TEHNIČKI GLASNIK

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 17 – BROJ / ISSUE 3

RUJAN 2023 / SEPTEMBER 2023 – STRANICA / PAGES 305-477

/

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

ISSN 1846-6168 (PRINT) / ISSN 1848-5588 (ONLINE)

TECHNICAL JOURNAL

ISSN 1846-6168 (Print) ISSN 1848-5588 (Online)

TEHNIČKI GLASNIK - TECHNICAL JOURNAL

Scientific-professional journal of University North

Volume 17 Issue 3
Varaždin, September 2023 Pages 305-477

Editorial Office:
Sveučilište Sjever / University North – Tehnički glasnik / Technical journal
Sveučilišni centar Varaždin / University Center Varaždin
Jurja Križanića 31b, 42000 Varaždin, Croatia
Tel. ++385 42 493 338, Fax.++385 42 493 336
E-mail: tehnickiglasnik@unin.hr
https://tehnickiglasnik.unin.hr
https://www.unin.hr/djelatnost/izdavastvo/tehnicki-glasnik/
https://hrcak.srce.hr/tehnickiglasnik

Founder and Publisher:
Sveučilište Sjever / University North

Print:
Centar za digitalno nakladništvo, Sveučilište Sjever

All manuscripts published in journal have been reviewed.
Manuscripts are not returned.

The journal is free of charge and four issues per year are published
(In March, June, September and December)
Circulation: 100 copies

Journal is indexed and abstracted in:
Web of Science Core Collection (Emerging Sources Citation Index - ESCI), Scopus, EBSCOhost Academic Search Complete, EBSCOhost – One Belt, One Road Reference Source Product, ERIH PLUS, CITEFACTOR – Academic Scientific Journals, DOAJ – Directory of Open Access Journals, Hrčak – Portal znanstvenih časopisa RH

Registration of journal:
The journal "Tehnički glasnik" is listed in the HGK Register on the issuance and distribution of printed editions on the 18th October 2007 under number 825.

Preparation ended:	Published (online):	Published (print):
July 12, 2023	July 15, 2023	September 15, 2023

Legend:
(1) University North, (2) University of Slavonski Brod, (3) Faculty of Graphic Arts Zagreb, (4) Faculty of Civil Engineering Osijek, (5) Faculty of Engineering Rijeka, (6) Faculty of Mechanical Engineering and Naval Architecture Zagreb, (7) Faculty of Metallurgy Sisak, (8) Tomas Bata University in Zlín, (9) Department of Physics of the University of Josip Juraj Strossmayer in Osijek, (10) Faculty of Humanities and Social Sciences Osijek, (11) Karlovac University of Applied Sciences, (12) University of Applied Sciences Velika Gorica, (13) Department of Polytechnics - Faculty of Humanities and Social Sciences Rijeka, (14) Faculty of Electrical Engineering and Computer Science - University of Maribor, (15) Faculty of Civil Engineering - University of Maribor, (16) University College of Teacher Education of Christian Churches Vienna/Krems, (17) Faculty of Mechanical Engineering - Poznan University of Technology (Poland), (18) Mechanical Engineering Faculty Sarajevo, (19) University of Travnik - Faculty of Technical Studies, (20) Higher Education Technical School of Professional Studies in Novi Sad, (21) University of Novi Sad - Faculty of Technical Sciences, (22) Faculty of Mechanical Engineering - University of Montenegro, (23) Brno University of Technology, (24) Odessa State Academy of Civil Engineering and Architecture, (25) Faculty of Civil Engineering - University of Mostar, (26) Faculty of Manufacturing Technologies with the seat in Prešov - Technical University in Košice, (27) Faculty of Mechanical Engineering - University of Maribor, (28) College of Engineering, IT & Environment - Charles Darwin University, (29) Universite Libre de Bruxelles, (30) Vishwakarma Institute of Information Technology (Pune, India), (31) AISSMS Institute of Information Technology (Pune, India), (32) Permtech Research Solutions (India), (33) University of Belgrade, (34) National Dong Hwa University - Taiwan, (35) Faculty of Mechanical Engineering - Opole University of Technology (Poland), (36) TU Berlin - Germany, (37) Shahid Chamran University of Ahvaz - Iran, (38) University of Bologna - Italy, (39) University of Defence in Belgrade - Military Academy - Serbia

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220819001236

Original scientific paper
Received: 2022-08-19, Accepted: 2022-12-01

Improving the Performance of Patch Antenna by Applying Bandwidth Enhancement Techniques for 5G Applications

Seda Ermiş*, Murat Demirci

Abstract: In this study, various Rectangular Microstrip Antenna (RMA) designs operating at 28 GHz frequency for 5G-communication system are performed. All designs are generated and analyzed using a 3D electromagnetic simulation program, ANSYS HFSS (High-Frequency Structure Simulator). Single and array type RMA designs are constructed by using non-contact inset-fed feeding technique. Subsequently, the bandwidth of RMAs is increased by slotting on the ground surface, and adding a parasitic element to the antenna structure. Because of these analyses, for single type RMA, the bandwidth increases from 2.09 GHz to 3.45 GHz. Moreover, for 1 × 2 and 1 × 4 array type RMAs, very wide bandwidths of 7.53 GHz and 4.53 GHz, respectively, are obtained by applying bandwidth enhancement techniques. The success of the study has been demonstrated by comparing outputs of the designs with the some similar, experimental or simulation studies published in the literature.

Keywords: DGS; Rectangular Patch Antenna; 5G; 28 GHz

1 INTRODUCTION

Due to widespread use of new technology devices in recent years, the growing demand for multimedia applications and wireless data creates a significant burden on existing cellular networks. After 4G mobile network, which has been available worldwide since 2009, the 5th Generation (5G) mobile communication technology is expected to show a revolutionary development in terms of network coverage, data rate, latency, network reliability and energy efficiency [1]. With the wide-scale deployment of 5G, mobile network will be required 1000 times higher capacity and 10-100 times faster data transmission rate than the current mobile technology. This is mainly because 5G is expected to obtain a reliable communication network and stable connection not just for phones and computers, but also various types of IoT devices such as self-driving vehicles, robots, cameras or smart home gadgets [1]. Since traditional 4G/LTE network do not provide large bandwidths in giga-bits for 5G applications, several new frequency bands between 20 and 70 GHz, also known millimeter wave bands, are identified in World Radio Communication Conference 2019 (WRC-19) report [2]. However, operational frequencies around Ku band and more specifically 28/38 GHz are prominent due to their low atmospheric attenuation [3].

Antenna design for 5G devices is very crucial to perform communication in specified millimeter wave frequencies with higher gain, enhanced bandwidth and lesser radiation losses [4]. In this sense, microstrip patch antennas emerge as a strong candidate because of their numerous attractive features such as small size, low profile, ease of production, high reliability etc. In addition, microstrip antennas can tolerate path loss in terms of gain and efficiency at higher frequencies of 5G technology. However, despite of these bountiful advantages, one major problem is their narrow bandwidth [4].

To overcome this disadvantage, bandwidth enhancement techniques such as adding parasitic element, slotting shapes on the patch surface, defecting ground structure (DGS), increasing substrate thickness or coupling type of feeding are commonly used in the designs.

Literature shows that various microstrip antennas for 5G have been studied by researches recently and in some of them, bandwidth enhancement techniques have been applied. For instance, Seyyedehelnaz Ershadi et al. designed a rectangular microstrip antenna with a 4-layer substrate. The use of multiple substrates increased the bandwidth up to 21% for 28 GHz resonance frequency [5]. In 2017, Saeed Ur Rahman et al. presented a single and a 1×2 array microstrip patch antenna with the quarter wave transformation method. The 1×2 array antenna resonated at 26.5 GHz and 28.8 GHz frequencies provided 28% impedance bandwidth [6]. In the study published by Nanae Yoon and Chulhun Seo (2017), a microstrip patch antenna that communicates at 28 GHz was designed. They investigated the effects of single and double U-shaped slits on bandwidth and gain. The study showed that opening slits on the single and 2×2 array rectangular patches increased bandwidth [7]. In 2018, Kyoseung Keum and Jaehoon Choi simulated a single and a 4×4 rectangular microstrip array antenna with double U-shaped slot on the patch. The bandwidths were 3.77 GHz, and 4.71 GHz for the single and 4×4 array antennas respectively [8]. Wahaj Abbas Awan et al. designed a microstrip patch antenna operating at 28 GHz frequency for 5G technology in 2019. It was observed that the bandwidth was increased from 1.33 GHz to 1.38 GHz and the return loss was decreased from -46.97 dB to -56.95 dB by using defected ground structure (DGS) [9]. In 2020, Sharaf et al. proposed a compact dual-frequency (38/60 GHz) microstrip patch antenna for dual-band 5G mobile applications. In the design, two electromagnetically coupled patches were used and experimental results showed that achieved impedance bandwidths are about 2 GHz and 3.2 GHz in the 38 GHz and 60 GHz bands, respectively [10]. In the scientific report published by Marasco et al., in 2022, a novel, miniaturized evolved patch antenna design was introduced for flexible and bendable 5G IoT devices and its radiation properties was enhanced by using a Split Ring Resonator (SRR) in the sub-6GHz frequency band [11]. In

2022, Ezzulddin et al. fabricated and analyzed rectangular, circular and triangular microstrip patch antennas operating at 28 GHz for 5G applications. Measurements showed that achieved bandwidths of rectangular, circular and triangular microstrip patches were 0.904, 0.848 and 0.744 GHz, with the gains of 6.44, 6.03 and 5.26 dB, respectively [12].

In the published studies, some of them are summarized above, bandwidth enhancement techniques have been used, different feeding methods and substrate materials have been tested, single-element or array-shaped microstrip antenna designs have been analyzed. However, the problem is important parameters such as return loss or antenna gain decrease significantly while increasing the bandwidth.

In this study, by using ANSYS HFSS simulation program, various Rectangular Microstrip Antennas (RMA) operating at 28 GHz frequency for 5G technology are designed and analyzed. To increase antenna bandwidth, parasite patch element is added to the antenna structure and DGS technique is applied on the ground surface. When outcomes of these designs are compared with the results of the experimental and simulation studies in the literature, it has been concluded that the study is quite successful.

2 THEORETICAL BACKGROUND

Microstrip antennas are widely used due to their advantages such as lightness, small volume and low cost. In its most basic form, a RMA geometry consists of the ground plane, dielectric layer and radiating patch as shown in Fig. 1.

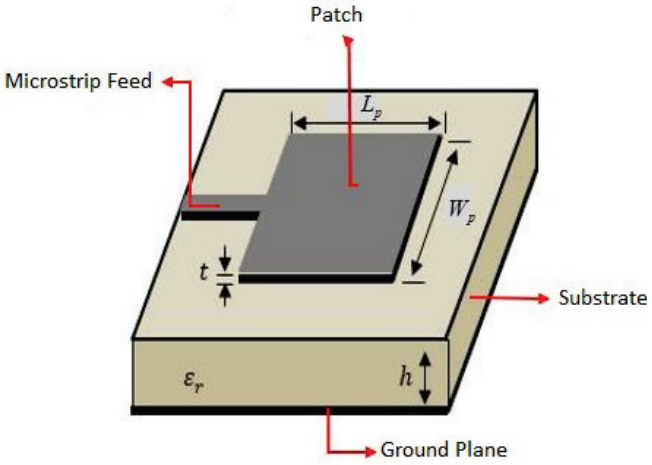


Figure 1 Rectangular Microstrip Antenna [13]

The most common patch geometry in terms of usage area is rectangular shaped patches. The width of the radiating patch, W_p, its length, L_p, the dielectric constant of the substrate (dielectric layer), ε_r and its thickness, h are shown in Fig. 1. Since thicker materials with low dielectric constant provide better radiation efficiency and wider bandwidth, a low-loss dielectric substrate Rogers RT/duroid 5880 with dielectric constant $\varepsilon_r = 2.2$ and loss tangent $\tan\delta = 0.0009$ is chosen.

As a first step, the patch width W_p and length L_p are calculated by using operation frequency f_r of the antenna and the dielectric constant of the substrate, as given in Eq. (1) [13].

$$W_p = \frac{1}{2f_r\sqrt{\mu_o\varepsilon_o}}\cdot\sqrt{\frac{2}{\varepsilon_r+1}} = \frac{\vartheta_o}{2f_r}\cdot\sqrt{\frac{2}{\varepsilon_r+1}} \tag{1}$$

Here, $\vartheta_o = c = 3\times10^8$ m/sn, $\mu_o = 4\pi\times10^{-7}$ H/m, $\varepsilon_o = 8.85\times10^{-12}$ F/m are speed of light, magnetic permeability and dielectric constant in free space respectively. The microstrip antenna has an inhomogeneous structure due to patch on the top surface, ground plane on the bottom surface and the dielectric layer between them. This structure causes the change of electrical conductivity and so, effective dielectric constant ε_{reff} is given as,

$$\varepsilon_{reff} = \frac{\varepsilon_r+1}{2} + \frac{\varepsilon_r-1}{2}\cdot\left[1+12\frac{h}{W_p}\right]^{-\frac{1}{2}} \quad \text{where } \frac{W_p}{h}>1 \tag{2}$$

Due to fringing field effect, the electrical dimension of the patch is greater than its physical dimension. The increment in length ΔL and electrical length of the patch L_{eff} are calculated by using Eqs. (3) and (4) respectively [13, 14].

$$\Delta L = \frac{0.412h\cdot(\varepsilon_{reff}+0.3)\cdot\left(\frac{W_p}{h}+0.264\right)}{(\varepsilon_{reff}-0.258)\cdot\left(\frac{W_p}{h}+0.8\right)} \tag{3}$$

$$L_{eff} = \frac{1}{2f_r\sqrt{\varepsilon_{reff}}\cdot\sqrt{\mu_o\varepsilon_o}} \tag{4}$$

Therefore, the actual length of the patch, L_p, is

$$L_p = L_{eff} - 2\Delta L = \frac{1}{2f_r\sqrt{\varepsilon_{reff}}\cdot\sqrt{\mu_o\varepsilon_o}} - 2\Delta L \tag{5}$$

Single element or array type RMAs are designed and analyzed in this study. Since non-contact inset-fed is preferred as the feeding technique, energy flow is provided indirectly by the contactless 50 Ω feeding line. The theoretical calculation of antenna and microstrip line impedances are explained in detail in Ref. 13 and 14. After the general theoretical calculations, the best gain, bandwidth and return loss values are obtained by impedance matching with the help of the software.

3 28 GHz SINGLE RMA DESIGN

RT Duroid 5880 material with dielectric constant $\varepsilon_r = 2.2$ and loss tangent $\tan\delta = 0.0009$ is used as a substrate for the 28 GHz rectangular microstrip antenna due to its low loss and low dielectric constant. The undertone substrate thickness is chosen as 0.508 mm.

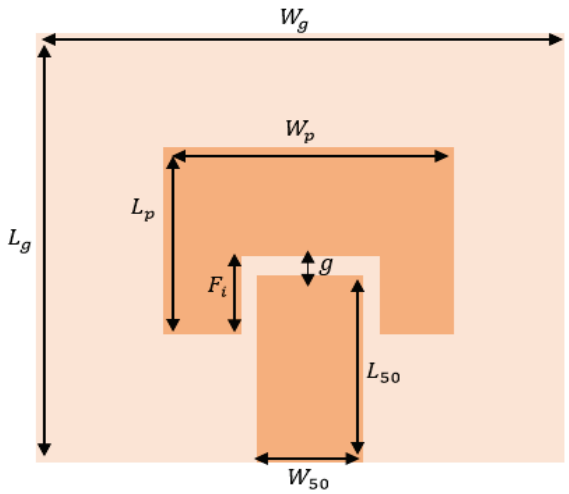


Figure 2 Top view of RMA

Table 1 Optimized dimensions of RMA (mm)

h	t	W_g	L_g	W_p	L_p	L_{50}	w_{50}	F_i	g
0.508	0.035	8	7.10	4.4	3.1	2.875	1.56	1.14	0.1

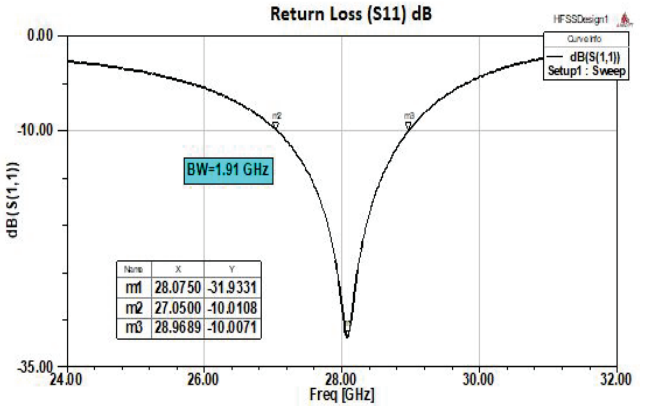


Figure 3 Return loss of RMA

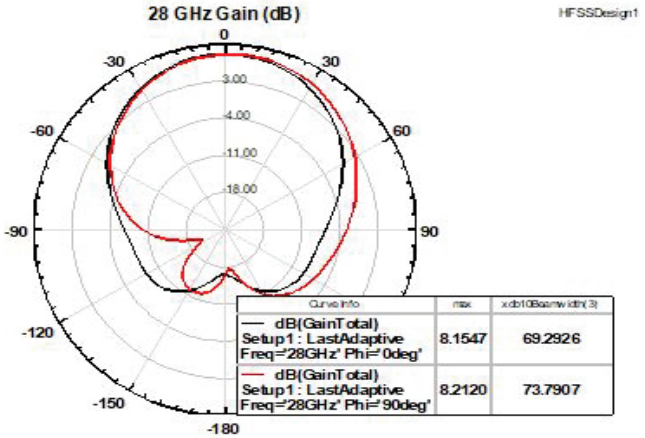


Figure 4 Polar view gain pattern of RMA

Fig. 2 shows the geometry of the RMA with non-contact inset-fed. As seen from the figure, W_p, and L_p are the patch width and length; F_i is the embedding distance of feeding line; g is gap between the feed line and the patch; h is substrate thickness; t is patch thickness; w_{50} and L_{50} are the feed line width and length; W_g and L_g are the ground surface width and length, respectively. Also, the width and length of substrate and ground surface are taken as equal. All dimensions of designed RMA are given in Tab. 1.

The return loss graph of RMA is shown in Fig. 3. RMA, emitting at a frequency of 28.075 GHz, has an impedance bandwidth of 1.91 GHz in the 27.05-28.96 GHz frequency range. As a result, designed RMA has a return loss of –31.93 dB at the frequency of 28.07 GHz, a bandwidth of 1.91 GHz and a gain of 8.20 dB (Fig. 4). As the next step, bandwidth enhancement techniques such as defecting ground structure or stacked patch technique are performed to enhance antenna bandwidth.

