooked. In their opinion, the shelf life of food depends on not exceeding temperatures during transport and storage [15].

Medicine is the next field of logistics where temperature measurement has an important role. As early as 1997, A. Wawryk, C. Mavromatis and M. Goldmeasured the temperature of long-stored vaccine using electronic data loggers. The study aimed to determine how long vaccines were stored at the temperature < 0 C. The vaccines in some cases were exposed to impermissible temperatures. In the commentary, the authors recommended using a temperature monitoring system and cold chain guidelines [16].

S. Monteleone, M. Sampaio, R. F. Maia presented an application of a smart temperature monitoring system in a medicine cold chain based on 2G-RFID-Sys. The authors pay attention to the importance of the sensitivity of medical products to the wrong temperature during the logistics cycle. The effects of exceeding the temperature conditions of storage and transport have an impact not only on financial losses but also directly on people's life. They proposed a conceptual model for temperature

monitoring in medicine cold chain. They also mentioned the technological progress of Wireless Sensor Network that cooperates with RFID (Radio-Frequency Identification) systems and provide better monitoring of parameters (e.g. location, temperature) [17].

Neural networks have been used in studies on the cold chain supply for some time now. The aim of the research done by W. Xu, Z. Zhang, D. Gong and X. Guan was to improve risk management by using prediction based on a neural network model. In the first step they presented Back Propagation neural network, but it did not produce the desired result. To improve the method, the authors increased the momentum term and adjusted the learning rate, and also modified the error and transformation function [18], [19].

Based on the above literature review, it is difficult to indicate one method that predicts the risk of temperature overrun to prevent it. Many systems allow data collection and monitoring. However, they are not used for analyses to improve the quality of services. A method for estimating environmental parameters should be implemented in control and registration software [20]. Another article identifies the current technologies trends in the cold chain, especially for the monitoring and control of the temperature through the production, transport and storage. The authors discuss the benefits of new technologies, such as significant time cost savings and reduction of the waste of food [21]. The effectiveness of the Distribution Center (DC) is strongly connected with the Transport between them, and the quality of the transport services impact the disruption of the cold chain [22]. European Union takes care of the energy consumption and waste, because of this it is the main goal of DC's to predict the disruption and reduce the energy and resources need in logistics processes [23].

3. RESEARCH METHODOLOGY

The research aimed to predict disruption in the cold chain and develop a method that could help avoid the disruption. In the first step of data analysis, authors found that it is hard to define one mathematical model which would enable the prediction of temperature changes in next fifteen minutes and help identify risky case occurred in this period. This problem is due to fuzzy boundaries between risky states. To solve the fuzzy-logic problem, and based on the previous research, the authors used ANN tools to achieve high efficiency of the developed method – high compliance of anticipated cases with reality. ANN is not treated as the method for this research, it is a part of a complex method that involves mathematical models, automatic analysis and ANN.

The scientific thesis for this research was formulated as:
- ST1: Automatic identification of disruption is possible based on the temperature curve analysis.
- ST2: The use of ANN tools enables them to achieve better results than the use of mathematical models.

In the study, a simulation was performed and its results were compared with real-life cases. The input data come from real transportation cases.

The research-based on real data collected from 105 coolboxes used for transportation of food between the distribution centre (DC) and the local depot. For data collection, high accuracy measurement hardware was used. For data preprocessing and first stage analysis, the Tableau software was used.

3.1 The process of callbox preparation

The logistics service process with the use of thermally insulated coolboxes (coolbox) is carried out in a 3-stage cycle. A typical logistics cycle with a thermally insulated coolbox is shown in figure 1.
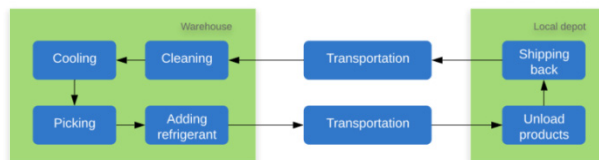

Figure 1. Collecting the data with readings from a sensor

Preparation - The coolbox preparation process usually consists of 4 steps:

Coolbox washing -after returning the coolbox to the warehouse it should be subjected to a washing and disinfection process (if required). In some cases, this step is skipped - procedures may involve periodic washing of the coolboxes once every several logistics cycles. Washing takes place using pressure and / or steam washers at the temperature above 800C. Failure to follow washing procedures may result in sanitary problems and thus damage to goods.

Cooling - the coolbox should be cooled to around + 60C before completing the assembly, especially if the temperature in which it was previously was much higher. If you skip this step or cool it down too short, the time for maintaining the temperature inside the coolbox can be significantly reduced.

Picking - the process of picking goods directly into the coolbox by the order to issue a GM.

Addition of a cooler. After completion of the assembly, a refrigerant such as a eutectic cartridge, compressed CO_2 or dry ice is added to the coolbox. The dosage of refrigerant has a key impact on the time of maintaining the correct temperature inside the coolbox.

Transport - coolbox delivery from the warehouse to the destination. In the simplest case, the coolbox is delivered directly from the warehouse to the destination (e.g. a local depot or store). Transport is frequently extensive and consists of several nodes. The coolbox must be delivered from the warehouse to the destination according to the assumed maximum delivery time. Exceeding this time may cause the temperature inside the coolbox to exceed the maximum storage temperature of the goods.

Coolbox returns - delivery of the empty coolbox back to the warehouse. After delivery of the goods to the destination, the thermally insulated coolboxes are returned to the warehouse. A coolbox can be returned by the same or subsequent transport. To ensure smooth circulation, coolboxes should be returned to storage as soon as possible.