3.1 Application of Defected Ground Structure (DGS) and Stacked Patch Techniques for Single RMA

In microwave circuits, DGS (Defected Ground Structure) is applied by etching slots on the ground surface. Defects on the ground plane may be in the form of a single cell or periodic/aperiodic configuration of slots which depends on the application. The well-known advantages of the DGS are reducing size of component, improving bandwidth, suppressing mutual coupling or cross polarization effect and using to adjust antenna impedance for matching and for maximum power transfer.

In this part of the study, DGS is applied to the ground surface of the single RMA with non-contact inset fed by opening a slot which is in the form of a ring line aligned at the center of the patch. Bottom view of the single RMA is shown in Fig. 5(a). The radius of the ring slot is R and its width is w_R. All dimensions of the designed RMA are given in Tab. 2.

Table 2 Dimensions of RMA with Ring Shaped DGS (mm)

h	t	W_g	L_g	W_p	L_p	F_i
0.508	0.035	8	7.1	4.4	2.92	1.14
R	w_R	g	L_{50}	w_{50}		
0.6	0.3	0.1	2.875	1.56		

The more complex shape imperfections at the ground surface provide further change the path of surface currents. To see the effect of changing current distribution on the ground surface to the antenna performance parameters such as bandwidth or gain, both ring and C-shaped slots are used together as shown in Fig. 5(b). Related dimensions of the RMA for this design are given in Tab. 3.

Table 3 Dimensions of RMA with Ring / Symmetrical C Shaped DGS (mm)

h	t	W_g	L_g	W_p	L_p	F_i	R
0.508	0.035	8	7.1	4.4	2.92	1.14	0.6
w_R	L_{50}	w_{50}	g	$L_1 = L_3$	L_2	w_c	
0.3	2.875	1.56	0.1	2	1.8	0.2	

Another method to increase bandwidth of microstrip antennas is to use a stacked patch. In this method, more than one dielectric material is used and a stacked patch is added on the antenna structure. By this design, the radiation of microwave is spread and wider bandwidth can be obtained. To see the effect of stacked patch technique, two dielectric layers with same thickness ($h_1 = h_2$) and same dielectric

constant ($\varepsilon_{r1} = \varepsilon_{r2} = 2.2$) used in the structure. Main patch and contactless feed line are placed between two dielectric layers, and stacked patch is at the top of the second dielectric layer, as seen from Fig. 5(c). In addition, to combine design with the DGS, a ring shape slut is cut on the ground surface. Dimensions of the design can be found in Tab. 4 where width and length of the stacked patch are represented by W_{pp} and L_{pp} respectively.

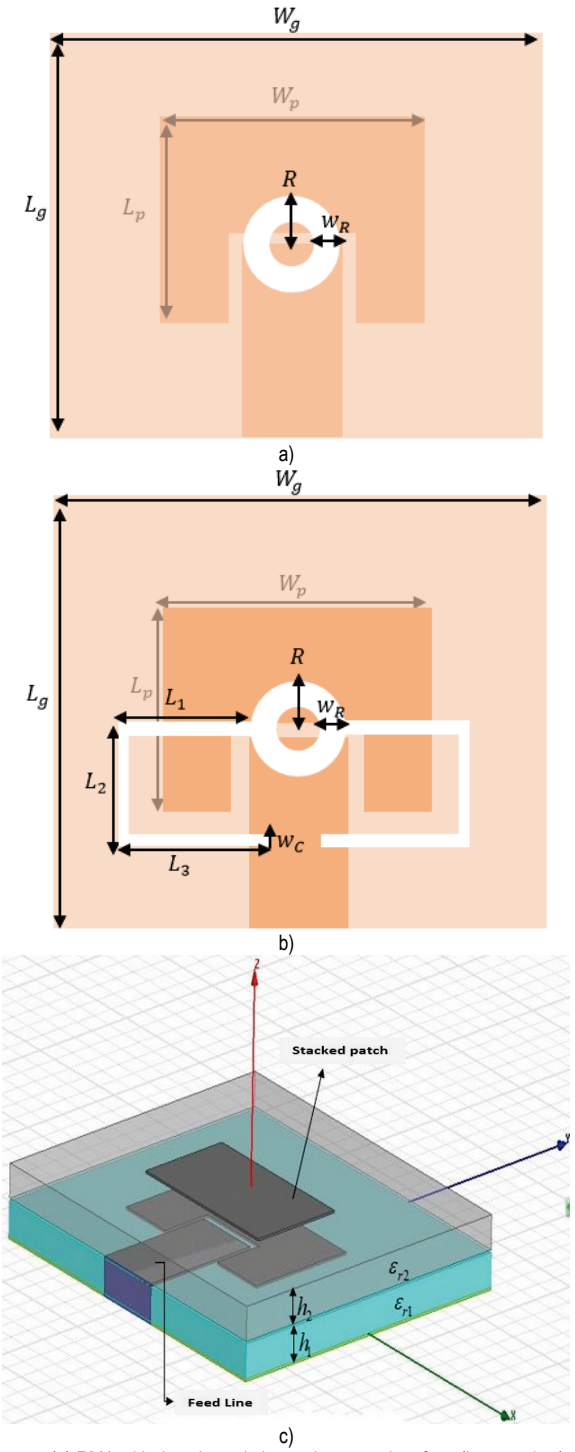


Figure 5 (a) RMA with ring shaped slot on the ground surface (bottom view); (b) RMA with ring and symmetrical C-shaped slots on the ground surface (bottom view); (c) Two-layered RMA structure combined with stacked patch and DGS.

Table 4 Dimensions of RMA with Stacked Patch and Ring Shaped DGS (mm)

h_1	h_2	t	L_{50}	w_{50}	W_p	L_p	g
0.508	0.508	0.035	2.875	1.56	4.4	2.6	0.1
L_{pp}	W_{pp}	L_g	W_g	F_i	R	w_R	
2.2	3.8	7.1	8	1.14	0.6	0.3	

Simulation results of designed RMAs given in Fig. 4, 5(a), 5(b) and 5(c) are compared and presented in Tab. 5. From the table, it is seen that the deformation on the ground surface of RMA increases bandwidth but also, it decreases antenna gain significantly. However, by the use of stacked patch technique and DGS together, clearly the widest bandwidth is obtained and significant decrease on gain is prevented.

Table 5 Comparison of Single RMA designs

Design Outputs	RMA without DGS and stacked patch	RMA with DGS		RMA with DGS and stacked patch
		RMA with ring shaped slot	RMA with ring and symmetrical C-shaped slots	
Resonance Frequency	28.075 GHz	28.03 GHz	27.99 GHz	27.96 GHz
Return Loss	–31.93 dB	–39.677dB	–28.92 dB	–26.06 dB
Gain	8.20 dB	8.08 dB	6.7 dB	8.0 dB
Bandwidth	1.91 GHz	2.09 GHz	2.69 GHz	3.45 GHz

As a next step of the study, bandwidth enhancement techniques are applied to 1 × 2 and 1 × 4 array antenna designs which are explained in the following sections.

4 28 GHz 1×2 ARRAY RMA DESIGN

The maximum gain of single microstrip patch antenna is obtained around 8 dB. As well-known from the antenna design studies, one way is creating an array structure to increase antenna directivity and gain. So, to provide higher gain, an array structure consists of two RMAs feeding by contactless microstrip lines is designed and analyzed. In Fig. 6, the geometry of the array design is shown. Dimensions of each array elements are same and width and length of the substrate and ground surface are equal.

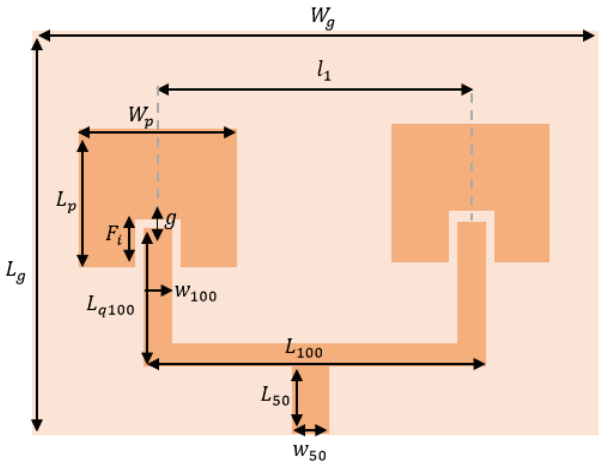


Figure 6 1×2 Array RMA dimensions

The distance (l_1) between the center points of the array elements effects the radiation pattern and changes the bandwidth and gain due to mutual coupling. Design analyses show that the widest bandwidth occurs when l_1 is chosen around 0.9λ. In addition, another issue about the design is to ensure impedance compatibility between the feed lines and patches. The impedance of the feeding line is 50 ohms and energy flow to the patches is obtained by the coupling effect of two equal 100-ohm transmission lines. The embedded distance of transmission lines into the patch has been decided by the help of simulation program to achieve impedance matching. Dimension parameters for the designed 1×2 array RMA are given in Tab. 6.

Table 6 Dimensions of 1×2 array RMA (mm)

h	t	W_g	L_g	W_p	L_p	F_i	g
0.508	0.035	18	9	4.4	2.92	0.7	0.1
L_{q100}	L_{100}	w_{100}	L_{50}	w_{50}	l_1		
1.9	9.63	1.56	2	1.56	0.9λ		

Analysis result shows that the return loss for the 28.08 GHz resonance frequency is around –44.62 dB and the bandwidth is 3.50 GHz. The gain pattern indicates that the 1×2 array RMA has 11.6 dB maximum realized gain, which is higher than the single element RMA, as expected. However, in this study, as it is aimed to increase antenna bandwidth as well as the gain, DGS and stacked patch techniques are applied to 1×2 array RMA designs which is explained in the next section.

4.1 Application of Defected Ground Structure (DGS) and Stacked Patch Techniques for 1×2 Array RMA

It has been observed that the widest bandwidth is obtained when the distance between centers of patch elements is selected as 0.9λ for the designed 1×2 Array RMA with non-contact inset-fed. In addition, ring and rod-shaped slots are cut on the ground surface and parasite patch elements are added antenna structure since it is aimed to increase the bandwidth of the 1×2 array RMA. As it can be seen from the Fig. 7(a) which is the bottom view of 1×2 array RMA, the ground surface is defected by ring and rod-shaped slots and they are located symmetrically and around the center of the two patch elements. In addition, from the Fig. 7(b), simulated 1×2 array RMA with stacked patch elements can be seen. In this design, two dielectric layers with same thickness ($h_1 = h_2$) and same dielectric constant ($\varepsilon_{r1} = \varepsilon_{r1} = 2.2$) are used.

Opening ring/rod-shaped slots on the ground surface causes a shift in the resonance frequency, and to keep resonance frequency at 28 GHz, the patch length has been reduced. While the patch length for each element is 2.92 mm in the array design without DGS, it is 2.67 mm in this design with DGS. Dimensions of the 1 × 2 array RMA represented by Fig. 7(a) and 7(b) are given in Tab. 7 where the width and length of stacked patches are represented by W_{pp} and L_{pp} respectively.

Analyses results show that by the application of stacked patches and DGS, a very wideband antenna design is made. As seen from the Fig. 8, the return loss of the design is –38.06 dB at 28 GHz and the impedance bandwidth is 7.53 GHz between 24.14 GHz and 31.66 GHz. The gain of the 1 × 2 array RMA was measured as 10.03 dB (Fig. 9). This indicates that a design that provides the desired antenna efficiency has been made, although there is a slight decrease in gain compared to previous designs.

Table 7 Dimensions of 1×2 array RMA with DGS and stacked patches (mm)

$h_2 = h_1$	t	W_g	L_g	W_p	L_p	l_1
0.508	0.035	18	9	4.4	2.7	0.9λ
L_{q100}	L_{100}	w_{100}	L_{50}	w_{50}	F_i	g
1.9	9.63	1.56	2	1.56	0.7	0.1
d	L	w_L	R	w_R	W_{pp}	L_{pp}
0.6	2.7	0.2	0.5	0.1	3.8	2

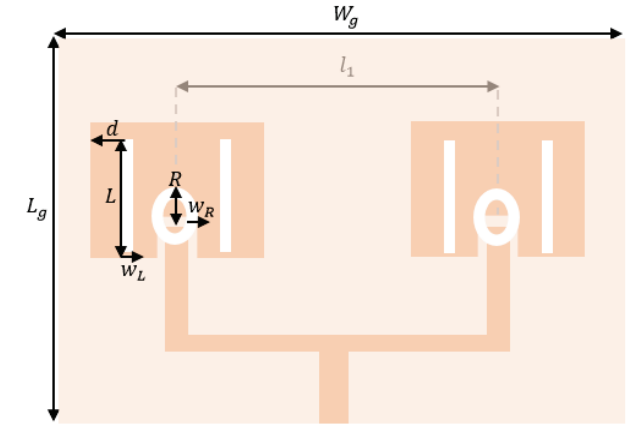


Figure 7 Bottom (a) and top (b) view of 1×2 array RMA with DGS and stacked patches

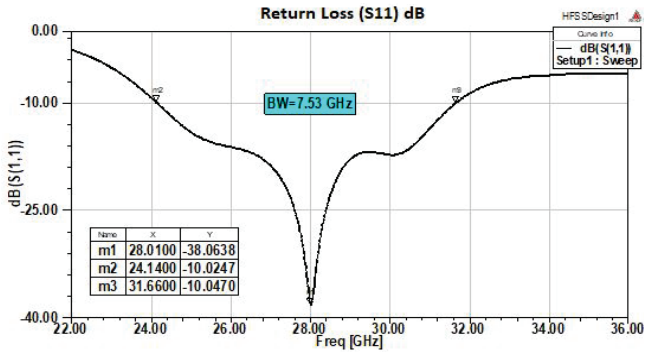


Figure 8 Return Loss of two Layer 1×2 array RMA with DGS

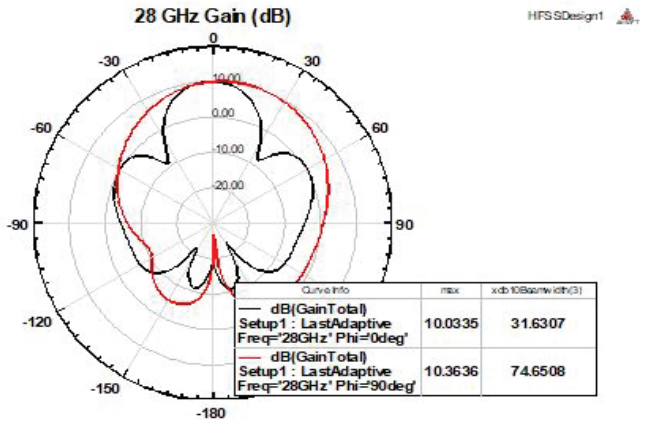


Figure 9 Gain pattern of two Layer 1 × 2 array RMA with DGS

5 28 GHz 1×4 ARRAY RMA DESIGN

In the design of 1×2 RMA array, antenna gain is varied between 10 and 11.5 dB. As a last step, 1×4 RMA array with contactless inset-fed is simulated and analyzed to increase antenna gain. Moreover, by the application of bandwidth enhancement techniques, wide band antenna is aimed at 28 GHz for 5G applications. Fig.10 shows the geometry of 1×4 array RMA design.

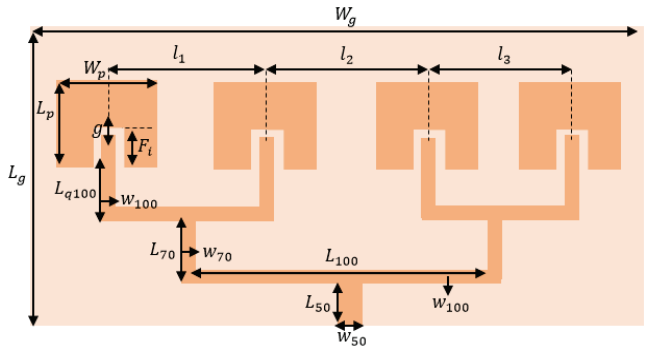


Figure 10 Geometry and dimensions of 1×4 RMA array

The distances between the centers of patch elements, given by l_1, l_2 and l_3, are defined by the help of software. The 50 ohm feed line is divided into two equivalent 100 ohm microstrip lines. Quarter wave transformation is performed by using 70 ohm microstrip line for impedance matching between two 100 ohm lines. The width and length of 70 ohm line is represented by w_{70} and L_{70}, respectively. All dimensions of the 1×4 array RMA illustrated in Fig. 10 are given in Tab. 8.

Table 8 Dimensions of 1×4 array RMA (mm)

h	t	L_g	W_g	L_p	W_p	F_i	g
0.508	0.035	11	29.7	2.9	4.35	0.8	0.1
L_{100}	L_{70}	L_{50}	L_{q100}	w_{100}	w_{70}	w_{50}	
9.63	2	2	1.9	0.45	0.9	1.56	
l_1	l_2	l_3					
0.7λ	0.5λ	0.7λ					

Analyses results show that the return loss of 1 × 4 array RMA is –43.56 dB at 27.92 GHz and –28.56 dB at 30.17 GHz. This indicates that antenna resonates at two frequencies. In addition, the bandwidth is found 4.18 GHz and maximum antenna gain is around 14 dB at 28 GHz.

5.1 Application of Defected Ground Structure (DGS) Technique for 1×4 Array RMA

In order to increase the bandwidth of the 1×4 RMA array, the ground is defected by four, concentric, double ring-shaped slots located each patch. As seen from Fig. 11, for each ring pair, the inner and outer ring radii are expressed by R_1 and R_2, with widths w_{R1} and w_{R2}, respectively. The widths and radii of slots are optimized by the HFSS software and the most suitable dimensions are defined for the best bandwidth.

The bottom view of 1×4 RMA array is given in Fig. 11 and the widths and radii of slots can be found from Tab. 9. All the other dimensions of the design are the same with the previous one, given in Tab. 8.

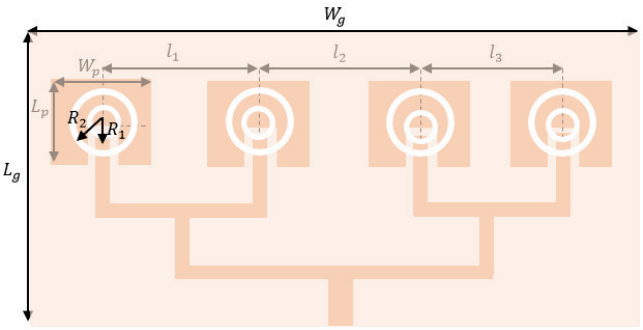


Figure 11 1×4 array RMA ground surface with DGS

Table 9 Slot dimensions of 1×4 array RMA with ring DGS (mm)

Inner Ring Radius	R_1	0.5
Inner Ring Width	w_{R1}	0.1
Outer Ring Radius	R_2	0.8
Outer Ring Width	w_{R2}	0.1

Like the previous one, in this design antenna resonates at two frequencies, which are 28.12 GHz, and 30.63 GHz. Return loss and antenna gain are –31.53 dB and 13.34 dB. As expected, gain slightly decreases due to defected ground surface. However, the bandwidth of RMA increases to 4.53 GHz.

6 RESULTS AND DISCUSSION

To see the effect of the bandwidth enhancement techniques, a comparison of all design cases with respect to antenna output parameters is given in Tab. 10 and important conclusions of the study are summarized as follows:

1) For all design cases, defected ground surface and parasitic patch element provide wider bandwidth at the operating frequency as seen from Tab. 10.
2) According to Tab. 5, if only DGS is applied to single RMA, antenna gain is reduced significantly due to complex shape imperfection and radiation loss. To avoid this situation, the combination of DGS and stacked patch technique with multiple layer dielectric substrates is performed in the design and it is observed that wider bandwidth can be achieved with slight decrease on gain. Therefore, it can be said that overall performance of

single RMA is improved by using these bandwidth enhancement techniques.

3) To provide higher gain, one well-known method is to increase antenna directivity by arraying. To do that, 1×2 and 1×4 RMA arrays are designed and analyzed. Both two array designs improve antenna gain with respect to the single RMA, as expected.
4) By applying DGS and adding parasitic patch element with multiple layer substrates to 1×2 RMA array structure, the best bandwidth (≈7.5 GHz), which is more than twice of the bandwidth of the design without using DGS and parasitic patch elements (≈3.5 GHz), is obtained. The gain for this design is around 10.3 dB which is slightly less (≈1.3dB) than 1 × 2 RMA array design without using any bandwidth enhancement methods.
5) The highest gain (≈14dB) is obtained by the 1 × 4 RMA array configuration, as expected. By applying DGS to the design, slight increase on bandwidth (9.7%) and decrease on gain (0.2 dB) is observed. Stacked patch technique is not implemented to the design to avoid high degree of design complexity, high loss and very dispersed radiation pattern which causes multiple resonance frequency.
6) For all designs, regardless of single or array RMA configurations, by the application of DGS and stacked patch, loss is increased because of adding some components such as multiple dielectric substrate or parasitic patch element to the design and deformation on the ground plane. However, for all designs, loss level is less than –10 dB so it is in acceptable range for application.

As a last step of this study, obtained results of the analyses of single and 1×2/1×4 array RMAs are compared with some similar publications in the literature and expressed in Tab. 11 and 12. As seen from Tab. 11, using the combination of DGS and stacked patch technique obviously increases bandwidth and it prevents significant decrease on gain.

Table 10 Comparison of outputs of Single and Array RMA designs in this work

Design Outputs	Single RMA without DGS and stacked patch	Single RMA with DGS and stacked patch	1 × 2 Array RMA without DGS and stacked patch	1 × 2 Array RMA with DGS and stacked patch	1 × 4 Array RMA without DGS	1 × 4 Array RMA with DGS
Resonance Frequency	28.075 GHz	27.96 GHz	28.08 GHz	28.01 GHz	27.92 GHz	28.12GHz
Return Loss	–31.93 dB	–26.06 dB	–44.62 dB	–38.06 dB	–43.57 dB	–31.53 dB
Gain	8.20 dB	8.0 dB	11.62 dB	10.36 dB	14.02 dB	13.80 dB
Bandwidth	1.91 GHz	3.45 GHz	3.50 GHz	7.53 GHz	4.13 GHz	4.53 GHz

Table 11 Comparison of Single RMA simulation results with similar studies in the literature

Reference Study	Resonance Frequency	Return Loss	Bandwidth	Gain
[12]	28.04 GHz (Rectangular Patch)	–14.81 dB	0.904 GHz	6.44 dB
	28.06 GHz (Circular Patch)	–24.12 dB	0.848 GHz	6.03 dB
	28.10 GHz (Triangular Patch)	–18.80 dB	0.744 GHz	5.26 dB
[15]	28 GHz	–39.36 dB	2.48 GHz	6.37 dB
[16]	28.3 GHz	≈–38 dB	2.4 GHz	5.56 dB
[17]	28 GHz	–40 dB	1.3GHz	7.6 dB
[18]	28 GHz	≈–28 dB	2.66 GHz	5.82 dB
[19]	28 GHz	–24 dB	2.24 GHz	7.86 dB
[20]	27.9 GHz	–15.35 dB	≈0.5 GHz	6.92 dB
[21]	28.96 GHz	–19.12 dB	3.93 GHz	6.05 dB
[22]	28 GHz	–59.17 dB	1.3 GHz	10.1 dB
Single RMA with DGS and Stacked Patch in this study	27.96 GHz	–26.06 dB	3.45 GHz	8.0 dB

Table 12 Comparison of Array RMA simulation results with similar studies in the literature

Reference Study	Array Type	Resonance Frequency	Return Loss	Bandwidth	Gain
[17]	1 × 4	28 GHz	–25.8 dB	1.2 GHz	13.5 dB
[20]	1 × 2	27.94 GHz	–13.77 dB	≈0.5 GHz	9.52 dB
	1 × 4	27.53 GHz	–21.44 dB	≈1 GHz	11.2 dB
[21]	1 × 2	29.06 GHz	–19.31 dB	2.36 GHz	7.15 dB
	1 × 4	28.85 GHz	–24.56 dB	1.17 GHz	10.27 dB
[22]	2 × 2	28 GHz	–61.19 dB	1.3 GHz	15.31 dB
[23]	1 × 2 (MIMO)	28 GHz	≈–17 dB	2.GHz	9.4 dB
Array RMA with DGS and Stacked Patch in this study	1 × 2	28.01 GHz	–38.06 dB	7.53 GHz	10.36 dB
Array RMA with DGS in this study	1 × 4	28.12 GHz	–31.53 dB	4.53 GHz	13.8 dB

Of course, it is possible to obtain wider bandwidth with lower gain or vice versa by applying different design techniques, which depends on the application and major consideration of the design. However, the overall aim of this work is to improve antenna performance by using bandwidth enhancement techniques in the design and results show that in the desired frequency range, efficient designs with wide bandwidth and good gain for single or array type RMAs are achieved for the applications of 5G.

7 REFERENCES

[1] Agiwal, M., Roy, A., & Saxena, N. (2016). Next Generation 5G Wireless Networks: A Comprehensive Survey. *IEEE Commun. Surv. Tutorials, 18*(3), 1617-1655.

https://doi.org/10.1109/COMST.2016.2532458
[2] Marcus, M. J. (2019). ITU WRC-19 Spectrum Policy Results. *IEEE Wireless Communications*, 26(6). 4-5. https://doi.org/10.1109/MWC.2019.8938175
[3] Banday, Y., Rather, G. Md., & Begh, G. R. (2019). Effect of Atmospheric Absorption on Millimeter Wave Frequencies for 5G Cellular Networks. *IET Communications, 13*(3), 265-270. https://doi.org/10.1049/iet-com.2018.5044
[4] Wang, X. et al. (2018). Millimeter Wave Communication: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials, 20*(3), 1616-1653. https://doi.org/10.1109/COMST.2018.284432
[5] Ershadi, S., Keshtkar, A., Ershadi, S., Abdelrahman, A. H., Yu, X., & Xin, H. (2016). Wideband subarray design for 5G an antenna arrays. *2016 URSI Asia-Pacific Radio Science Conference (URSI AP-RASC)*, 185-187. https://doi.org/10.1109/URSIAP-RASC.2016.7601266
[6] Rahman, S. U., Cao, Q., Hussain, I., Khalil, H., Zeeshan, M., & Nazar, W. (2017). Design of rectangular patch antenna array for 5G wireless communication. *Progress in Electromagnetics Research Symposium - Spring (PIERS2017)*, 1558-1562. https://doi.org/10.1109/PIERS.2017.8261995
[7] Yoon, N. & Seo, C. (2017). A 28-GHz Wideband 2×2 U-slot patch array antenna. *J. Electromagn. Eng. Sci.*, 17(1), 133-137. https://doi.org/10.5515/JKIEES.2017.17.3.133
[8] Keum, K. & Choi, J. (2018). A 28 GHz 4 × 4 U-Slot Patch Array Antenna for mm-wave Communication. *International Symposium on Antennas and Propagation (ISAP2018)*, 1-2.
[9] Awan, W. A., Zaidi, A., & Baghdad, A. (2019). Patch Antenna with Improved Performance Using DGS for 28 GHz Applications. *International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS2019)*, 1-4. https://doi.org/10.1109/WITS.2019.8723828
[10] Sharaf, M. H., Zaki, A. I., Hamad, R. K., & Omar, M. M. M. (2020). A Novel Dual-Band (38/60 GHz) Patch Antenna for 5G Mobile Handsets. *Sensors, 20*(9), 2541. https://doi.org/10.3390/s20092541
[11] Marasco, I., Niro, G., Mastronardi, V. M. et al. (2022). A Compact Evolved Antenna for 5G Communications. *Sci Rep 12*, 10327. https://doi.org/10.1038/s41598-022-14447-9
[12] Ezzulddin, S. K., Hasan, S. O., & Ameen, M. M. (2022). Microstrip Patch Antenna Design, Simulation and Fabrication for 5G Applications. *Simulation Modelling Practice and Theory, 116*, 102497. https://doi.org/10.1016/j.simpat.2022.102497
[13] Balanis, C. (1982). *Antenna Theory: analysis and design*. Balanis, C. A. SERBIULA (sistema Libr. 2.0).
[14] Pozar, D. M. (2012). *Microwave Engineering*. Fourth edition. Hoboken, NJ, Wiley.
[15] Kaeib, A. F., Shebani, N. M., & Zarek, A. R. (2019). Design and Analysis of a Slotted Microstrip Antenna for 5G Communication Networks at 28 GHz. *The 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, 648-653. https://doi.org/10.1109/STA.2019.8717292
[16] Adedotun, O. K. & Johnbosco, A. I. E. (2017). Design and Optimization of a Square Patch Atenna for Milimeter Wave Applications. *International Journal of Science and Research (IJSR), 6*(4), 2192-2194.
[17] Khattak, M. I., Sohail, A., Khan, U., Barki, Z., & Witjaksono, G. (2019). Elliptical Slot Circular Patch Antenna Array with Dual Band Behavior for Future 5G Mobile Communication Networks. *Progress in Electromagnetics Research C, 89*, 133-147. https://doi.org/10.2528/PIERC18101401
[18] Rahman, A., Ng, M., Yi, A., Ahmed, U., Alam, T., Singh, M. J., & Islam, M. T. (2016). A Compact 5G Antenna Printed on Manganese Zinc Ferrite Substrate Material. *IEICE Electronics Express, J-STAGE, 13*(11), 20160377. https://doi.org/10.1587/elex.13.20160377
[19] Rahman, S. U., Cao, Q., Hussain, I., Khalil, H., Zeeshan, M., & Nazar, W. (2017). Design of Rectangular Patch Antenna Array for 5G Wireless Communication. *Progress in Electromagnetics Research Symposium - Spring (PIERS2017)*, 1558-1562. https://doi.org/10.1109/PIERS.2017.8261995
[20] EL Mashade, M. B. & Hegazy, E. A. (2018). Design and Analysis of 28 GHz Rectangular Microstrip Patch Array Antenna. *WSEAS Transactions on Communications, 17*(1), 9.
[21] Prabu, R. T. & Bai, V. T. (2019). Design of Eight Element Phased Array Microstrip Patch Antenna for 5G Application. *International Journal of Recent Technology and Engineering (IJRTE), 8*(1), 936-941.
[22] Soliman, Md. M., Imran, A. Z. Md., & Ullah, K. S. (2019). Highly Efficient 2×2 Antenna Array at 28 GHz and 38 GHz for 5G Applications. *In 2019 International Conference on Innovation in Engineering and Technology (ICIET)*, 23-24 December. https://doi.org/10.1109/ICIET48527.2019.9290692
[23] Ardianto, F. W., Lanang, F. F., Renaldy, S., & Yunita, T. (2018). Design Mimo Antenna with U-Slot Rectangular Patch Array for 5G Applications. In *2018 International Symposium on Antennas and Propagation (ISAP)* IEEE, 1-2.

Authors' contacts:

Seda Ermiş, Asst. Prof. Dr.
(Corresponding author)
Osmaniye Korkut Ata University, Engineering Faculty,
Electrical and Electronics Engineering Department,
80000, Merkez, Osmaniye, Turkey
sedaermis@osmaniye.edu.tr

Murat Demirci
Osmaniye Korkut Ata University, Engineering Faculty,
Electrical and Electronics Engineering Department,
80000, Merkez, Osmaniye, Turkey

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220828220446

Original scientific paper
Received: 2022-08-28, Accepted: 2022-10-22

Wrapper and Hybrid Feature Selection Methods Using Metaheuristic Algorithm for Chest X-Ray Images Classification: COVID-19 as a Case Study

Ali Yasar

Abstract: Covid-19 virus has led to a tremendous pandemic in more than 200 countries across the globe, leading to severe impacts on the lives and health of a large number of people globally. The emergence of Omicron (SARS-CoV-2), which is a coronavirus 2 variant, an acute respiratory syndrome which is highly mutated, has again caused social limitations around the world because of infectious and vaccine escape mutations. One of the most significant steps in the fight against covid-19 is to identify those who were infected with the virus as early as possible, to start their treatment and to minimize the risk of transmission. Detection of this disease from radiographic and radiological images is perhaps one of the quickest and most accessible methods of diagnosing patients. In this study, a computer aided system based on deep learning is proposed for rapid diagnosis of COVID-19 from chest x-ray images. First, a dataset of 5380 Chest x-ray images was collected from publicly available datasets. In the first step, the deep features of the images in the dataset are extracted by using the dataset pre-trained convolutional neural network (CNN) model. In the second step, Differential Evolution (DE), Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) algorithms were used for feature selection in order to find the features that are effective for classification of these deep features. Finally, the features obtained in two stages, Decision Tree (DT), Naive Bayes (NB), support vector machine (SVM), k-Nearest Neighbours (k-NN) and Neural Network (NN) classifiers are used for binary, triple and quadruple classification. In order to measure the success of the models objectively, 10 folds cross validation was used. As a result, 1000 features were extracted with the SqueezeNet CNN model. In the binary, triple and quadruple classification process using these features, the SVM method was found to be the best classifier. The classification successes of the SVM model are 96.02%, 86.84% and 79.87%, respectively. The results obtained from the classification process with deep feature extraction were achieved by selecting the features in the proposed method in less time and with less features. While the performance achieved is very good, further analysis is required on a larger set of COVID-19 images to obtain higher estimates of accuracy.

Keywords: Chest; Covid-19; Feature Extraction; Feature Selection; Optimization

1 INTRODUCTION

The new type of coronavirus (COVID-19), which first appeared in Wuhan Province of China, is a respiratory virus that shows symptoms such as cough, fever and shortness of breath in some patients. The virus, which has had a tremendous impact on people's daily lives and health, has affected people in more than 200 countries around the world [1].

Simple CT and chest x-ray imaging play an important role in the diagnosis of SARS-CoV-2 pneumonia, detection of complications and follow-up of the disease. These imaging techniques are also important devices in determining the severity of the disease. Chest x-ray may remain of low sensitivity for detecting pulmonary leaks in infected individuals in their early stages and in mild survivors. However, it is widely used because it is accessible even in most small settlements, and its capacity to raise the suspicion of pneumonia makes this device valuable. Chest x-ray imaging plays an important role in monitoring the course of pulmonary lesions, even in patients who have a severe course of the disease and are hospitalized in the intensive care unit [2-4]

In general, the diagnosis of COVID-19 can be made by a detailed image analysis of coronavirus symptoms, computed tomography (CT) or chest x-ray image. Otherwise, manual analysis will not be sufficient to distinguish between pneumonia and COVID-19 disease [5]. Radiological imaging is an important tool for diagnosing COVID-19. Most cases of COVID-19 have similar features on radiographic images in the early stage, particularly in the lower lobes, including bilateral, multifocal, ground-glass opacities with posterior or peripheral distribution, and pulmonary consolidation in the late stage [6].

Chest x-ray images are known to have potential in monitoring and examining various lung diseases such as infiltration, tuberculosis, pneumonia, atelectasis and hernia [7]. Chest x-ray imaging technique is considered one of the most powerful medical imaging techniques in hospital for detecting chest abnormalities. However, the biggest problem in using chest x-ray imaging technique is that it takes a long time for radiologists to read and interpret chest radiograph images. Considering the increase in the number of chest x-rays taken from patients with the Covid19 pandemic, more workload falls on radiologists [8]. Lung inflammation called pulmonary opacification, which occurs in many of the respiratory tract diseases, is also seen in the new type of Corona virus, COVID-19. The formation of such opacities renders the lung regions unrecognizable, making automatic image analysis on chest x-ray images difficult [9].

The increase in this workload may cause the spread of Covid19, an infectious disease, the increase in the workload of hospitals, the increase in the degree of illness of the patients and even the loss of the patient. The spread of these situations can lead to consequences that can lead to health and economic problems in countries, victimization of their citizens and curfews. Therefore, rapid and accurate evaluation of chest x-rays is very important Especially with the emergence of highly mutated (SARS-CoV-2) variant Omicron (B.1.1.529), its contagiousness is much faster and it has caused worldwide panic because of its vaccine-escapable mutations [10].

Today, artificial intelligence methods, image processing techniques and deep learning methods are used in the

detection of many diseases (e.g. stomach cancer detection, breast cancer detection, brain tumour detection, etc.) [11-13].

In their study, Aslan et al. (2021) proposed AlexNet and hybrid AlexNet + BiLSTM methods to automatically detect positive COVID-19 cases by using artificial neural networks on chest CT X-ray images [14].

Arpaci et al. In their study in 2021, they classified 14 features obtained from clinical studies using 6 different classifiers: BayesNet, Logistic, IBk, CR, PART and J48 [15].

Taspinar et al. They classified SVM, LR and ANN from a total of 3486 healthy, Covid-19 and viral pneumonia chest X-ray images and achieved the highest classification success of 96.7% [13].