3.2 Data collection

There are many different thermal recorders on the market that can be used for process measurement. When choosing a device, special attention should be paid to:

- acceptable temperature and humidity range (e.g. -30°C to +60°C)
- temperature measurement range (e.g. -20°C to +50°C)
- measurement accuracy (especially in the temperature range inside the coolbox, e.g. from -20 to 0 +/- 0.2^{o}C)
- battery life (from 15 days up to several years)
- IP housing leakproofness (e.g. IP67)
- method of downloading data from the recorder.

A system dedicated to this type of applications from SKK was chosen to collect research data. For the research, the data for analysis were exported from the measuringsystem collecting data from the wireless temperature sensor. The wireless temperature sensor is a small measuring device based on the Bluetooth Low Energy technology. The device has non-volatile memory and sensors, i.e. Swiss temperature module.

The sensor is the basic technical elements of the measuring system that is responsible for the collection and pre-analysis of measurement data. The range of stable operation of the device is 15 m at standard transmitter settings.

The main parameters of the sensor are described in table 1.

Table 1. The main parameters of the used sensor

Parameter	Value
Memory	100 – 20 000 records of one sensor data
TX Power	-12 do 2 [dBm]
Measurement interval	2 [s] – 1 [h]
Battery life	Up to 2 years
Working temperature	-25 ÷ 65 [°C]
Range	-20 ÷ 60 [°C]
Resolution	0.3 [°C]
Accuracy	0.2 [°C]

The sensor works in the Wireless Data Collector that automatically downloads data from nearby devices. In the next step, the sensors' measurements are sent to the central system with the database. This system enables visualization and data export.

The data collection process is carried out according to the attached architecture – figure 2.

Figure 2. Collecting the data with readings from the sensor

Thermally insulated coolboxes were equipped with sensors. The sensors were mounted in the middle of the back wall of the coolbox. In the upper part of the coolbox, a refrigerant is most often applied and too close installation of sensors can affect the accuracy of the measurement.

Wireless data collectors were installed in the warehouse for automatically downloading data from the sensors. Next, the data from the wireless collectors were transmitted to the database.

The sensors measure and store data throughout the study - that is, at each of the three stages of the logistics cycle. The data are downloaded from the sensor's memory each time the logistic cycle is completed and the coolbox returns to the warehouse.

The temperature measurement interval in the research was 15 minutes.

3.3 The data used in research and understanding the disruptions

For data analysiscovered the period of 11th May 2019 to 5th July 2019. In the analysed period 105 coolboxes were used in transport. The typical number of measurement for each sensor was about 2891 readings so the research was based on the 289441 temperature readings. A sample of exported data from the measuring system is presented in table 2. The PartitionKey means the MAC address of the measured sensor. This MAC address is connected with the coolbox number. The RowKey means the time of the measure in the DateTime format, and Value is the temperature.

Table 2. The structure of row data from the measuring system

PartitionKey	RowKey	Value
00:81:f9:83:43:1a	2019-06-20T00:00:00	21.3
00:81:f9:83:43:1a	2019-06-20T00:15:00	21.3
00:81:f9:83:43:1a	2019-06-20T00:30:00	21.3
00:81:f9:83:43:1a	2019-06-20T00:45:00	21.3
00:81:f9:83:43:1a	2019-06-20T01:00:00	20.9
00:81:f9:83:43:1a	2019-06-20T01:15:00	20.9

The sensors used for collecting the data are active 24 hours each day. So the data include readings from coolbox preparation, input products into them, transport, products unloading and returning empty coolboxes. The data from the sensors are combined with the data from TMS (Transport Management System) and with a list of sensors MAC addresses with coolbox name. The correlation of reading data with other data sources is presented in figure 3.
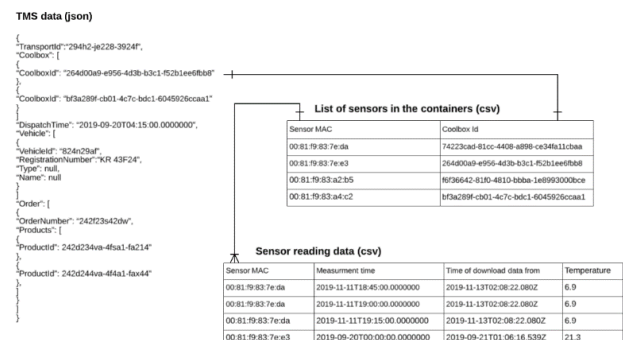

Figure 3. Collecting the data

The transportation temperature is required to be below -15oC. The products in the coolbox are trans–

ported over short and medium distances so the typical distribution time is between 10 and 24 hours. Depen–ding on the predicted distribution time an adequate refrigerant is selected – the volume of dry ice put into coolbox. One of the main problems during this process is matching the amount of refrigerant to the mass of goods and the length of the route as well as the damage of the coolbox itself i.e. coolbox leak, seal wear. An example of achieving the required temperature below -15oC is presented in figure 4 as a box-and-whisker plot.
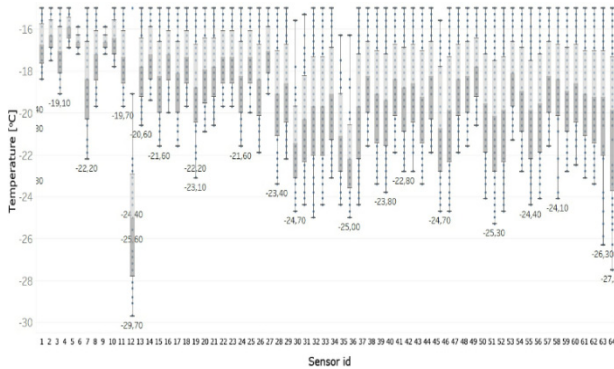
Figure 4. The coolbox and whisker plot for the temperature above -15oC for selected sensors (proper conditions in the cold chain)