Jaiswal et al. They performed transfer learning by fine-tuning the DenseNet201 model, one of the pre-trained CNN models using chest CT images, and proposed this model to diagnose COVID-19 positive or Covid-19 negative. They found 99.82%, 97.4% and 96.25% [16].

In this study, firstly, transfer learning was applied on the CNN pre-trained model on the publicly available Covid-19 Chest X-Ray Database, and deep features were extracted. Secondly, the obtained features were given to the NB, DT, SVM, k-NN and NN classifiers for binary, triple and quadruple classification, and their classification success was measured. Finally, features obtained from deep learning were selected with DE, ACO and PSO algorithms. The features obtained after feature selection were given to the SVM classifier for binary, triple and quadruple classification, and their classification success was measured and their success was compared. In order to measure the success of the models objectively, 10 folds cross validation was used. The main contributions of this study are listed below.

1) The proposed system adopts a new framework based on the diagnoses of COVID-19, Viral Pneumonia, and of Lung Opacification, a classification system, deep feature extraction, classification with extracted features, feature selection and classification strategies from obtained features.
2) COVID-19, Viral Pneumonia, Lung Opacifications disorders, due to the momentary carelessness of doctors in chest x-ray images, may result in the diagnosis of COVID-19 to be made on those who do not have Covid-19 or the diagnosis of Viral Pneumonia, Lung Opacifications to be made on those who have COVID-19. This causes misdiagnosis, makes people nervous, and increases health costs. The study will assist in making the correct diagnosis.
3) In the binary, triple and quadruple class experiments the SVM classifier achieved, 96.02%, 86.84% and 79.87% accuracy respectively.
4) In classification very similar results were obtained with fewer features obtained from the DE feature selection process. This will save us time in cases where time is important for diagnosis.
5) The proposed system of detection and classification of COVID-19 and other lung conditions can be efficiently used for clinical applications.
6) The proposed system will be used as a decision support system to help medical professionals, so that the global fight against Covid-19 will be made faster.
7) This article continued as follows. Chapter 2 describes the materials and methods used in this research. Chapter 3 presents the experimental results for the binary, triple and quadruple class problems. In Chapter 4, a comprehensive comparison of study performance with previous studies was made.

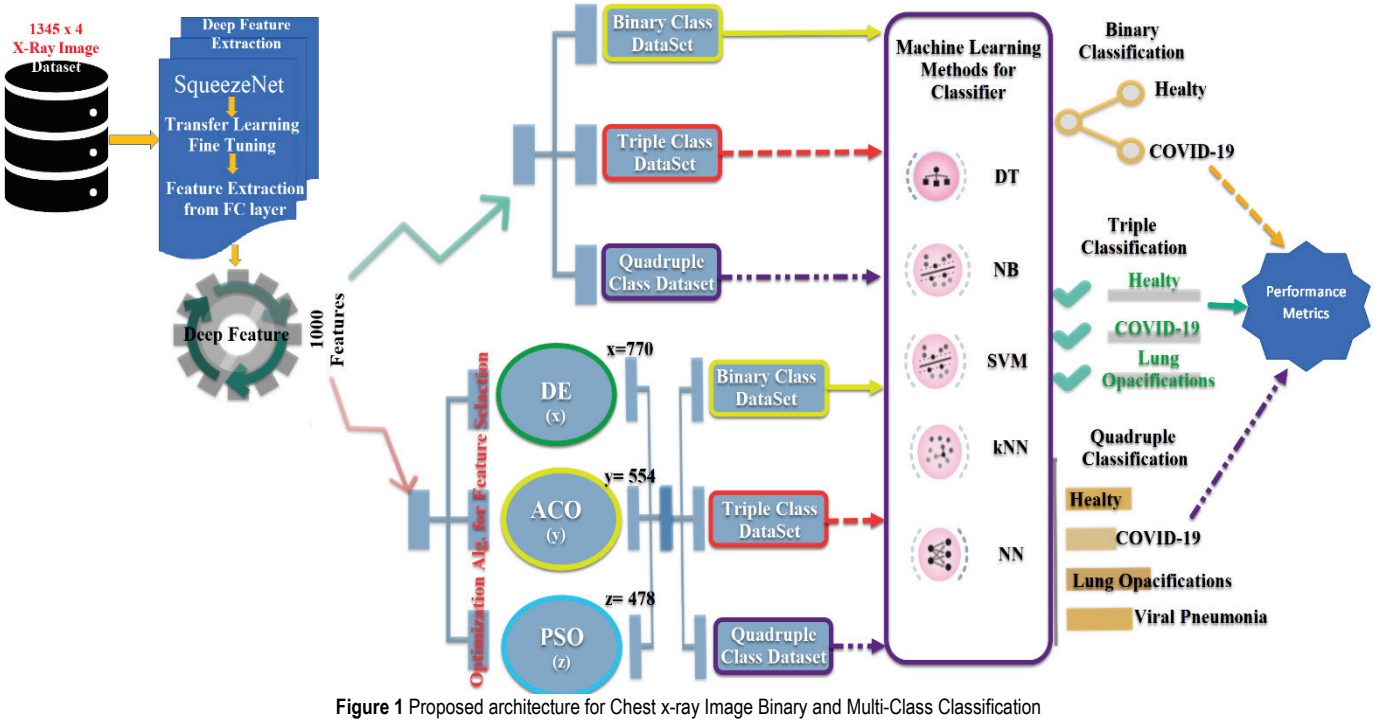


Figure 1 Proposed architecture for Chest x-ray Image Binary and Multi-Class Classification

2 MATERIALS AND METHODS

In this study, firstly Chest x-ray images for deep feature extraction procedure, the SqueezeNet model was used. Machine learning (DT, NB, SVM, K-NN and NN) algorithms classifiers are used. Secondly, redundant and redundant features, optimization methods (DE, ACO and PSO) and feature selection process were performed among the features extracted in the first stage. By using the features obtained from the process, it was subjected to the classification process with the most successful classification method in the first step. Finally, classification successes were compared. Using the chest x-ray images suggested in Figure 1, deep learning-based and feature selection-based binary (Covid-19 or Normal), triple (Covid-19, Normal, Viral Pneumonia) and quadruple (Covid-19, Normal, Viral Pneumonia, Lung Opacifications) illustrates the classification process.

2.1 Data Set and Properties

X-ray images were used in this research. Images were downloaded from the public online Kaggle dataset Repository [12]. The data set consists of chest x-ray images approved by different institutions (COVID-19, Viral Pneumonia, Lung Opacifications and Health (Normal)). There are 1345 Covid-19, 1345 Viral Pneumonia, 1345 Lung Opacifications and 1345 Health (Normal) chest x-ray images in the dataset. Each chest x-ray image is 229×229 in size and consists of a total of 5380 images. Chest x-ray images of each class are given in Fig. 2. How the chest x-ray images used for binary classification, triple classification and quadruple classification were created in the study are given in the Tab. 1.

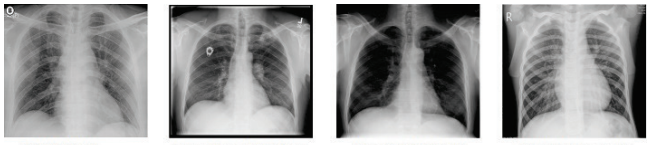


Figure 2 Chest x-rays of COVID-19, Lung Opacifications, Healthy (Normal) and Viral Pneumonia

Table 1 Contents of the dataset

Binary Classification	Triple Classification	Quadruple Classification
Healthy (Normal) + COVID-19	Healthy (Normal) + COVID-19 + Lung Opacifications	Healthy (Normal) + COVID-19 + Lung Opacifications + Viral Pneumonia

2.2 Deep Transfer Learning and Convolutional Neural Network (CNN)

It is a convolutional neural network (CNN), which is used in image recognition problems and is one of the most well-established algorithms of deep learning models[1]. Deep learning models are used in many medical applications such as classifying medical images, segmenting images, and detecting lesions from images. Medical imaging techniques such as x-ray, magnetic resonance imaging (MRI) and computed tomography (CT) are also used for signal processing and analysis of image data using CNN models. As a result of these analyses, it provides great convenience to doctors in studies such as the diagnosis of diseases such as stomach cancer, breast cancer, diabetes mellitus, skin cancer and brain tumours [17-19].

CNN consists of three layers: convolution, pooling and fully connected layer to effectively perform the process of learning from image data and testing it on new images. Feature extraction is performed in both the convolutional and pooling layers, and finally classification is performed in the fully connected layer. SqueezeNet, one of the pre-trained CNN models, was used in the study. Transfer learning enables the models to be transferred to the new model by fine-tuning the weight coefficients and parameters obtained on the previously trained data sets. Designing a new CNN model instead of transfer learning and training the model from scratch requires both very time-consuming, costly and high-performance equipment [20]. Therefore, with the transfer learning method, the researchers preferred to fine-tune the pre-designed system and transfer the information obtained by the model used on a large pre-trained dataset to the newly trained model with fewer samples. One of these models, the SqueezeNet model, was designed by training on the ImageNET dataset with a total of 1000 data classes [21].

2.3 Feature Selection

In feature selection, informative features are selected from the feature space such that the redundant and irrelevant deep features of the feature vector obtained from deep feature extraction are reduced and are effective in remote classes [22]. In this study, three features were selected from the feature space optimization algorithms (DE, ACO, PSO). Selected feature vectors obtained from all three optimization algorithms were classified by machine learning algorithms and two-class and multi-class classification methods.

2.3.1 Differential Evolution (DE)

The Differential Evolution (DE) algorithm is a population-based algorithm proposed by Price in 1995. It is promising in optimization problems with its advantages such as finding the true global minimum, using a small number of control parameters and fast convergence, independent of the initial parameter values. Important parameters include population size, scaling factor, and crossover constant. It is similar to the genetic algorithm with operators such as crossover, mutation and selection in its structure [23-25].

2.3.2 Ant Colony Optimization (ACO)

Ant Colony Optimization (ACO) is a meta-heuristic approach inspired by the pheromone trailing and tracking behaviors of some ant species [26]. It is an optimization algorithm inspired by the method used by colony-dwelling ants to find the shortest and right path between their nests and

food in their search for food. Important studies have been carried out to find solutions with this optimization algorithm in solving problems such as discrete-continuous optimization problems, traveling salesman problem, load balancing and routing in telecommunication [27].

2.3.3 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a heuristic computational technique developed by Kennedy and Eberhart in 1995 to solve problems and behavior of flying birds [28, 29]. In the selection of important features that affect the PSO classification, it can contribute to the increase of classification accuracy and performance by reducing the workload of the classifier, since PSO has a powerful discovery and the ability to explore different parts of the different particle's solution space[29]. PSO, which has the memory of the particle swarm, is very useful in the feature selection process since all particles in the problem space hold the solution information for the solution of the problem [25].

2.4 Classification Algorithms

Classification problems are one of the most classical problems that researchers use for binary or multiclass decomposition of data in various fields[30, 31]. There are studies in many fields such as agriculture, medicine, education and military. In this study, five machine learning methods NB, DT, SVM, k-NN and NN, binary (Healthy or Covid-19), triple (Viral Pneumonia, Health or Covid-19) and quadruple (Lung Opacifications) for binary and multiple classifier model were used. , Viral Pneumonia, Health or Covid-19) were used to perform diagnostic determinations. Below is a brief description of each of the algorithms.

2.4.1 Decision Tree (DT) Algorithm

Decision tree classification provides a fast and convenient solution to classify samples in large datasets containing a large number of variables. There are two key elements for constructing decision trees: (a) growing the tree to ensure it categorizes the training dataset correctly, and (b) the pruning phase where unnecessary nodes and branches are removed to improve classification accuracy [32].

2.4.2 Naive Bayes (NB) Algorithm

Naive Bayes algorithm is a probability-based classifier. For this, it calculates the probability set according to the relevance between the value frequency and combinations for binary and multiclass classification of the data in the dataset. The NB algorithm internally uses the polynomial model, Bernoulli or Gaussian model for training and testing [33, 34].

2.4.3 Multi-Class Support Vector (mSVM) Algorithm

Support Vector Machines is the model proposed by Vapnik (Widodo & Yang, 2007) as a supervised statistical learning method based on the principle of inherent risk reduction. SVM is a method for finding hyperplanes between various training data classes for classification of test data by hyperplanes in a d-dimensional feature space [7].

2.4.4 k-Nearest Neighbors (KNN) Algorithm

A grouping approach that depends on the learning information closest to the item under consideration is the K-nearest neighbors (k-NN) algorithm. The k-NN algorithm is a generalization algorithm that uses the class label of the k-example with the inductive offset for the nearest neighbor rules, the class label to be tested most similar to the closest one. It differs from the nearest neighbor in that it extends the nearest neighbor to k in the decision-making phase. This extension allows the k-NN algorithm to retrieve and use more information. Unlike other classification algorithms, they skip the learning process [35, 36].

2.4.5 Neural Network (NN) Algorithm

Effective use of ANN as a powerful machine learning-based classification method is particularly useful because of its non-linear mapping capabilities. The ANN classifier has the ability to classify the weights and biases used by the neurons in the layers in their connections with each other, using artificial neurons connected to each other. ANN performance depends on factors such as the structure of the network, the activation and transfer function, and the number of hidden layers [37].

2.5 k-Fold Cross-Validation

By dividing the data set by *k*, one of them is selected as the test data set, while the others (*k*–1) are repeated k times to be used for training purposes, and all data are tested. As a result, *k* different accuracy is obtained from all the processed data. The variance of the estimation obtained from the cross-correction process decreases as the k-folds value increases, and the *k*-value is generally used as 10 in studies. The purpose of this method is to prevent randomness in the prediction results. The disadvantage is the longer duration of the training algorithm. In this study, validation was performed as k = 10 [38]. Fig. 3 shows the 10 fold cross validation process.

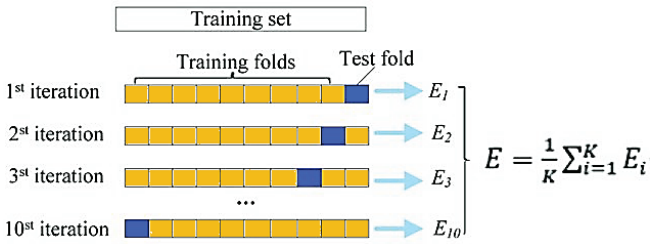


Figure 3 *k* = 10 Fold Cross-Validation

2.6 Comparative Analysis

Confusion matrix was used to calculate the performance the complexity matrix was used to calculate the metrics of classification performances and to test the usability of the proposed methods. We obtain information about the real

class and the predicted class numbers from the confusion matrix. Fig. 4 shows the confusion matrix for dual class confusion matrix and multi class classification. The Accuracy, Sensitivity, Specificity, Precision, F1-Score and Correlation Coefficient (*CC*) results of each method were calculated with the Complexity matrices we obtained as a result of the classification processes. These calculations are given in equations 1-6 for binary classification.

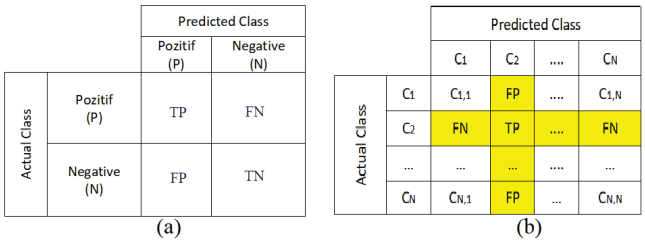


Figure 4 Confusion matrix examples. (a) Binary classification problem confusion matrix. (b) Multiclass classification problem confusion matrix.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN}, \quad (2)$$

$$Specificity = \frac{TN}{TN + FP}, \quad (3)$$

$$Precision = \frac{TP}{TP + FP}, \quad (4)$$

$$F1 - Score = 2 \times \frac{Sensitivity \times Precision}{Sensitivity + Precision}, \quad (5)$$

$$CC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN) + (TP + FP) + (FN + TN) \times (FP + TN)}}. \quad (6)$$

3 EXPERIMENTAL WORK AND RESULTS

A laptop computer with 32 GB RAM (3200 MHz), Intel (R) Core (TM) i7-10750H CPU @ 2.60GHz and 500 GB NVM-2 SSD HDD was used for the present study. The program has a user-friendly interface designed in Matlab. Experiments are performed in two different scenarios to classify and detect COVID-19 using X-ray images. First, the CNN pre-trained model of X-ray images is trained separately to classify in two, three, and four categories using deep features obtained with SqueezeNet: Machine learning methods (NB, DT, SVM, k-NN, NN) classifier for classifying these feature vectors and (Healty or Covid-19), (Viral Pneumonia, Health or Covid-19) and (Lung Opacifications, Viral Pneumonia, Health or Covid-19) are used to train/test data with 10 fold cross validation for the classification process. Second, deep features were extracted from the CNN pre-trained model, SqueezeNet, and feature selection was performed using optimization techniques (DE, ACO and PSO) and a final feature vector was created. These last feature vectors were used to train/test the mSVM method, which was the most successful classification method in the first stage. In the classification, (Healthy or Covid-19), (Viral Pneumonia, Healthy or Covid-19) and (Lung Opacifications, Viral Pneumonia, Healthy or Covid-19) were classified separately. In order to measure the success of the models objectively, 10 folds cross validation was used.

3.1 Classification Evaluation without Feature Selection

Five prediction models were created using the training dataset for machine learning methods (NB, DT, SVM, k-NN, NN) classifiers. The models were constructed using all the features obtained from deep feature extraction of the Chest x-ray image dataset without any feature selection process.

3.1.1 Binary Classification

In the first phase, Health or Covid-19 x-ray image data was used for binary classification. Our images consist of 1345 healthy and 1345 COVID-19 images. Classification performances of the X-ray images in the dataset using the feature vector (2690x1000) obtained with the deep features extracted with the CNN pre-trained model SqueezeNet are given in Tab. 1.

3.1.2 Triple Classification

In the second phase, it was performed for triple classification among Covid-19, Normal, and Viral Pneumonia x-ray image datasets. It consists of 1345 Normal images, 1345 COVID-19 images, and 1345 Viral Pneumonia images. Classification performances of the X-ray images in the dataset using the feature vector (3035x1000) obtained with the deep features extracted with the CNN pre-trained model SqueezeNet are given in Tab. 2.

3.1.3 Quadruple Classification

In the third phase, it was performed for quadruple classification among X-ray image datasets of Covid-19, Normal, Viral Pneumonia, and Lung Opacifications. It consists of 1345 Normal images, 1345 COVID-19 images, 1345 Viral Pneumonia images and 1345 Lung Opacifications images. Classification performances of the X-ray images in the dataset using the feature vector (5380×1000) obtained with the deep features extracted with the CNN pre-trained model SqueezeNet are given in Tab. 3.