As shown in figure 4, products in some coolboxes spend more time under -5oC than in other boxes. On the other hand, in some coolboxes, the temperature is near that required. This can be accounted for by:
 – very rare use of the toolbox for this kind of
    transportation,
 – damaged/ untight coolbox,
 – unclosed coolbox,
 – a high temperature of the products in the coolbox,
 – lack of/too low volume of refrigerant.

This simple data analysis indicates whether there is a need to develop a way to avoid these problems. The solution presented below is dedicated to disruptions prediction in transport, which helps avoid breaks in the cold chain and improve the quality of services provided by career carriers and cargo forwarders.

3.4 Analysis of temperature curve

At the beginning of the research, the collected readings from the sensor were manually analysed. A sample of readings is presented in the chart in figure 5.
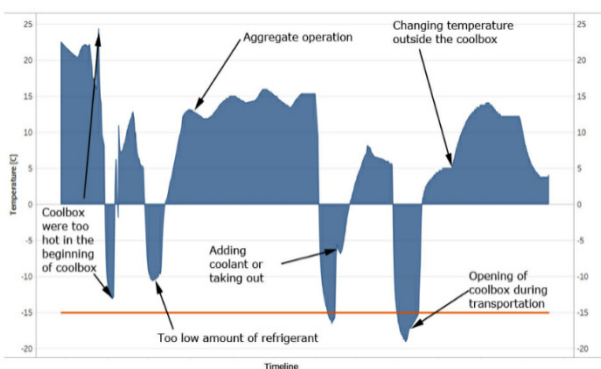
Figure 5. Sample of analysed data with identification of causes of temperature change

3.5 Automatic analysis of current states/disruptions

The goal of the research was to predict the next state, which will be noticed in the 15 minutes interval. To achieve it eight states were identified, half of which are disruptions. These disruption states are based on real problems of companies, which are presented in the intro-duction of the paper. The problems companies have to deal with concern mainly a bad condition of the coolbox (worn or damaged), identification of properly cooled coolboxes (bad management in coolbox preparation process), the opening of coolbox during transportation without reason, incorrect quantity of refrigerant (too high –unnecessary costs, too low – the temperature of the product is higher than required). Based on the analysis of these problems disruption states were tabulated (Table 3).

**Table 3. The disruption states**

No.	Problems reported by companies	Cases in the analyzed period	Disruption	States
1	The bad condition of coolbox (worn or damaged). It is hard to identify coolboxes which need servicing. In these coolboxes, proper temperature is not maintained.	25634	improper cooling	cooling Slow
2	A too low quantity of refrigerant.			
3	Problem with identification of properly cooled coolboxes. Coolbox was too hot at the beginning of the coolbox cooling process (preparation) or time of cooling was too short.	4740	temperature is higher than required	Cooling FromHigh
4	Opening of coolbox during transportation without reason.	25802	unexpected temperature normalizatio n or heating during the cooling process	Cooling Anomaly
5	Reaching the maximum possible temperature with refrigerant but this temperature is higher than required in transport	6441	temperature is higher than required	minimum Cool But ToHigh

For automatic data analysis – identification of the state of the current reading analytics formulas were devised. For each state the following state number was assigned:
 – state number = 1 means tempState = constant,
 – state number = 2 means tempState = cooling,
 – state number = 3 means tempState = coolingSlow,
 – state number = 4 means tempState =
    coolingFromHigh,
 – state number = 5 means tempState =
    coolingAnomaly,
 – state number = 6 means tempState =
    minimumCoolReached,
 – state number = 7 means tempState =
    minimumCoolButTooHigh,
 – state number = 8 means tempState = heating.

The bolded states 3, 4, 5, 7 denote disruption and state 1, 2, 6 are normally used as a support for automatic data analysis. State 8 can be found in both of them (disruption or normal state) - what the state depends on: whether the temperature rose before the end of transport (disruption) or heating occurred after unloading of products (normal state).

The key disruptions were identified and the temperature curve was referred to the disruptions. The sample temperature curve with disruption is presented in figure 6.
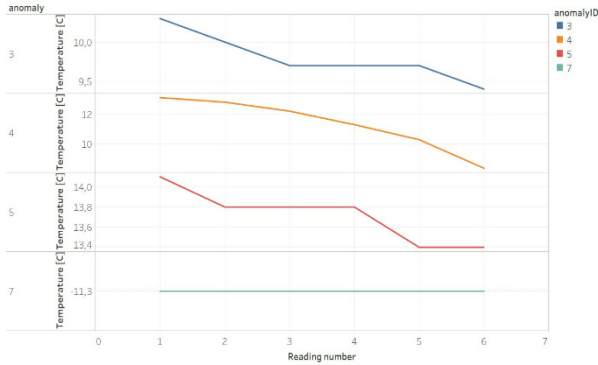
Figure 6. Sample of temperature curve typical of disruption

Based on the temperature curve and the problems companies face as mentioned in the introduction, mathematical models for automatic identification were developed. On the basis of managers' experience and the analysis of the temperature curve for the analysis of the current state the following assumptions were made:
  – temperature gradient during cooling (Tc): 0.3°C,
  – temperature gradient during heating (Th): 0.3 °C,
  – proper temperature of the coolbox before loading products (Tb): 8 °C,
  – the permissible amplitude of temperature (Ta): 0.4 °C,
  – proper cold chain temperature in transport (Treq): -15 °C,
  – the temperature of very slow cooling for an insufficient quantity of dry ice (Ts): 0.05 °C.

The states (numbers 1÷8) are determined by meeting the conditions (1÷8).

The constant temperature (state: 1- constant) is determined by the formula (1):

$$\left(\left|\left(T_{i-1} - T_i\right)\right| \leq T_a \wedge \left|T_{i-2} - T_i\right| \leq T_a \wedge \left|T_{i-3} - T_i\right| \leq T_a\right) \Rightarrow tempState = 1 \tag{1}$$

The temperature drop (state: 2 - cooling) inside the coolbox is determined by (2):

$$\left(T_{i-1} - T_i \geq T_c \wedge T_{i-2} - T_i \geq 2 * T_c\right) \Rightarrow tempState = 2 \tag{2}$$

The too-slow cooling - a slight decrease in temperature gradient (state: 3- coolingSlow) is determined by (3):

$$\left(T_{i-1} - T_i \geq T_s \wedge T_{i-2} - T_i \geq 2 * T_s\right) \Rightarrow tempState = 3 \tag{3}$$

The identification of a coolbox not ready for transportation – too high temperature of the coolbox while starting loading products into it (state: 4-coolingFromHigh) is determined by the formula (4):

$$\left(T_{i-1} - T_i \geq T_c \wedge T_{i-2} - T_i \geq 2 * T_c \wedge T_{i-2} > T_b\right) \Rightarrow tempState = 4 \tag{4}$$

The identification of unusual temperature changes that could be caused by an unexpected opening of the coolbox (state: 5- coolingAnomaly) is determined by (5):

$$\left(TempMax \neq 0 \wedge tempState(i) \notin \{2;3;4\}\right) \Rightarrow tempState = 5 \tag{5}$$

where:
TempMax – the maximum temperature of coolboxes in the short period before the cooling process (before the temperature decrease: tempState $\in \{2,3,4\}$).

The state of achieving the required temperature (state: 6 - minimumCoolReached) is defined by (6):

$$\left(T_i \leq T_{i-1} + T_a \wedge T_i \leq T_{i-2} + T_a + 1 \wedge T_i \leq T_{i-3} + T_a + 3 \wedge T_i \leq T_{req}\right) \Rightarrow tempState = 6 \tag{6}$$

Reaching minimum possible temperature (state: 7-minimumCoolButTooHigh) but higher than required is determined by the formula (7):

$$\left(T_i < 0 \wedge tempState(i) \neq 1 \wedge T_i \geq T_{req}\right) \Rightarrow tempState = 7 \tag{7}$$

Heating - the growing temperature inside the coolbox (state: 8 - heating) is determined by the formula (8):

$$\left(T_i \geq T_{i-1} + T_h \wedge T_i \geq T_{i-2} + 2 * T_h\right) \Rightarrow tempState = 8 \tag{8}$$

Based on the presented formulas the states were defined sample results of automatic state identification algorithms are presented in Figure 7.
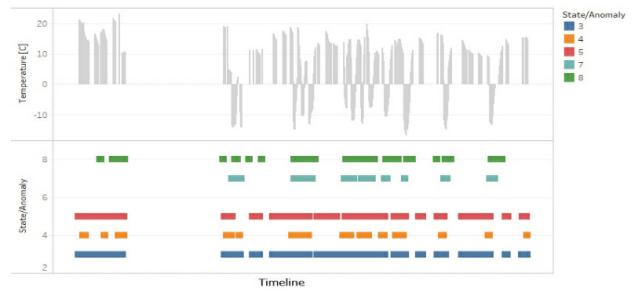
Figure 7. A sample of results of automatic data analysis

The example presents the results of automatic data analysis for the sensor 00:81:f9:a2:e7 in the filtered period from 11th May 2019 to 5th July 2019. The colour shows details about state/anomaly. The data zooming enables the validation of proper state identification.

Owing to the automatic analysis based on the presented method the density of disruptions in the analysed period can be specified. Frequent disruptions in transport are: too slow cooling (10.5% cases of the total; 35.7% cases of disruptions), incorrect cooling, i.e. too low refrigerant, damaged coolbox and cooling anomaly (12.3% cases of the total; 41.8% cases of disruptions) or unexpected temperature normalisation or heating during

the cooling process. The problem of reaching the maximum possible temperature by means of the refrigerant but the temperature exceeding that required in transport is typical of 4.1% cases of the total; 13.9% cases of disruptions. The disruption caused by using not ready coolbox (heat coolbox) not reaching the proper temperature of the process is found in 2.5% cases of the total; 8.6% cases of disruptions. The results are presented in a graphic form in figure 8.
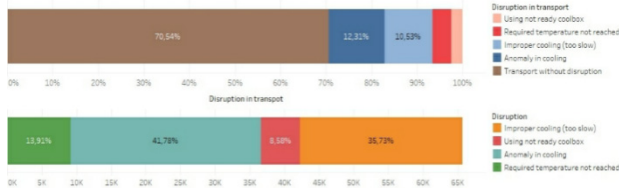

Figure 8. Disruptions in the analysed transport cases

After the identification of disruption states by the presented mathematical models (formulas 1÷8) was followed by a comparison of the results obtained with real-life causes. The effectiveness of these models is shown in table 4.