3.2 Classification Evaluation with Feature Selection

This section consists of two stages. In the first stage, feature selection was made with optimization algorithms (DE, ACO and PSO). The second step was to evaluate the performance of the models based on the features selected from all three optimization algorithms. Classification was made with the mSVM classifier method, which is the best classifier in classification processes, without feature selection in the evaluation process. The X-ray images in the dataset were extracted with the pre-trained CNN model, SqueezeNet, and a 770-featured vector was obtained with the DE feature selection process, a 478-featured vector with the

PSO feature selection process, and a 554-featured vector with the ACO feature selection process.

3.2.1 Binary Classifications with Selected Attributes

The first experiment was for binary classification between Covid-19 or Normal x-ray image datasets. It consists of 1345 Normal images and 1345 COVID-19 images. Finally, feature vectors (2690×770), (2690×554) and (2690×478) were created for DE, ACO and PSO, respectively. These feature vectors were classified with the mSVM classifier and their classification performances were measured. The classification performances of the data formed with the selected features are given in Tab. 4.

3.2.2 Triple Classification

The second experiment was performed for triple classification among Covid-19, Normal, and Viral Pneumonia x-ray image datasets. It consists of 1345 Normal images, 1345 COVID-19 images, and 1345 Viral Pneumonia images. Feature vectors (4035×770), (4035×554) and (4035×478) were created for DE, ACO and PSO, respectively. These feature vectors were classified with the mSVM classifier and their classification performances were measured. The classification performances of the data formed with the selected features are given in Tab. 5.

Table 2 Binary Class Classification Performance Values

Machine Learning Algorithms for Classifier	Performances					
	Accuracy	Sensitivity	Specificity	Precision	F1-Score	CC
Decision Tree	94.28	92.71	95.84	95.70	91.18	88.59
Naive Bayes	94.83	91.82	97.84	97.71	94.67	89.83
mSVM	96.02	93.68	98.36	98.28	95.93	92.15
KNN	95.28	93.23	97.32	97.21	95.18	90.63
NN	95.84	94.94	96.73	96.67	95.80	91.69

Table 3 Triple Binary Class Classification Performance Values

Machine Learning Algorithms for Classifier	Performances					
	Accuracy	Sensitivity	Specificity	Precision	F1-Score	CC
Decision Tree	82.11	82.11	91.05	82.25	82.14	73.23
Naive Bayes	81.19	81.19	90.59	81.08	80.94	71.80
mSVM	86.84	86.84	93.42	86.90	86.80	80.31
KNN	79.53	79.52	89.76	80.59	78.54	70.03
NN	84.83	84.83	92.42	84.79	84.79	77.24

Table 4 Quadruple Class Classification Performance Values

Machine Learning Algorithms for Classifier	Performances					
	Accuracy	Sensitivity	Specificity	Precision	F1-Score	CC
Decision Tree	74.80	74.80	91.60	74.88	74.82	66.43
Naive Bayes	71.60	71.60	90.53	71.64	71.47	62.15
mSVM	79.87	79.87	93.29	80.20	79.94	73.30
KNN	70.17	70.17	90.06	71.99	69.30	60.93
NN	76.43	76.43	92.14	76.44	76.43	68.58

Table 5 Binary classification performance with 10 folds cross-validation with selected features

Machine Learning Algorithms for Classifier	Performances						
	Number of Class	Accuracy	Sensivity	Spesificity	Precision	F1-Score	CC
ACO + DT	2	93.94	93.46	94.42	94.37	93.91	87.89
ACO + NB		92.68	90.93	94.42	94.22	92.55	85.41
ACO + kNN		94.05	92.42	95.69	95.54	93.95	88.15
ACO + NN		94.05	93.23	94.87	94.78	94.00	88.12
ACO + SVM		**95.76**	**93.38**	**98.14**	**98.05**	**95.66**	**91.63**
PSO + DT		93.61	92.42	94.80	94.67	93.53	87.24
PSO + NB		93.75	91.23	96.28	96.08	93.59	87.62
PSO + kNN		94.61	93.16	96.06	95.94	94.53	89.26
PSO + NN		94.94	94.20	95.69	95.62	94.91	89.90
PSO + SVM		**95.46**	**93.98**	**96.95**	**96.86**	**95.40**	**90.97**
DE + DT		94.28	92.12	96.43	96.27	94.15	88.63
DE + NB		94.13	91.82	96.43	96.26	93.99	88.35
DE + kNN		95.35	93.31	97.40	97.29	95.26	90.78
DE + NN		95.43	94.72	96.13	96.08	95.39	90.86
DE + SVM		**95.83**	**93.31**	**98.36**	**98.28**	**95.73**	**91.79**

3.2.3 Quadruple Classification

The third experiment was performed for quadruple classification among the Covid-19, Normal, Viral Pneumonia, and Lung Opacifications x-ray image datasets. It consists of 1345 Normal images, 1345 COVID-19 images, 1345 Viral Pneumonia images, and 1345 Lung Opacifications images. Feature vectors (5380×770),

(5380×554) and (5380×478) were created for DE, ACO and PSO, respectively. This feature was classified with the mSVM classifier and its classification performances were measured. The classification performances of the data formed with the selected features are given in Tab. 6.

Fig. 3 shows the radar graph of Accuracy, Sensitivity, Specificity, Precision, F1-Score and Correlation coefficient properties according to the selected features.

Table 6 Triple classification performance with 10 folds cross-validation with selected features

Machine Learning Algorithms for Classifier	Performances						
	Number of Class	Accuracy	Sensivity	Spesificity	Precision	F1-Score	CC
ACO + DT	3	81.24	81.24	90.62	81.23	80.23	71.86
ACO + NB		73.46	73.46	86.73	73.31	7.05	60.22
ACO + kNN		74.20	74.20	87.10	73.81	73.49	61.31
ACO + NN		79.53	79.53	89.76	79.46	79.48	69.27
ACO + mSVM		**84.39**	**84.39**	**92.19**	**84.53**	**84.39**	**76.66**
PSO + DT		79.50	79.50	89.75	79.60	79.49	69.31
PSO + NB		77.32	77.32	88.66	77.25	76.94	66.05
PSO + kNN		76.80	76.80	88.40	76.58	76.04	65.31
PSO + NN		79.90	79.90	89.95	79.87	79.88	69.84
PSO + mSVM		**84.83**	**84.83**	**92.42**	**84.9**	**84.82**	**77.29**
DE + DT		81.69	81.69	90.84	81.73	81.70	72.55
DE + NB		78.74	78.74	89.37	78.65	78.46	68.13
DE + kNN		80.20	80.20	90.10	80.29	79.61	70.52
DE + NN		83.54	83.54	91.77	83.47	83.50	75.29
DE + mSVM		**86.12**	**86.12**	**93.06**	**86.16**	**86.07**	**79.22**

Table 7 Quadruple classification performance with 10 folds cross-validation with selected features

Machine Learning Algorithms for Classifier	Performances						
	Number of Class	Accuracy	Sensivity	Spesificity	Precision	F1-Score	CC
ACO + DT	4	72.73	72.73	90.91	72.84	72.76	63.69
ACO + NB		63.25	63.25	87.75	63.12	62.90	50.95
ACO + kNN		61.54	61.54	87.18	61.92	60.63	40.85
ACO + NN		69.00	69.00	89.67	68.96	68.96	58.65
ACO + mSVM		**74.28**	**74.28**	**91.43**	**74.63**	**74.35**	**65.86**
PSO + DT		71.04	71.04	90.35	71.31	71.08	61.51
PSO + NB		66.84	66.84	88.95	66.85	66.54	55.79
PSO + kNN		64.26	64.26	88.09	65.25	63.37	52.72
PSO + NN		70.13	70.13	90.04	70.13	70.12	60.17
PSO + mSVM		**77.08**	**77.08**	**92.36**	**77.38**	**77.16**	**69.58**
DE + DT		75.07	75.07	91.69	75.26	75.13	66.85
DE + NB		69.37	69.37	89.79	69.38	69.16	59.16
DE + kNN		68.42	68.42	89.47	69.80	67.60	58.44
DE + NN		73.96	73.96	91.32	74.00	93.97	65.30
DE + mSVM		**79.67**	**79.67**	**93.22**	**79.93**	**79.73**	**73.01**

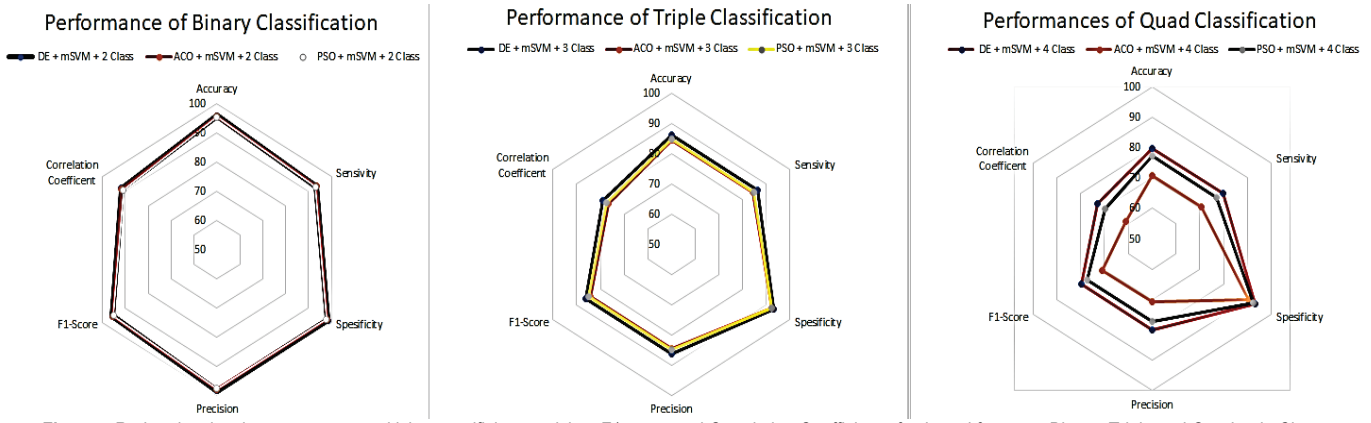


Figure 5 Radar plot showing accuracy, sensitivity, specificity, precision, F1-score and Correlation Coefficient of selected features, Binary, Triple and Quadruple Class

3.3 Comparative Analysis Based on Accuracy Measures of Each Classifier

In this section, the graphs of the accuracy values of the binary, triple and quadruple classification results of the selected features by using the classification operations and optimization algorithms using the features obtained by deep feature extraction are given in Figs. 6-8.

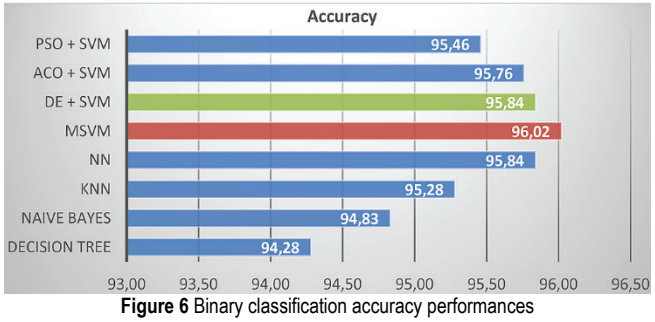


Figure 6 Binary classification accuracy performances

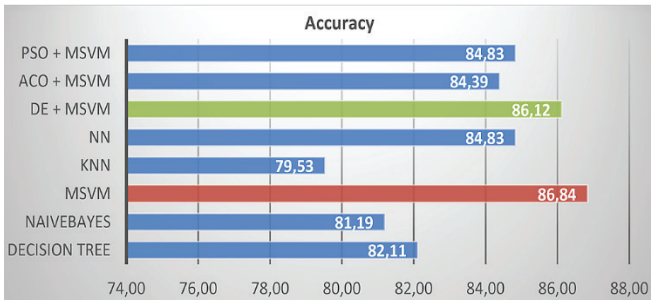


Figure 7 Triple class classification accuracy performances

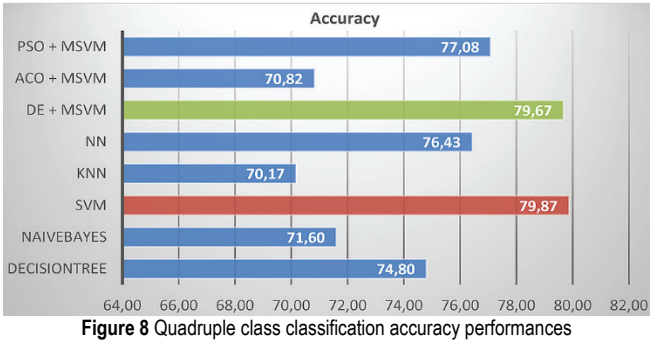


Figure 8 Quadruple class classification accuracy performances

The classification of the features obtained by extracting the deep features and the confusion matrices of the 2nd, 3rd and 4th class best performance are given Fig. 9.

The confusion matrices of the performance of the best DE algorithm as a result of the classification of the features obtained by the extraction of deep features and the features obtained by the feature selection of the optimization algorithms classified as 2, 3 and 4 classes are given Fig. 10.

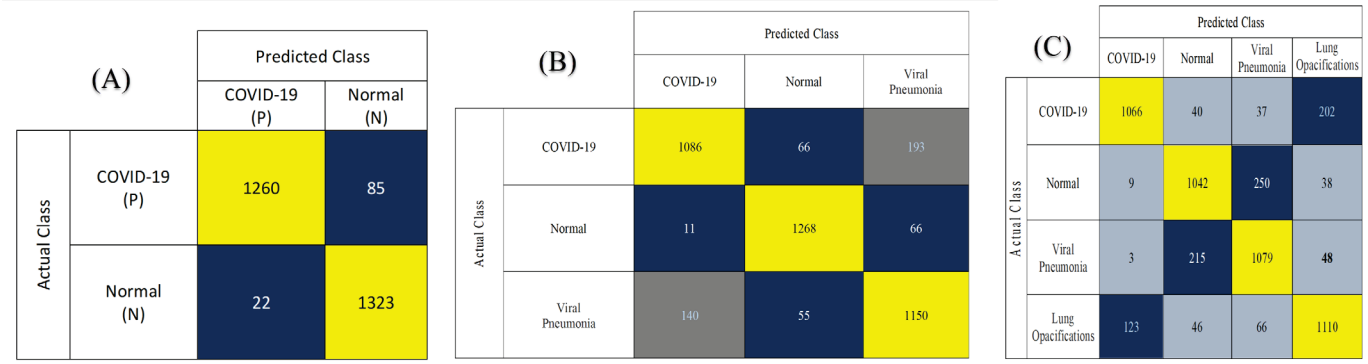


Figure 9 Confusion matrices of the (A)-2nd, (B)-3rd and (C)-4th class best performance

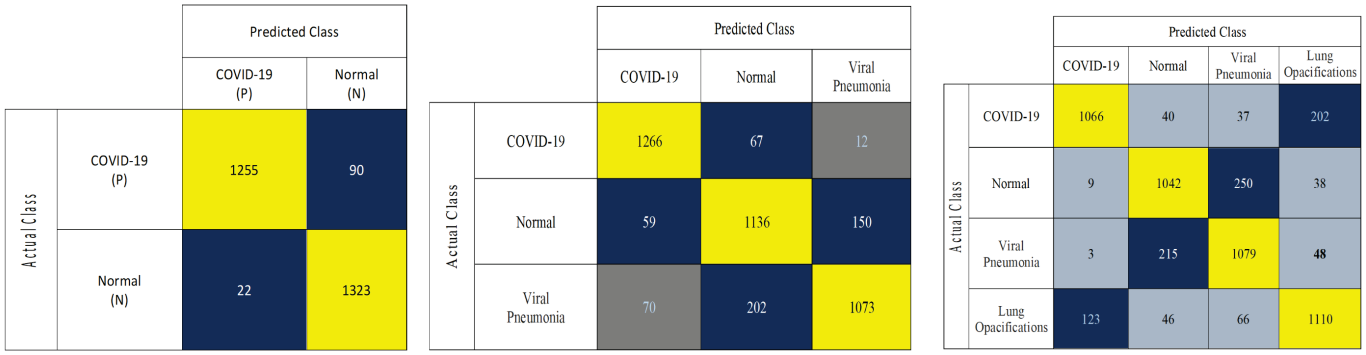


Figure 10 Confusion matrices of the (A)-2nd, (B)-3rd and (C)-4th class best performance for DE Feature Selection

4 DISCUSSION

When the recent literature studies on X-ray images in the diagnosis of COVID-19 are examined, it is seen that deep CNN is one of the most preferred techniques [13]. While some of these studies have dual classifications, some have multiple classifications. When the studies are examined, it is seen that there are different numbers of chest X-rays and different numbers of classification. Tab. 8 contains the latest studies on chest X-ray in the literature.

5 CONCLUSION

Early and rapid diagnosis of Covid-19 is very important to prevent the spread of the disease to other people and to prevent the spread of the pandemic. Separating Covid-19 and

respiratory diseases will also eliminate the anxiety that may occur in people.

In this study, binary, triple and quadruple computer-assisted diagnosis was made using deep CNN approach from chest x-rays (COVID-19, Normal), (COVID-19, Viral Pneumonia, Normal) and (COVID-19, Viral Pneumonia, Lung Opacifications, Normal).