Table 4. The results obtained using a mathematical model vs. real-life cases

Disruption state	Real cases	Correct identification	Effectiveness [%]
3	25634	21348	83.3
4	4740	2001	42.2
5	25802	24540	95.1
7	6441	5570	86.5
Total	62617	53459	85.4

Based on table 4 it can be concluded that the effectiveness between /in mathematical models is about 85.4% What must be noticed is that the mathematical model makes many mistakes in the identifications of the coolboxes which are not ready for loading products into them.

Because of the low quality of the results a better method should be provided. To solve the fuzzy-logic issue with states analysis, which is the main problem, ANN tools could be used to achieve high efficiency of a developed method, i.e. high compatibility between anticipated cases and reality. ANN is part of a complex method used as a core and the mathematical models, automatic analysis and the ANN.

3.6 A states prediction model

As the input data for the ANN training, six consecutive readings and the temperature at the beginning of cooling (TempMax) were used as input data. The StateId was used as the target. The StateId means one of the states based on six reading and the maximum temperature of coolbox before these readings. A sample of the input data set is presented in table 5. For ANN training 66417 input sets were used.

A feed-forward network was used in the simulation. The selection of an artificial neural network structure was based on the method of subsequent approximations. For simulations the following set of training data sets was made:

  − 70% of the collection - training data,

  − 15% of the collection - validation data,

  − 15% of the set - test data.

In data preparation the following Matlab functions were used:

  − Mapminmax - normalises input values to the range of -1, 1 (acceleration of calculations),

  − Removeconstantrows - removal of input vectors consisting of the same values.

Table 5. The structure of input data for ANN training

Input							Target
TempMax	t1	t2	t3	t4	t5	t6	State ID
20.9	21.3	21.3	21.3	21.3	20.9	20.9	5

The network structure selection was made based on the mean square error (mse). The network structure selection was made based on the mse which was about 0,24849 at epoch 81. Figure 6 shows the best validation performance. Over fifteen different structures were examined to choose the best structure.
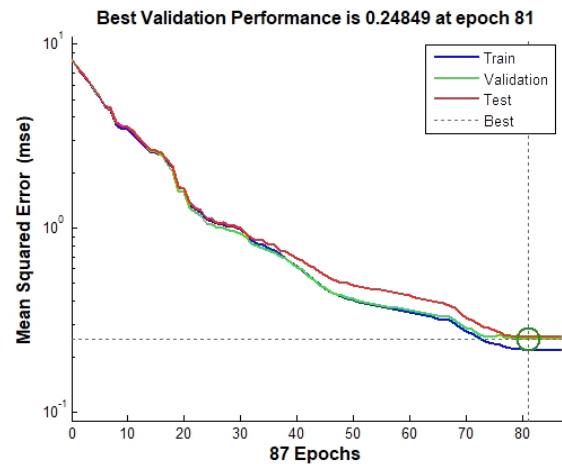

Figure 9. Structure of artificial neural network

Figure 10 shows the structure of the network. The created artificial neural network (ANN) consists of three hidden layers (containing respectively 5, 10 and 5 neurons) and one output layer. As activation, the tangential function was implemented The Levenberg-Marquardt backpropagation method was used as the learning algorithm.
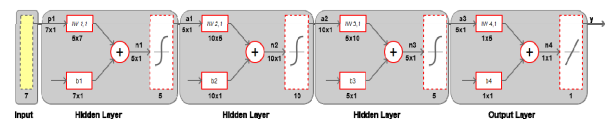

Figure 10. Structure of artificial neural network

The training parameters for the best structure were defined as:

  − The maximum number of epochs to train (net.trainParam.epochs): 200

  − Performance goal (net.trainParam.goal): 0

  − Minimum performance gradient (net.trainParam.min_grad): 1e-5

  − Maximum validation failures (net.trainParam.max_fail): 6

  − Training gain - Mu (net.trainParam.mu): 0.001

  − Mu decrease factor (net.trainParam.mu_dec): 0.1

− Mu increase factor (net.trainParam.mu_inc): 10
− Maximum mu (net.trainParam.mu_max): 1e10

The Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares, then the Hessian matrix can be approximated as:

$$H = J^T J \qquad (9)$$

and the gradient can be computed as:

$$H = J^T J \qquad (10)$$

where J is the Jacobian matrix that contains first derivatives of the network errors concerning the weights and biases and e is a vector of network errors. The Jacobian matrix can be computed through a backpropagation technique. The Levenberg-Marquardt algorithm uses this approximation to the Hessian matrix in the following update:

$$x_{k+1} = x_k - \left[J^T J + \mu I \right]^{-1} J^T e \qquad (11)$$

When the scalar μ is zero it is Newton's method, using the approximate Hessian matrix. When μ is large it becomes gradient descent with a small step size. Newton's method is faster and more accurate near an error minimum, so the aim is to shift toward Newton's method as quickly as possible. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a tentative step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm.

Backpropagation is used to calculate the Jacobian jX of the performance part concerning the weight and bias variables X. Each variable is adjusted according to the Levenberg-Marquardt algorithm:

$$\begin{aligned} ji &= jX * jX \\ je &= jX * E \qquad (12) \\ dX &= -\frac{ji + I * mu}{je} \end{aligned}$$

where E is all errors and I is the identity matrix.

The adaptive value mu is increased by mu_inc. The change is then made to the network and mu is decreased by mu_dec.

In figure 11 the regression plots for training, validation and testing of ANN are presented. The ANN returns the float value instead of the whole so the prediction of the state is similar to fuzzy logic. And these results could be processed in the second stage.

As can be seen the average error is less than 0.02. This indicates that the chosen structure of the ANN is well fitted to real data cases. The circle in figure 11 represents the results of the comparison of the target value with the predicted one. If the model reflects real cases well, the regression line is tilted 45 degrees to the axis and the results (circle markers) are located near the regression line. If the model is overlearned, the results are usually above the line, if the model is not learned

enough, the results are below. If the results are on both sides of the line at a considerable distance, despite the inclination of 45 degrees, it means that the analysed cases are characterised by high uncertainty. The regression for the presented ANN structure is determined by the formula (13):

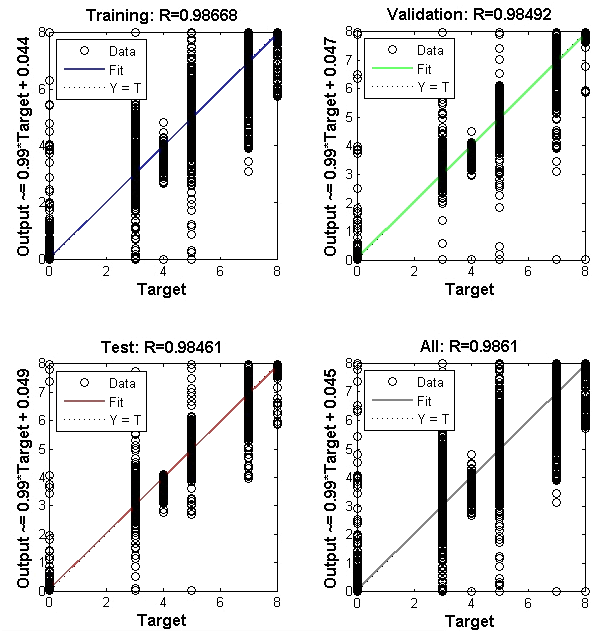$$Prediction = 0.99 * TargetStateId + 0.045 \qquad (13)$$

Figure 11. Regression plot for training, validation and testing of ANN

Temperature prediction is not an easy process because it depends on many factors. In the analysed cases these factors include heat capacity of the products in the coolbox, air humidity and technical condition of the coolbox. Heat capacity and temperature of the products as well as temperature outside the coolbox are also crucial. There is also airflow inside and outside, number of coolboxes near the analysed one and more to be taken into cosiderartion. These factors are not measured because of complexity and costliness. Consequently, it is not possible to reduce the uncertain impact of factors on temperature changes.

4. MODEL CONVERGENCE ANALYSIS

The regression plot presented in figure 11 for the selected ANN structure shows that the model gives quite good results. In the next step, the real cases were used and compared with the predicted ones. For this simulation 223019 cases were used. These cases are derived from real temperature readings during the transport in the analysed company. For these cases, the model returns 12 563 wrong predicted states so it was about 5.63% of cases. Among these states there were also ones not classified as disruptions. Disruption model returns 5 600 incorrectly specified states (about 14.6% of total).

A detailed analysis of effectiveness is presented in table 6.

The cases of slow cooling (StateID=3), anomaly during cooling (StateID=5) and reaching minimum possible temperature, still higher than required, (Stated=7)

were predicted correctly in over 90%. The cases of identifying the use of an unprepared coolbox (StateID =4) were predicted correctly in 49%.

Table 6. Results obtained using mathematical model vs. ANN method and real cases

Disruption state	Real cases	Mathematical model		ANN method	
		Correct identifica-tions	Effecti-veness [%]	Correct identifica-tions	Effective-ness [%]
3	25634	21348	83.3	23563	91.9
4	4740	2001	42.2	2363	49.9
5	25802	24540	95.1	25207	97.7
7	6441	5570	86.5	5884	91.4
Total	62617	53459	85.4	57017	91.1

The effectiveness analysis proves that using the ANN tools as a core of the presented method yields better results than classical mathematical modelling. The results obtained were better by about 5.7% so the prediction of the disruption is more effective. The presented research proves the scientific thesis:
- ST1: Automatic identification of disruption can be based on the temperature curve analysis.
- ST2: The use of ANN tools yields better results than mathematical models.

5. COST REDUCTION RESULTING FROM THE APPLICATION OF THE PRESENTED METHOD

During the research, the authors asked themselves whether it would make sense to implement such system in practice, taking into account the results of the analysis. To answer this question the Return on investment – ROI would have to be counted. ROI is defined as:

$$ROI = Gp\text{-}Ti \qquad \backslash \qquad (14)$$

where: Gp is Gross profit = savings due to the lack of loss of goods

It was denoted the total investment costs over a given period. In our case, we assumed 3 years for ROI.