Table 8 Previous studies on the issue

References	Model	Dataset	Accuracy (%)
Bhattacharyya et al. [39]	Conditional generative adversarial network (C-GAN) model has been proposed.	C:342 N:341 P:347	96.6
Aziz et al. [1]	Computer-aided diagnosis of COVID-19 disease from chest x-ray images ResNET50v2 model has been proposed.	C:254 N:310 P:310	99.5% for binary 95.5% for triple
Mahmud et al. [40]	CNN-base+A3:E42d CovXNet model has been proposed.	VP: 305 BP: 305 N: 305 C: 305	90.3
Ismael et al. [7]	ResNet50 model and SVM classifier with the Linear kernel function model has been proposed.	N: 200 C: 180	94.7
Ouchicha et al. [41]	CNN-based CVDNet model has been proposed.	C: 219 N: 1341 VP: 1345	97.20
Khan et al. [42]	CoroNet model has been proposed by using Deep Convolutional Neural Networks (DCNN)-based Xception architecture.	N: 310 BP: 330 VP: 327 C: 284	89.6
Ozturk et al. [43]	CNN-based DarkCOVIDNet model has been proposed.	C:125 P: 500 NoFindings: 500	87.02
Our Study	CNN-based SequeezeNet + ML Feature Selection (DE, ACO, PSO) +ML	C:1345 P:1345 VP:1345 LO:1345	96.2% for binary 86.84% for triple 79.87% for quadruple - (DE) 95.76% for binary 84.83% for triple 79.67% for quadruple

More specifically, deep features were extracted with pre-trained Deep CNN, and machine learning techniques were used as classifiers for these extracted features classification. Afterwards, these features were selected with optimization algorithms and fewer feature vectors were obtained, and their performance was measured by classifying them with mSVM, and the following results were obtained:

1) The mSVM method stands out as the best classifier in the binary, triple and quadruple classification process using 1000 features obtained by feature extraction with deep CNN learning. In binary, triple and quadruple classification, accuracy values of 96.02%, 86.84% and 79.87% were calculated, respectively.
2) The features obtained by deep learning were selected using metaheuristic optimization algorithms DE, ACO and PSO. 1000 attributes were reduced to 770 with DE, 554 with ACO, and 478 with PSO.
3) These features obtained in two stages were classified separately with the mSVM classifier. Very good results were obtained with the best accuracy DE + mSVM with 95.83, 86.12 and 79.67, respectively.
4) The classification performance obtained from the classification process with deep feature extraction was achieved by using feature selection in the proposed method, by using less features and by spending less time for training time. Classification of features obtained from feature selection with DE, ACO and PSO was performed with 31.5%, 72.6% and 44.9% less training time, respectively, compared to classification using all features.
5) DE feature selection algorithm and mSVM classifier produced more successful results in deep feature extraction.

6 REFERENCES

[1] Aziz, S., et al. (2021). Computer-aided diagnosis of COVID-19 disease from chest x-ray images integrating deep feature extraction. *Expert Systems*, e12919. https://doi.org/10.1111/exsy.12919

[2] Landete, P., et al. (2020). Clinical features and radiological manifestations of COVID-19 disease. *World Journal of Radiology, 12*(11), p. 247. https://doi.org/10.4329/wjr.v12.i11.247

[3] Kim, H. (2020). Outbreak of novel coronavirus (COVID-19): What is the role of radiologists? *Eur Radio., 30*(6), 3266-3267. https://doi.org/10.1007/s00330-020-06748-2

[4] Yang, W., et al. (2020). The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). *Eur Radio., 30*(6), 4874-4882. https://doi.org/10.1007/s00330-020-06827-4

[5] Moujahid, H., Cherradi, B., Al-Sarem, M., & Bahatti, L. (2021). Diagnosis of COVID-19 Disease Using Convolutional Neural Network Models Based Transfer Learning. In: Saeed, F., Mohammed, F., Al-Nahari, A. (eds) Innovative Systems for Intelligent Health Informatics. *IRICT 2020. Lecture Notes on Data Engineering and Communications Technologies, vol 72*. Springer, Cham. https://doi.org/10.1007/978-3-030-70713-2_16

[6] Chowdhury, M. E. H., et al. (2020). Can AI help in screening viral and COVID-19 pneumonia? *IEEE Access, 8*, 132665-132676. https://doi.org/10.1109/ACCESS.2020.3010287

[7] Ismael, A. M. & Şengür, A. (2021). Deep learning approaches for COVID-19 detection based on chest X-ray images. *Expert Systems with Applications, 164*, p. 114054.

https://doi.org/10.1016/j.eswa.2020.114054
[8] Al-Waisy, A. S., et al. (2020). COVID-CheXNet: hybrid deep learning framework for identifying COVID-19 virus in chest X-rays images. *Soft Computing*, 1-16. https://doi.org/10.1007/s00500-020-05424-3
[9] Selvan, R., et al. (2020). Lung segmentation from chest X-rays using variational data imputation. *arXiv preprint*, arXiv: 2005.10052.
[10] Khan, N. A., Al-Thani, H., & El-Menyar, A. (2022). The emergence of new SARS-CoV-2 variant (Omicron) and increasing calls for COVID-19 vaccine boosters-The debate continues. *Travel Medicine and Infectious Disease, 45*, p. 102246. https://doi.org/10.1016/j.tmaid.2021.102246
[11] Yasar, A., Saritas, I., & Korkmaz, H. (2019). Computer-aided diagnosis system for detection of stomach cancer with image processing techniques. *Journal of Medical Systems, 43*(4), 1-11. https://doi.org/10.1007/s10916-019-1203-y
[12] Rahman, T., et al. (2020). Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray. *Applied Sciences, 10*(9), p. 3233. https://doi.org/10.3390/app10093233
[13] Taspinar, Y. S., Cinar, I., & Koklu, M. (2021). Classification by a stacking model using CNN features for COVID-19 infection diagnosis. *Journal of X-ray Science and Technology*, (Preprint), 1-16. https://doi.org/10.3233/XST-211031
[14] Aslan, M. F., et al. (2021). CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection. *Applied Soft Computing, 98*, p. 106912. https://doi.org/10.1016/j.asoc.2020.106912
[15] Arpaci, I., et al. (2021). Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms. Multimedia *Tools and Applications, 80*(8), 11943-11957. https://doi.org/10.1007/s11042-020-10340-7
[16] Jaiswal, A., et al. (2021). Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *Journal of Biomolecular Structure and Dynamics, 39*(15), 5682-5689. https://doi.org/10.1080/07391102.2020.1788642
[17] Narin, A., Kaya, C., & Pamuk, Z. (2021). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, 1-14. https://doi.org/10.1007/s10044-021-00984-y
[18] Saba, T., et al. (2020). Brain tumor detection using fusion of hand crafted and deep learning features. *Cognitive Systems Research, 59*, 221-230. https://doi.org/10.1016/j.cogsys.2019.09.007
[19] Celik, Y., et al. (2020). Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognition Letters, 133*, 232-239. https://doi.org/10.1016/j.patrec.2020.03.011
[20] Lu, J., et al. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems, 80*, 14-23. https://doi.org/10.1016/j.knosys.2015.01.010
[21] Sivaraman, K. & Murthy, A. (2018). Object recognition under lighting variations using pre-trained networks. In *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. https://doi.org/10.1109/AIPR.2018.8707399
[22] Al-Tashi, Q., et al. (2020). Approaches to multi-objective feature selection: A systematic literature review. *IEEE Access, 8*, 125076-125096. https://doi.org/10.1109/ACCESS.2020.3007291
[23] Opara, K. R. & Arabas, J. ((2019). Differential Evolution: A survey of theoretical analyses. *Swarm and Evolutionary Computation, 44*, 546-558. https://doi.org/10.1016/j.swevo.2018.06.010
[24] Salehpour, M., et al. (2017). A new adaptive differential evolution optimization algorithm based on fuzzy inference system. *Engineering Science and Technology, 20*(2), 587-597. https://doi.org/10.1016/j.jestch.2017.01.004
[25] Karaboğa, D. & Ökdem, S. (2004). A simple and global optimization algorithm for engineering problems: differential evolution algorithm. *Turkish Journal of Electrical Engineering & Computer Sciences, 12*(1), 53-60.
[26] Dorigo, M. (1992). Ant Colony Optimization for vehicle routing problem. *Doctoral dissertation*, Politecnico di Milano, Milan, Italy.
[27] Blum, C. (2005). Ant colony optimization: Introduction and recent trends. *Physics of Life Reviews, 2*(4), 353-373. https://doi.org/10.1016/j.plrev.2005.10.001
[28] Shi, Y. (2001). Particle swarm optimization: developments, applications and resources. In Proceedings of the *2001 IEEE Congress on Evolutionary Computation* (IEEE Cat. No. 01TH8546). https://doi.org/10.1109/CEC.2001.934377
[29] Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access, 6*, 29637-29647. https://doi.org/10.1109/ACCESS.2018.2843443
[30] Yan, J., et al. (2020). A hybrid scheme-based one-vs-all decision trees for multi-class classification tasks. *Knowledge-Based Systems, 198*, p. 105922. https://doi.org/10.1016/j.knosys.2020.105922
[31] Ozkan, I. A., Koklu, M., & Sert, I. U. (2018). Diagnosis of urinary tract infection based on artificial intelligence methods. *Computer Methods and Programs in Biomedicine, 166*, 51-59. https://doi.org/10.1016/j.cmpb.2018.10.007
[32] Siar, M. & Teshnehlab, M. (2019). Brain tumor detection using deep neural network and machine learning algorithm. In *The 9th IEEE International Conference on Computer and Knowledge Engineering (ICCKE)*. https://doi.org/10.1109/ICCKE48569.2019.8964846
[33] Saritas, M. M. & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering, 7*(2), 88-91. https://doi.org/10.18201/ijisae.2019252786
[34] Balaji, V., et al. (2020). Skin disease detection and segmentation using dynamic graph cut algorithm and classification through Naive Bayes classifier. *Measurement, 163*, p. 107922. https://doi.org/10.1016/j.measurement.2020.107922
[35] Xing, W. & Bei, Y. (2019). Medical health big data classification based on KNN classification algorithm. *IEEE Access, 8*, 28808-28819. https://doi.org/10.1109/ACCESS.2019.2955754
[36] Koklu, M. & Sabanci, K. (2016). Estimation of credit card customers payment status by using kNN and MLP. *International Journal of Intelligent Systems and Applications in Engineering*, 249-251. https://doi.org/10.18201/ijisae.2016SpecialIssue-146983
[37] Rasekh, M., et al. (2021). Performance analysis of mau-9 electronic-nose MOS sensor array components and ANN classification methods for discrimination of herb and fruit essential oils. *Chemosensors, 9*(9), p. 243. https://doi.org/10.3390/chemosensors9090243
[38] Koul, A., Becchio, C., & Cavallo, A. (2018). Cross-validation approaches for replicability in psychology. *Frontiers in Psychology, 9*, p. 1117. https://doi.org/10.3389/fpsyg.2018.01117
[39] Bhattacharyya, A., et al. (2022). A deep learning based approach for automatic detection of COVID-19 cases using chest X-ray images. *Biomedical Signal Processing and Control, 71*, p. 103182. https://doi.org/10.1016/j.bspc.2021.103182
[40] Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature

optimization. *Computers in Biology and Medicine, 122*, p. 103869. https://doi.org/10.1016/j.compbiomed.2020.103869

[41] Ouchicha, C., Ammor, O., & Meknassi, M. (2020). CVDNet: A novel deep learning architecture for detection of coronavirus (Covid-19) from chest x-ray images. *Chaos, Solitons & Fractals, 140*, p. 110245. https://doi.org/10.1016/j.chaos.2020.110245

[42] Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine, 196*, p. 105581. https://doi.org/10.1016/j.cmpb.2020.105581

[43] Ozturk, T., et al. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine, 121*, p. 103792. https://doi.org/10.1016/j.compbiomed.2020.103792

Author's contacts:

Ali Yasar
Department of Mechatronics Engineering,
Selcuk University,
Yeni Istanbul Caddesi 363,
42130 Konya/Ardıclı, Türkiye
+905055204696
ORCID: 0000-0001-9012-7950
aliyasar@selcuk.edu.tr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220920172008

Original scientific paper
Received: 2022-09-20, Accepted: 2022-11-02

Unbalance Response Analysis of a Rotor Kit with Two Identical Discs Located Between Bearings

Marko Katinić*, Josip Kolar, Pejo Konjatić, Mladen Bošnjaković

Abstract: In this paper, the dynamic behavior of a rotor kit with two identical disks located between the plain bearings was analyzed. Modal and harmonic analysis of this rotor kit configuration were performed in the Ansys software package. To calibrate the bearing parameters (stiffness and damping) in the numerical model, experimental measurements of the rotor kit with a disc mounted at the midspan of the shaft were performed. As a result of modal analysis, natural frequencies and models were obtained. Using the Campbell's diagram, the critical speeds and the influence of the gyroscopic effects on the natural frequencies were determined. The responses of the rotor kit to different unbalance distributions were considered by harmonic analysis.

Keywords: Campbell diagram; critical speed; harmonic analysis; modal analysis; rotor kit; unbalance

1 INTRODUCTION

The vibration of a rotating machine occurs as a result of dynamic forces. Vibration usually appears as radial (lateral), axial and torsional. Radial vibration causes the most common machine problem, and is therefore commonly measured. Axial vibration occurs only in some special machine operating conditions (for example, compressor surge). Measuring torsional vibration is quite difficult, so this vibration is usually neglected.

Machine vibration causes cyclic stresses in machine components, which can result in high-cycle fatigue failure and forced shutdown of the machine. If the displacements due to vibration are large enough, it can cause unwanted contact between rotating and stationary machine parts, causing wear or damage.

The most common dynamic force that produces radial vibration in rotating machines is unbalance. The non-uniform distribution of the rotor mass during the rotation of the rotor creates a centrifugal force. The rotating centrifugal force has a frequency equal to the rotor speed and causes a synchronous vibration of the rotor. The vibration amplitude increases when the rotor speed approaches the value of the rotor's natural frequency. When the rotor speed becomes equal to the natural frequency, the rotor has reached a balance resonance which is known as the critical speed or critical [1]. At the critical speed, the vibration amplitude reaches its maximum value.

Modern rotating machines, such as gas and steam turbines, centrifugal compressors, have high operating speeds and vibration problems are much more pronounced than for other machines. Some of these machines operate below the first critical, while most fairly large machines in the process industry operate above the first or even more critical speeds. This means that it is important to consider the amplification of rotor vibration in resonance during the start-up or shutdown of the machine. High vibration at resonance results in high rotor stresses, danger of rotor-to-stator contact and wear of seals and bearings [1-3].

Because of the above mentioned, the main areas of consideration in rotary machine design are critical speeds and unbalance response. As a first step in rotor design, an analysis is performed to determine critical speeds and mode shapes. Then the analysis of unbalance response is carried out. This analysis provides an answer as to whether an unbalance response problem can be expected for a particular mode. For this purpose, various commercial software packages (Abaqus, Ansys, etc.) or specialized programs for rotor dynamics (Dynamics R4, MADYN 2000, DeRoBeS, etc.) are used.

To achieve a better understanding of rotor dynamic behaviour, experimental research is often carried out on a rotor kit that is specially designed to simulate behaviour of the actual rotor [4-8]. Investigations of unbalance response were mainly carried out on a single-disc rotor kit. However, most rotors have two or more discs or impellers mounted on the rotor shaft. Therefore, it is of interest to investigate the unbalance response of a rotor kit with two or more discs mounted on a shaft.

In this paper, the dynamic behavior of a rotor kit with two identical discs located between the plain bearings was analyzed. The analysis was performed by the Ansys software package [9]. Bearing parameters used in the numerical model were determined using data obtained from experimental measurements of the response of a rotor kit with a disc mounted at the midspan of the shaft. By modal analysis of a rotor kit with two discs, natural frequencies and mode shapes were obtained. The critical speeds of the rotor kit were determined using Campbell's diagram. Finally, a harmonic analysis was performed to determine the response to unbalance. The responses to different unbalance distributions were considered.

2 MODAL AND HARMONIC ANALYSIS

The analyzed two-disc rotor kit is shown in Fig. 1 [10]. Each rotor disc has 16 threaded holes for mounting M5 screws of different masses. The holes are located equidistantly on a disc radius of 30 mm.

The material of the rotor kit is steel with a density of 7850 kg/m^3. Its modulus of elasticity is 210 GPa and Poisson's ratio is 0.3.

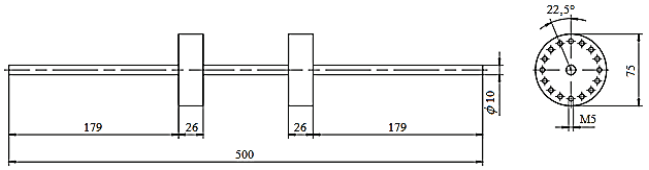


Figure 1 Rotor kit with two discs [10]

The rotor shaft is supported by two plain bearings located at the ends of the shaft. The bearing parameters (stiffness and damping) for the numerical model of the rotor kit were determined experimentally using a rotor kit with a disc located at the midspan of the shaft.

2.1 Determination of Bearing Parameters

The dimensions of the rotor kit with a disc are shown in Fig. 2.

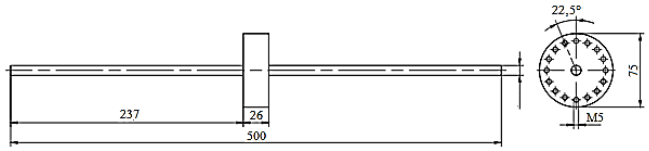


Figure 2 Rotor kit with a disc [10]

Fig. 3 shows the experimental measuring scheme of the rotor kit. Labels *r1x*, *r1y*, *r2x*, *r2y* indicate proximity vibration probes in the *x* and *y* directions for the first and second measuring location, respectively. Bearing stiffness constants in the *x* and *y* directions are denoted by K_{bx} and K_{by}. C_{bx} and C_{by} denote the damping constants of the bearing in the *x* and *y* directions.

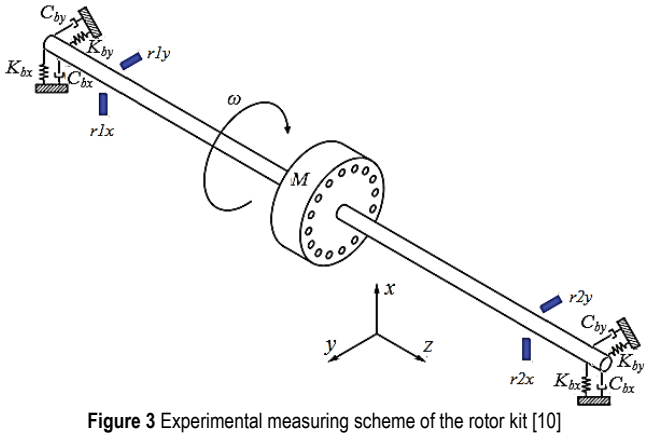


Figure 3 Experimental measuring scheme of the rotor kit [10]

Experimental vibration measurements were performed on the RK4 produced by Bently Nevada [4] (Fig. 4). The measurements were carried out using a specialized system for data acquisition and processing, which is a product of Turbocom company from Croatia. At two measuring points, which are at a distance of 65 mm from the centre of each bearing (to avoid the occurrence of a nodal point), proximity probes of the Bently Nevada system 3300 – 8 mm were installed. These probes measure the peak-to-peak (pp) vibration amplitude in µm. The measurements were carried out in a transient condition, by simulating the start-up of the rotor from 0 rpm to 10000 rpm. The rotor speed was increased by constant angular acceleration.