To calculate the investment costs, it was assumed that the system would work in a medium-sized company with the following parameters:
- number of icebox containers:

$$Ib = 1000 \qquad (15)$$

- average container rotation time:

$$Ro = 5 \text{ days} \qquad (16)$$

- annual number of transport orders carried out using the icebox:

$$No = 25{,}000 \qquad (17)$$

- average cost of a single icebox content:

$$Ic = 250 \text{ EUR} \qquad (18)$$

The investment costs will consist of:
- the cost of purchasing 1000 sensors with assembly, estimated 55 EUR / icebox which gives a total of 55 000 EUR
- the cost of purchasing WirelessDataCollector devices in one warehouse with assembly, estimated EUR 5,000

- the software cost including implementation services, estimated at EUR 40,000
- the annual cost of maintaining the system (service fee), estimated EUR 12,000

The total estimated investment cost over 3 years will be:

$$Ti = 55000+5000+40000+3x12000 = 136000 \text{ EUR} \quad (19)$$

The estimate of savings is much more difficult. During the tests it was found that in 37.5% of the samples tested the temperature was exceeded. Based on the tests carried out, however, the question which of the real losses in goods are caused by the detected temperature overruns cannot be answered. It is not known what types of goods were transported, how sensitive they were to incorrect storage and how long they were kept under wrong temperatures. There is also no data on actual losses of goods in the company in which dispatches were carried out in a given period. This is why we propose to count ROI in reverse.

Assuming a scenario accepted by many managers, that to make an investment decision, it must pay for itself within 3 years, i.e.:

$$ROI = 0 \qquad (20)$$

The savings due to the lack of losses of goods would have to amount to:

$$Gp = Ti=136\ 000 \text{ EUR} \qquad (21)$$

Translating this into the number of iceboxes whose content would have to be saved over 3 years (Ics):

$$Ics = Ti \ / \ Ic = 544 \qquad (22)$$

The percentage of transport orders (iceboxes) that would have to be saved (Nos) is:

$$Nos = Ics \ / \ No * 100\% = 440 \ / \ 3x25000 = 0.73 \qquad (23)$$

This means that it is enough for the implemented system to improve effectively the prevention of goods loss at the level of 0.73% or higher. Then the profitability of implementation of the system will be sensible and will achieve the assumed ROI. By comparing the calculated value of temperature disruptions (35.7%) with the minimum number of saved transport orders numbers, the minimum detection efficiency Ef for the described ANN method can be determined as:

$$Ef=(Nos*100)/35.7=2.05\% \qquad (24)$$

6. CONCLUSIONS

The data analysis enables the identification of the density of disruptions in cold chain transport. Frequent disruptions in transport include too slow cooling (35.7% cases) because of improper cooling, i.e. too low refrigerant, damaged coolbox and cooling anomaly (41.8% cases) - unexpected temperature normalisation or heating during the cooling process. The problem with reaching the maximum possible temperature with refrigerant,but this temperature being higher than required in transport, is typical of 13.9% cases. Using not ready coolbox (heat coolbox) cause not reaching the proper temperature of the process (is found in 8.6% of cases).

The presented method enables the disruption in the cold chain to be predicted in over 91% cases. Consequently, preventive action can be undertaken by reaching the required temperature condition. The identification of the cause of the problems is the key issue for each transport manager. The presented method enables the major problems of companies to be identified:

– worn or damaged coolbox. In such coolboxes proper temperature is not maintained, so using them in transportation causes breaks in the transport chain, increases the cost of refrigerant, increases the risk of lower quality of products and complaints. Owing to the presented method it is not hard to identify coolboxes which need servicing;

– insufficient quantity of refrigerant. Next time the picker or the driver can add more refrigerant into the coolbox when the system alerts him and sends prompts;

– elimination not ready coolboxes, ie. those with high temperature, from being used;

– opening of coolbox during transportation without reason.

Once these cases are identified, the cold chain can be improved. This method increases the quality of products delivered to the customer and makes supply chain resilience for risky cases and disruptions.

The presented methodology can be used by cargo forwarders, carriers and 3PL companies. The implementation of this method will increase the quality of transport and reduce the complaints of clients. What is more, it helps to reduce the waste of goods.

Future research will focus on other disruption cases, employing additional data from humidity, vibration and light sensors. Using other factors correlated with temperature could increase the probability of state prediction.

REFERENCES

[1] Ivanov, D., Dolgui, A., Sokolov, B. (2019). 'The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics'. International Journal of Production Research, 57(3), 829-846

[2] Das A., Gottlieb S., Ivanov, D. (2019). 'Managing disruptions and the ripple effect in digital sup-ply chains: empirical case studies', in Ivanov D. et al. (Eds.) Handbook of the Ripple Effects in the Supply Chain. Springer, New York, pp. 261-285.

[3] Dolgui, A., Ivanov, D., Potryasaev, S., Sokolov, B., Ivanova, M., Werner, F. (2020). 'Blockchain-oriented dynamic modelling of smart contract design and execution in the supply chain'. International Journal of Production Research. 58. pp. 2184-2199.

[4] Dubey, R., Altay, N., Gunasekaran, A., Blome, C., Papadopoulos, T., Childe SJ (2018) 'Supply chain agility, adaptability and alignment: empirical evidence from the Indian auto components industry'. International Journal of Operations & Production Management 38 (1), 129-148.

[5] Tabakow, M., Korczak, J., Franczyk, B. et al. (2014) 'Big Data - Definitions, Challenges and Information Technologies', Publishing House of Wroclaw University of Economics, vol. 1, no. 31.

[6] D. Laney. (2001) '3D Data Management: Controlling Data Volume, Velocity, and Variety' META Group Research Note, 6.

[7] Corchado, J. M. and Fyfe, C. (1999) 'Unsupervised neural method for temperature forecasting', Artificial Intelligence in Engineering. Elsevier Science Ltd, 13(4), pp. 351–357. doi: 10.1016/S0954-1810(99)00007-2.