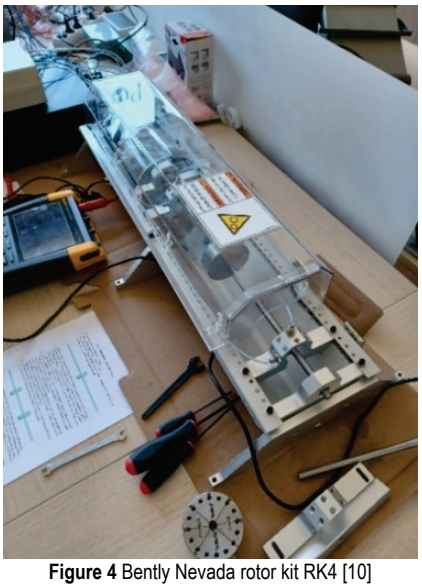


Figure 4 Bently Nevada rotor kit RK4 [10]

Fig. 5 shows the obtained Bode plot of the start-up, which shows the change in amplitude and phase of synchronous vibration (1X) versus rotor speed.

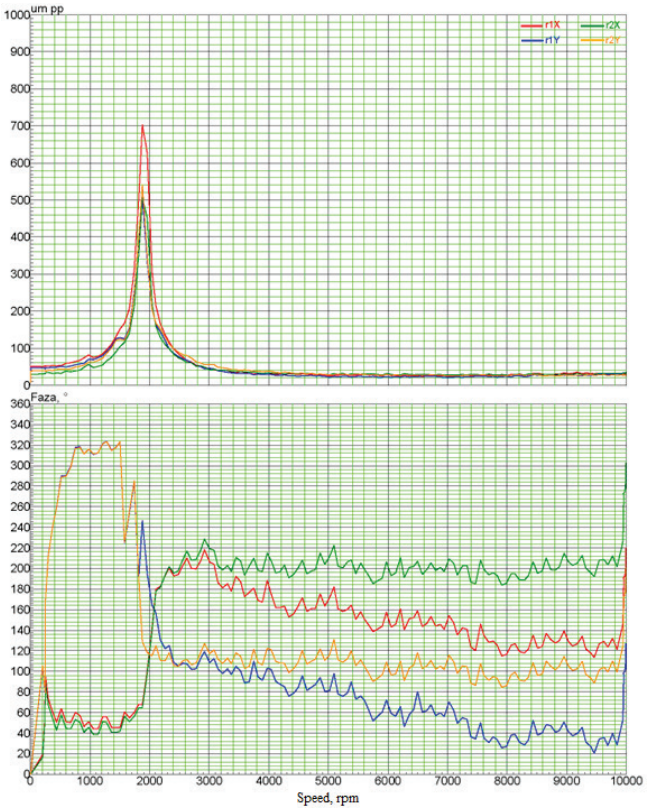


Figure 5 Bode plot [10]

The data from the Bode plot were used to determine the stiffness and damping constants of the bearings. According to [10], the obtained constants are as follows: $K_{bx} = K_{by} = 149$ kN/m and $C_{bx} = C_{by} = 9.445$ Ns/m.

These values of constants were confirmed by the modal analysis of the rotor kit in Ansys software. The geometric model of the rotor-bearing system is shown in Fig. 6, where the bearing model in Ansys is implemented as a mechanical connection type defined at selected locations of the shaft. Bearing modelling using the Bearing Connection option allows defining the stiffness and damping constants of the bearings.

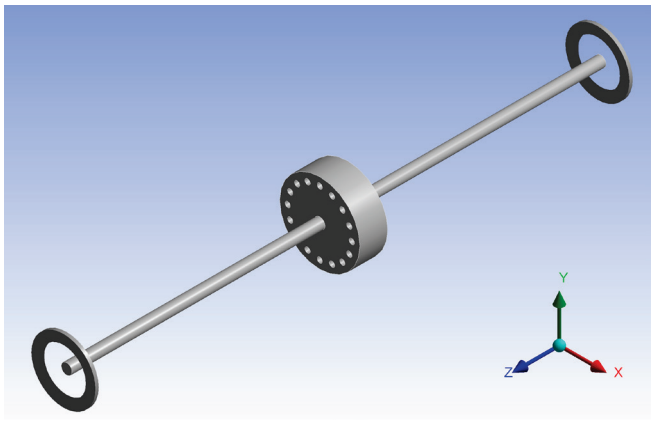


Figure 6 Rotor-bearing system in Ansys [10]

Discretization of the model was performed automatically, with a mesh of finite elements of size 3 mm. The finite element mesh was generated with SOLID 187 finite elements. Fig. 7 shows the meshed model of the rotor kit.

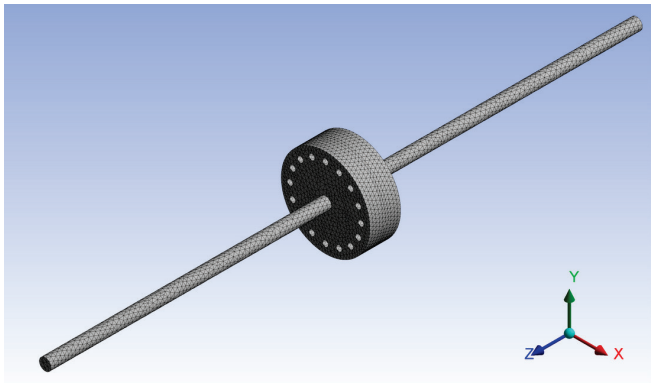


Figure 7 Meshed model of the rotor kit [10]

The boundary conditions are set so that the rotation in relation to the longitudinal axis of the rotor was fixed, and the rotations around the two remaining axes remained free to avoid the so-called "zero pivot" error in the solver. The axial displacement of the rotor is also fixed. The considered rotor speeds from 0 to 10000 rpm with a step of 1000 rpm are defined tabularly in Ansys. In the analysis settings, the first five (5) modes of vibration are defined, and the creation of a Campbell diagram is also included. Natural frequencies are calculated for each considered rotor speed.

Tab. 1 gives the calculated values of the first five natural frequencies for zero rotor speed [10]. Modes 1 and 2 represent the first bending mode with forward (FW) and backward (BW) whirling, respectively. Mode 3 is the first torsional mode, while modes 4 and 5 represent the second bending mode with forward (FW) and backward (BW) whirling, respectively.

Table 1 Natural frequencies of the single-disc rotor kit at zero speed [10]

Mode	Natural frequency, Hz
1	31.66
2	31.67
3	120.67
4	203.85
5	203.85

Fig. 8 shows the Campbell diagram, which is one of the most important engineering tools for determining rotor critical speeds. The diagram shows the dependence of the natural frequencies of the system and the excitation frequencies as a function of the rotor speed. On the abscissa of the diagram is the rotor speed in revolutions per minute (rpm), and on the ordinate is the frequency in Hertz (Hz).

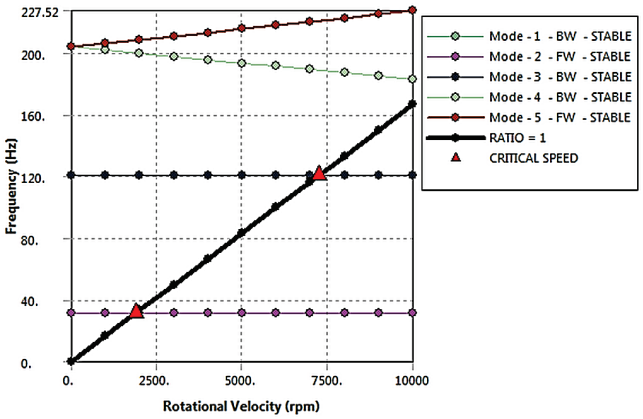


Figure 8 Campbell diagram for the single-disc rotor kit [10]

It is evident from the diagram that the natural frequencies of the first three modes (the first bending with FW and BW and the first torsional) do not change with the rotor speed. This means that there is no influence of the gyroscopic effect on the spring stiffness of the rotor system. However, for the remaining two modes (the second bending with FW and BW), the change of natural frequencies with the rotor speed is obvious. For the second bending mode with FW, the natural frequency increases linearly with increasing rotor speed, which is the result of increasing the stiffness of the system due to the gyroscopic effect. This phenomenon is known as gyroscopic stiffening. On the other hand, for the second bending mode with BW, the natural frequency decreases linearly with increasing rotor speed, which is the result of the reduction of system stiffness due to the gyroscopic effect. This phenomenon is known as gyroscopic softening. In the diagram, one additional straight line can be seen (marked as RATIO = 1), which is called the excitation line. This line corresponds to an excitation with a frequency that coincides with the rotation frequency of the rotor. It is

evident that the excitation line intersects the natural frequency lines of the first bending mode with FW and BW. The point of intersection of these lines determines the critical speed of the rotor, which is 1900 rpm. There is no deviation between the numerically and experimentally determined critical speed. This gives full confidence in the numerical model of the rotor kit. The excitation line also intersects the natural frequency of the first torsional mode, so there is also a torsional critical speed of 7240 rpm.

2.2 Modal Analysis of a Two-Disc Rotor Kit

A modal analysis of the two-disc rotor kit shown in Fig. 2 was performed with the aim of determining the natural frequencies and associated mode shapes. To create a numerical model of the rotor kit, the bearing parameters obtained in chapter 2.1 were used. The geometric model of the rotor-bearing system is shown in Fig. 9.

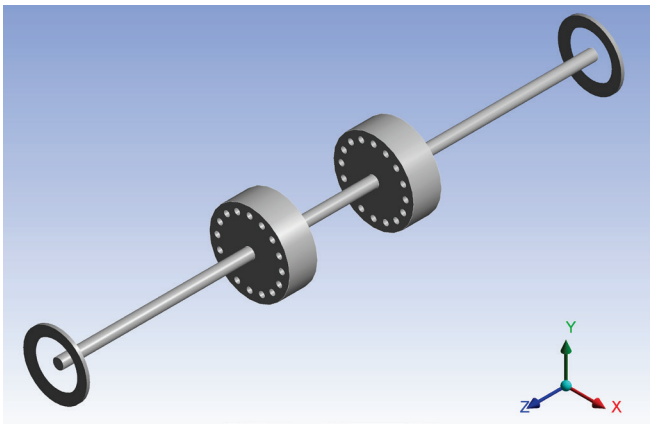


Figure 9 Rotor-bearing system in Ansys [10]

Fig. 10 shows the meshed model of the rotor kit. The finite element mesh was generated automatically using SOLID 187 finite elements of size 3 mm.

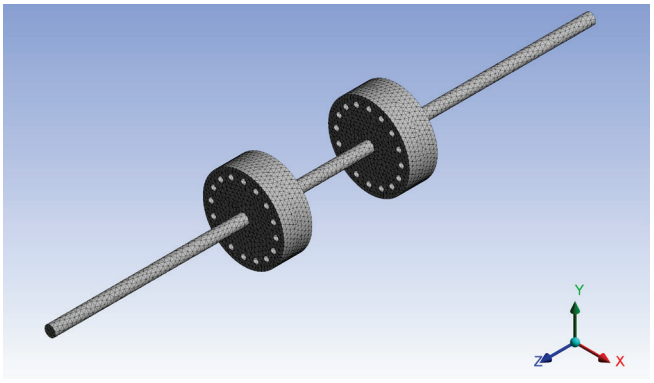


Figure 10 Meshed model of the rotor kit [10]

Identical boundary conditions are set as in the case of single-disc rotors. The considered rotor speeds from 0 to 10000 rpm with a step of 1000 rpm are defined tabularly in Ansys. In the analysis settings, the first 6 modes of vibration are selected. The creation of a Campbell diagram is also included. Natural frequencies are calculated for each considered rotor speed.

The calculated values of natural frequencies for zero rotor speed are given in Tab. 2 [10].

Table 2 Natural frequencies of the two-disc rotor kit at zero speed [10]

Mode	Natural frequency, Hz
1	25.38
2	25.38
3	91.77
4	116.89
5	116.90
6	293.12

The first bending modes with BW and FW are shown in Fig. 11 and Fig. 12, respectively.

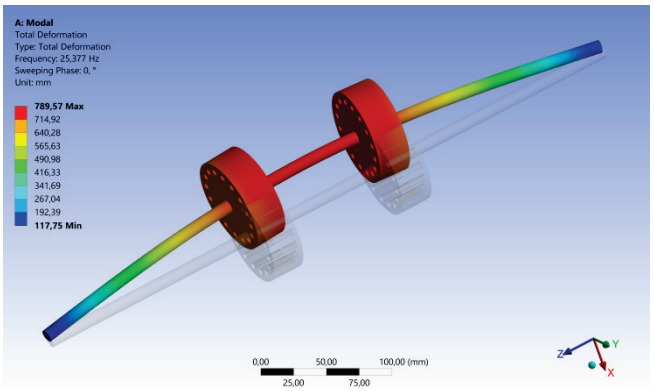


Figure 11 The first bending mode with BW

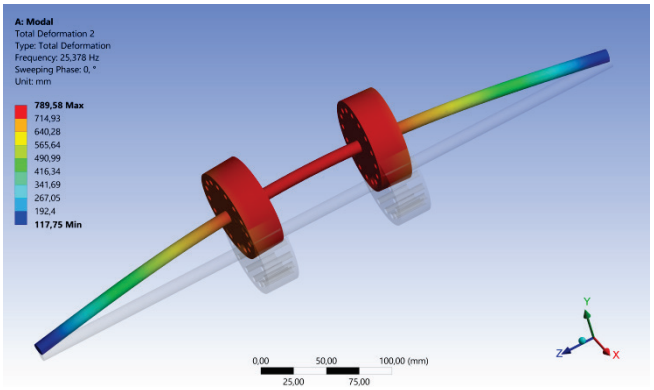


Figure 12 The first bending mode with FW

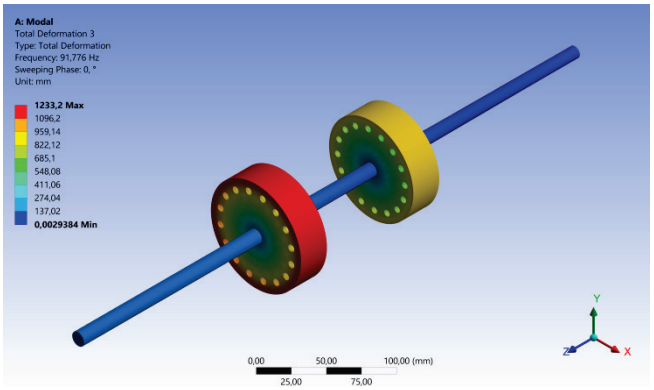


Figure 13 The first torsional mode

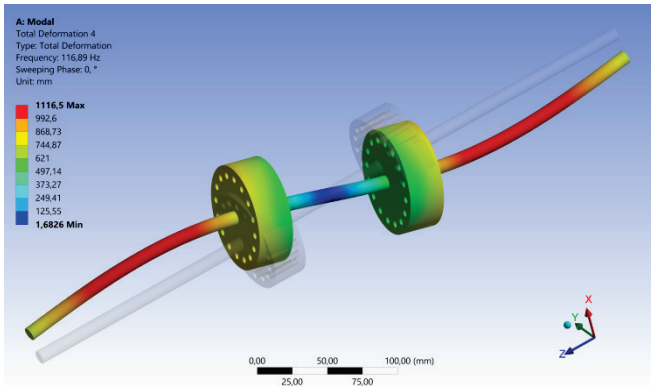


Figure 14 The second bending mode with BW

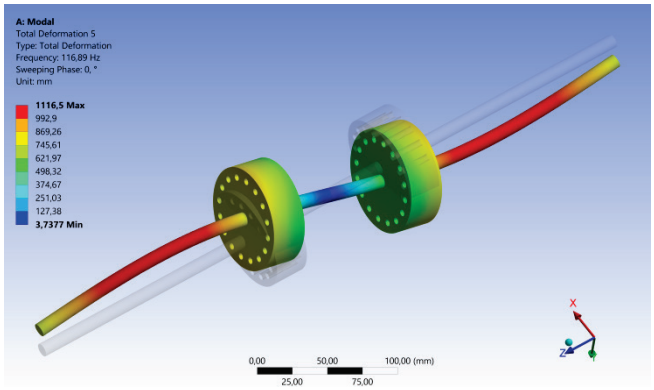


Figure 15 The second bending mode with FW

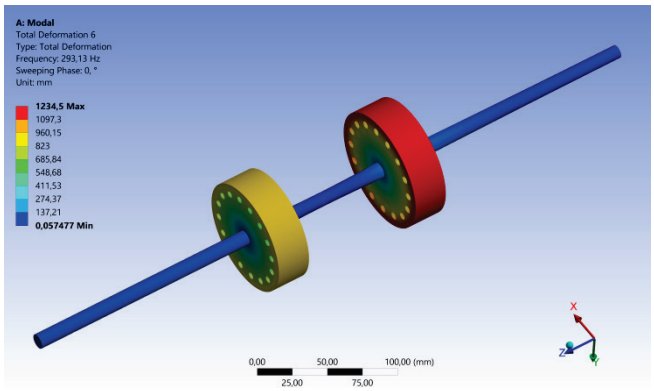


Figure 16 The second torsional mode

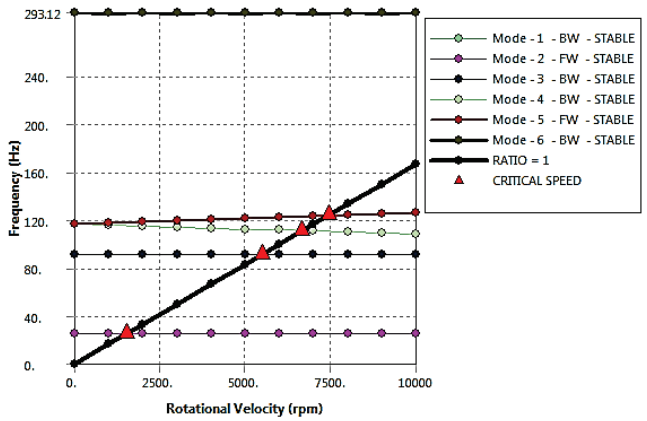


Figure 17 Campbell diagram for the two-disc rotor kit [10]

The first and second torsional modes are shown in Fig. 13 and Fig. 16, respectively.

The second bending modes with BW and FW are shown in Fig. 14 and Fig. 15, respectively.

Campbell diagram is shown in Fig. 17.