[8] Abdel-Aal, R. E. (2004) 'Hourly temperature forecasting using abductive networks', Engineering Applications of Artificial Intelligence, 17(5), pp. 543–556. doi: 10.1016/j.engappai.2004.04.002.

[9] Lee, L. W., Wang, L. H. and Chen, S. M. (2007) 'Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms', Expert Systems with Applications, 33(3), pp. 539–550.

[10] Lee, L. W., Wang, L. H. and Chen, S. M. (2008) 'Temperature prediction and TAIFEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques', Expert Systems with Applications. Elsevier Ltd, 34(1), pp. 328–336.

[11] Smith, B. A., Hoogenboom, G. and McClendon, R. W. (2009) 'Artificial neural networks for automated year-round temperature prediction', Computers and Electronics in Agriculture, 68(1), pp. 52–61.

[12] Chevalier, R. F. et al. (2011) 'Support vector regression with reduced training sets for air temperature prediction: A comparison with artificial neural networks', Neural Computing and Applications, 20(1), pp. 151–159. doi: 10.1007/s00521-010-0363-y.

[13] Hu lsmann, M. and Brenner, V. (2011) Causes and effects of cold chain ruptures performance of fragmented versus integrated cold chains. Jacobs Univ., School of Engineering and Science, Internat. Logistics, Systems Management.

[14] Jedermann, R., Praeger, U., et al. (2014) 'Remote quality monitoring in the banana chain', Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. Royal Society, 372(2017).

[15] Jedermann, R., Nicometo, M., et al. (2014) 'Reducing food losses by intelligent food logistics', Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. Royal Society, 372(2017).

[16] Wawryk, A., Mavromalis, C. and Gold, M. (1997) 'Electronic monitoring of vaccine cold chain in a metropolitan area', British Medical Journal, 315(7107), p. 518.

[17] Monteleone, S., Sampaio, M. and Maia, R. F. (2017) 'A novel deployment of smart Cold Chain system using 2G-RFID-Sys temperature monitoring in medicine Cold Chain based on Internet of Things', in Proceedings - 2017 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2017. Institute of Electrical and Electronics Engineers Inc., pp. 205–210.

[18] Xu, W. et al. (2014) 'Neural network model for the risk prediction in cold chain logistics', International Journal of Multimedia and Ubiquitous Engineering. Science and Engineering Research Support Society, 9(8), pp. 111–124.

[19] Panetto, H., Iung, B., Ivanov, D., Weichhart, G., Wang, X. (2019). 'Challenges for the Cyber-Physical Manufacturing Enterprises of the Future'. Annual Reviews in Control. Volume 47, pp 200-213, ISSN 1367-5788.

[20] Campos, Y., Villa, J. L. (2018). 'Technologies applied in the monitoring and control of the temperature in the Cold Chain'. 2018 IEEE 2nd Colombian Conference on Robotics and Automation, CCRA 2018.

[21] Ivanov, D., Sokolov, B., Käschel, J. (2009). 'Structure dynamics control-based framework for adaptive reconfiguration of collaborative enterprise networks'. International Journal of Manufacturing Technology and Management, Vol.17, No.1/2. pp. 23-41.

[22] Lorenc, A., Burinskiene, A. (2021). 'Improve the orders picking in e-commerce by using WMS data and BigData analysis'. FME Transactions, Vol. 49, No. 1. pp. 233-243.

[23] Rücker, A., Rief, J., Fottner, J., (2020). 'Development of a Method for the Energy Efficiency Determination of Stacker Cranes in Automated High-Bay Warehouses'. FME Transactions, Vol. 48, No. 4. pp. 753-760.

АНАЛИТИКА ВЕЛИКИХ ПОДАТАКА И ПРЕДВИЂАЊЕ АНОМАЛИЈА КОД ХЛАДНОГ ЛАНЦА У ЦИЉУ ПОСТИЗАЊА ЕЛАСТИЧНОСТИ ЛАНЦА СНАБДЕВАЊА

А. Лоренц, М. Чуба, Ј. Шарата

Циљ истраживања је развој методе предвиђања за спречавање поремећаја изазваних температурним аномалијама у хладном ланцу снабдевања. Анализирани подаци обухватају цео радни циклус термалног контејнера. Анализа великих података и математичко моделирање су коришћени за откривање поремећаја. Помоћу вештачких неуронских мрежа извршено је предвиђање могућих поремећаја у транспорту везаних за температуру. Истраживање је показало да је могуће спречити 82% поремећаја у хладном ланцу. Вештачке неуронске мреже омогућавају да се изврши анализа температурне криве и предвиђање поремећаја пре него што настану. Истраживање је било ограничено на транспорт хране у хладним кутијама испод 20^0C, али би се метод могао користити и за транспорт пуног утовара код транспорта хладњачама. Истраживање је засновано на реалним подацима, развијени метод је од помоћи у смањењу шкарта код хладног ланца, побољшава квалитет транспорта и еластичност ланца снабдевања.

Такође, избегавају се прекиди у хладном ланцу и смањује оштећење производа, чиме се побољшава процес транспорта. Могу га користити шпедитери, као и независни добављачи логистике у циљу смањења трошкова и шкарта. Преглед литературе потврђује да не постоји сличан метод за спречавање поремећаја у транспортном ланцу. Коришћење Интернета ствари (IoT) за прикупљање података повезаних са анализом великих података и вештачким неуронским мрежама обезбеђује еластичност ланца снабдевања.