It is evident from the diagram that the natural frequencies of the first bending FW and BW, and the first and second torsional modes do not change with the rotor speed. This means that there is no influence of gyroscopic effect on the spring stiffness of the rotor system. However, for the second bending mode with FW and BW, there is a change in natural frequencies with changing rotor speed. For the second bending mode with FW, the natural frequency increases linearly with increasing rotor speed, which is the result of gyroscopic stiffening. On the other hand, for the second bending mode with BW, the natural frequency decreases linearly with increasing rotor speed, which is the result of gyroscopic softening. The excitation line (marked with RATIO = 1) intersects the lines of the first and second bending modes and the first torsional mode. The points of intersection of these lines indicate the critical speeds of the rotor. The intersection point with the natural frequency lines of the first bending mode (FW and BW lines) determines the first critical speed of the rotor, which is 1525 rpm. Since the change of the natural frequency of the second bending mode is affected by the gyroscopic effect, the excitation line intersects the natural frequency lines of this mode (FW and BW lines) at two different points, which represent the second critical speed of the rotor. In the case of FW, the second critical speed is 7436 rpm, and in the case of BW, it is 6656 rpm. Torsional critical speed is 5506 rpm.

2.3 Harmonic Analysis of a Two-Disc Rotor Kit

Actual rotor systems in operation are subjected to forced vibration. Rotating unbalance is the most common form of rotor excitation. The distribution of rotating unbalance creates an axially distributed force system that can excite the natural frequencies. This force system has a form that is conceptually similar to a mode shape. How much a natural mode will be excited depends on the unbalance distribution and how well it matches a particular mode shape. Good unbalance matching results in relatively high mode excitation (a rotor resonance). A poor unbalance match produces little or no mode excitation, and little or no resonance [1].

Using the example of the described rotor kit with two discs (Fig. 1), a harmonic analysis was performed with the aim of determining the influence of different axial distributions of unbalance on the excitation of vibration modes. Unbalance is simulated by adding concentrated masses to the rotor discs at a radius of 30 mm. The analysis was performed using two masses of 0.8 g. Three different harmonic analyses were performed. The first analysis consisted in determining the response of the rotor kit to the in-phase unbalance distribution (Fig. 18). Namely, a mass of 0.8 g was added to each disc and the added masses are in phase (at the same angle). The response of the rotor kit to the out-of-phase unbalance distribution was found in the second analysis (Fig. 19). A mass of 0.8 g is added to each disc, but

the added masses are out-of-phase (the phase difference is 180°). In the third analysis, the response of the rotor kit to the asymmetric distribution of the imbalance, which was achieved by adding a single mass of 0.8 g to only one disc, was investigated. In all analyses, it was chosen that the unbalance has the same direction of rotation as the rotor kit.

The finite element mesh and boundary conditions are identical as in modal analysis. The analysis was performed using the direct integration method [9]. The frequency response was taken on the rotor discs. The selected frequency range for considering the response is 0 to 170 Hz.

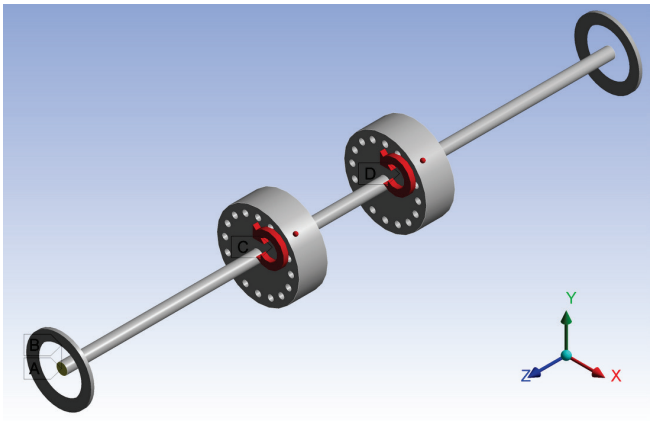


Figure 18 The in-phase unbalance distribution [10]

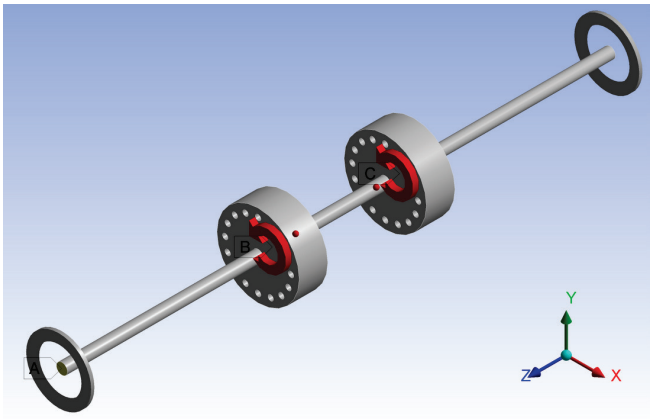


Figure 19 The out-of-phase unbalance distribution [10]

The frequency response is displayed using a Bode plot, which consists of two separate diagrams. The first diagram shows the dependence of the amplitude of the response on the frequency, and the second diagram shows the dependence of the phase of the response on the frequency. Figs. 20 and 21 show the frequency response plots for the *x* and *y* axis, respectively, when the unbalance distribution is in-phase.

The peak of the vibration amplitude at a frequency of 25 Hz is clearly visible on both plots. At this frequency, the phase angle of the response changes by 180°, and the frequency 25 Hz is obviously the resonant frequency. At this frequency, the amplitude is limited only by the damping of the system. From the modal analysis, it is evident that it is the natural frequency of the first bending mode with FW. It means the in-phase unbalance distribution excites only the first mode, because it matches the first mode. The second mode is not excited because the unbalance distribution does not match the second mode shape.

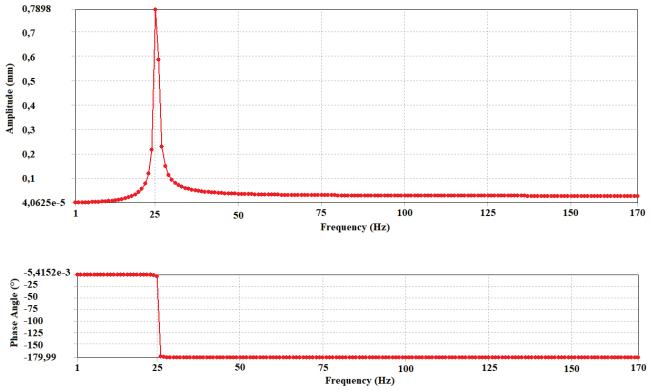


Figure 20 Response to in-phase unbalance distribution for the *x* axis [10]

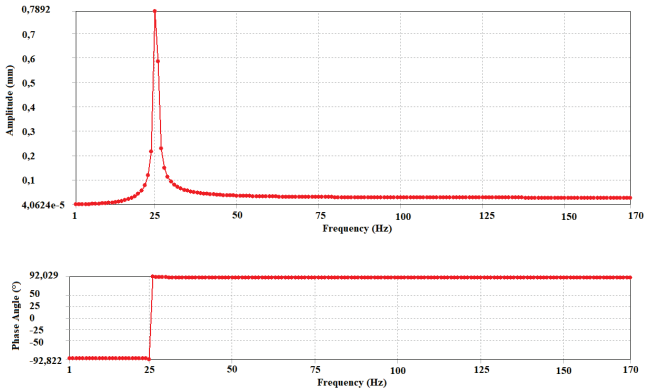


Figure 21 Response to in-phase unbalance distribution for the *y* axis [10]

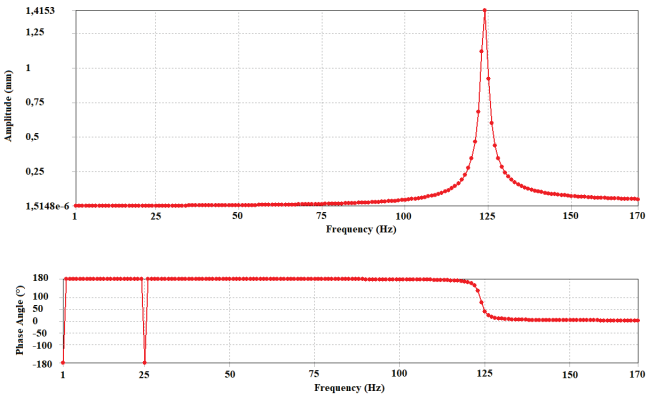


Figure 22 Response to out-of-phase unbalance distribution for the *x* axis [10]

Figs. 22 and 23 show the frequency response plots for the *x* and *y* axis, respectively, when the unbalance distribution is out-of-phase. The peak of the vibration amplitude at a frequency of 124 Hz is clearly visible on both plots. This frequency is obviously the resonant frequency. From the modal analysis, it is evident that it is the natural frequency of the second bending mode with FW. It means the out-of-phase unbalance distribution excites only the second mode. The first mode is not excited because the unbalance distribution does not fit the first mode shape.

In the general case of unbalance distribution (asymmetric unbalance distribution) both bending modes of

vibration are excited. This was confirmed by harmonic analysis which obtained the response of a two-disc rotor for an unbalance mass of 0.8 g added to a 30 mm radius of only one disc. Figs. 24 and 25 show the corresponding frequency responses. The frequency response for *x* axis is shown in Fig. 24, and for the *y* axis in Fig. 25.

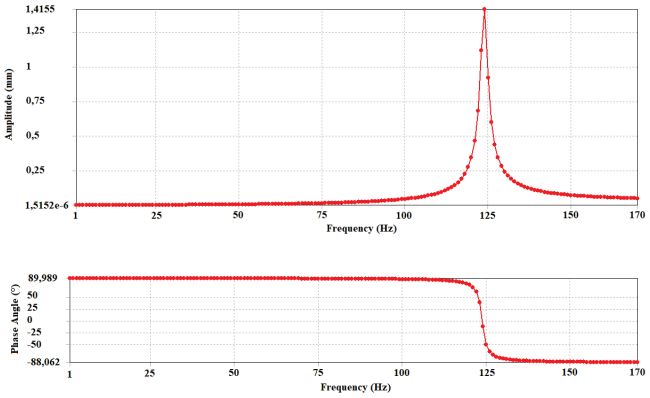


Figure 23 Response to out-of-phase unbalance distribution for the *y* axis [10]

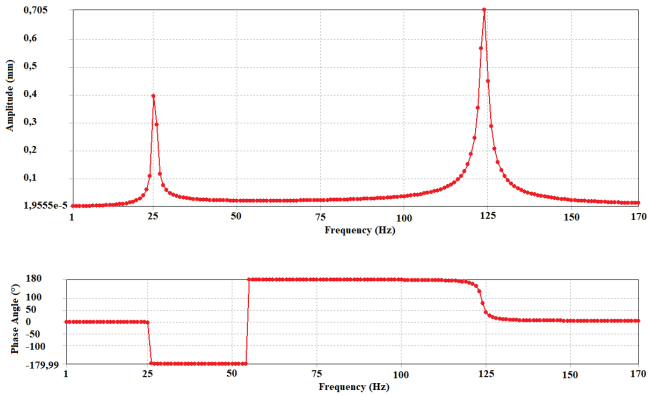


Figure 24 Response to asymmetric unbalance distribution for the *x* axis [10]

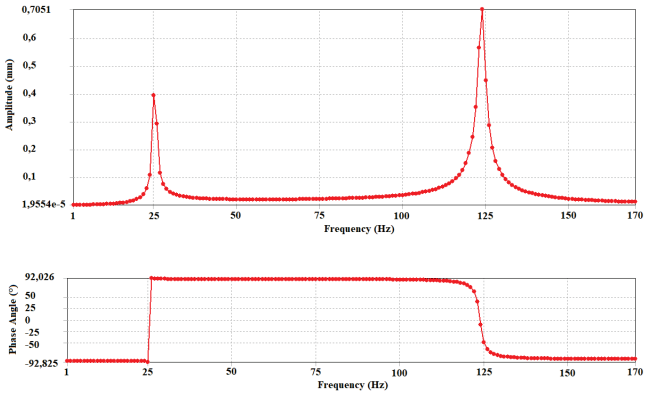


Figure 25 Response to asymmetric unbalance distribution for the *y* axis [10]

As can be clearly seen from both responses, resonant amplitude peaks appear at frequencies of 25 Hz and 124 Hz. In this way, it was proved that the general unbalance, which appears in practical cases, excites both bending modes of vibration.

3 CONCLUSIONS

Modal and harmonic analysis of the two-disc rotor kit was performed. The discs are located between the rotor bearings and at an identical distance from the adjacent bearing. To perform these analyses, a numerical model was created in the Ansys software package. The numerical model was calibrated and verified using the results of experimental measurements on the rotor kit with a disc mounted at the midspan of the rotor shaft.

Natural frequencies, modes and Campbell diagram were obtained by modal analysis. From the Campbell diagram it was concluded that the two-disc rotor kit in the considered speed range has two critical speeds, has the first and second critical speeds for the first and second bending modes, respectively. The first critical speed is the same for both FW and BW of the rotor, because there is no influence of the gyroscopic effect. However, due to the gyroscopic effect, there are two different second critical speeds: a higher value for FW and a lower value for BW of the rotor.

The responses of the two-disc rotor kit to different unbalance distributions were obtained by harmonic analysis. The in-phase unbalance distribution excites only the first mode, because it matches the first mode. The out-of-phase unbalance distribution excites only the second mode, because it fits the second mode. The asymmetric distribution of the unbalance excites both the first and second modes.

4 REFERENCES

[1] Bently, D. E. (2002). *Fundamentals of Rotating Machinery Diagnostics*. Minden, Bently Pressurized Bearing Press

[2] Jacquet-Richardet, G. Torkhani, M., Cartraud, P., Thouverez, F., Nouri-Baranger, T., Herran, M., Gibert, C., Baguet, S., Almeida, P., & Peletan, L. (2013). Rotor to stator contacts in turbomachines. Review and application. *Mechanical Systems and Signal Processing, 40*(2), 401-420. https://doi.org/10.1016/j.ymssp.2013.05.010

[3] Blair, B. J. & Pethybridge, G. (2010). Hydrodynamic Bearing Damage and Remediation of Contributing Factors in Rotating Machinery. *The 9th EDF/Pprime (LMS) Poitiers Workshop "Improvement of Bearing Performance and Evaluation of Adverse Conditions"*, 1-14.

[4] RK4 Rotor Kit, Datasheet, Bently Nevada. Retrieved from https://dam.bakerhughesds.com/m/5af2591c64daec8f/original/RK4-Rotor-Kit-Datasheet-141592-pdf.pdf

[5] Alsaleh, A., Sedighi, H. M., & Ouakad, H. M. (2020). Experimental and theoretical investigations of the lateral vibrations of an unbalanced Jeffcott rotor. *Frontiers of Structural and Civil Engineering, 14*, 1024-1032. https://doi.org/10.1007/s11709-020-0647-y

[6] Tchomeni, B. X. & Alugongo, A. (2019). Experimental diagnosis of multiple faults on a rotor-stator system by fast Fourier transform and wavelet scalogram. *Journal of Vibroengineering, 21*(4), 911-926. https://doi.org/10.21595/jve.2018.19639

[7] Ryu, J. Y. (2020). Diagnostics of Steam Turbine High Vibration from Coupling Failure. *Orbit Newsletter*, Q4. Retrieved from https://dam.bakerhughesds.com/m/6f504c6a0dea307b/original/Q4-2020-KOSPO-RMC-Orbit-Article.pdf?_ga=2.49610852.332631911.1667326481-1286916522.1667326481

[8] Ishibashi, T. & Kawai, T. (2019). Modeling of Rotating Shaft with Partial Rubbing. *Proceedings of the 13th International Modelica Conference*, Regensburg, Germany, 343-352. https://doi.org/10.3384/ecp19157343
[9] Ansys Inc.: ANSYS 17.2
[10] Kolar, J. (2022). Analysis of rotor dynamics for simulation of dynamic behaviour of rotary machines. *Graduate Thesis.* Mechanical Engineering Faculty in Slavonski Brod (in Croatian)
[11] http://www.turbocom.hr/

Authors' contacts:

Marko Katinić, PhD, Associate Professor
(Corresponding author)
Mechanical Engineering Faculty, University of Slavonski Brod,
Trg I. B. Mažuranić 2, 35000 Slavonski Brod, Croatia
035/493417, mkatinic@unisb.hr

Josip Kolar, Mag. Ing. Mech.
Mechanical Engineering Faculty, University of Slavonski Brod,
Trg I. B. Mažuranić 2, 35000 Slavonski Brod, Croatia
0958948995, jkolar@unisb.hr

Pejo Konjatić, PhD, Full Professor
Mechanical Engineering Faculty, University of Slavonski Brod,
Trg I. B. Mažuranić 2, 35000 Slavonski Brod, Croatia
035/493424, pkonjatic@unisb.hr

Mladen Bošnjaković, PhD, Assistant Professor
Technical Department, University of Slavonski Brod,
Trg I. B. Mažuranić 2, 35000 Slavonski Brod, Croatia
035/492634, mbosnjakovic@unisb.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220921070537

Original scientific paper
Received: 2022-09-21, Accepted: 2022-10-17

Approach of Solving Multi-objective Programming Problem by Means of Probability Theory and Uniform Experimental Design

Maosheng Zheng*, Haipeng Teng, Yi Wang

Abstract: In this paper, an approach to deal with the multi-objective programming problem is regulated by means of probability-based multi-objective optimization, discrete uniform experimental design, and sequential algorithm for optimization. The probability-based method for multi-objective optimization is used to conduct conversion of the multi-objective optimization problem into a single-objective optimization one in the viewpoint of probability theory. The discrete uniform experimental design is used to supply an efficient sampling to simplify the conversion. The sequential algorithm for optimization is employed to carry out further optimization. The corresponding treatments reveal the essence of the multi-objective programming, and consideration of the simultaneous optimization of each objective of multi-objective programming problem rationally. Two examples are conducted to illuminate the rationality of the approach.

Keywords: discretization; favorable probability; multi-objective programming; probability theory; sequential algorithm for optimization

1 INTRODUCTION

Multi-objective programming (MOP) is a branch of mathematical programming, which studies the optimization of more than one objective function [1]. The idea of multi-objective programming sprouts in 1776 in the study of utility theory in economics. In 1896, economist Pareto first proposed the multi-objective programming problem in the study of economic balance, and gave a simple idea, which was later called Pareto optimal solution. In 1947, von Neumann and Morgenstern mentioned the multi-objective programming problem in their work on game theory, which attracted much more attentions to this problem. In 1951, Koopmans proposed the multi-objective optimization problem in the activity analysis of production and distribution, and proposed the concept of Pareto optimal solution for the first time. In the same year, Kuhn and Tucker gave the concept of Pareto optimal solution of vector extremum problem from the perspective of mathematical programming. The sufficient and necessary conditions for the existence of this solution are also studied. Debreu's discussion on evaluation equilibrium in 1954 and Harwicz's research on multi-objective optimization problems in topological vector spaces in 1958 laid the foundation for the establishment of this discipline. In 1968, Johnsen published the first monograph on multi-objective decision-making models. Until the 1970s and 1980s, the basic theory of multi-objective programming was finally established after the efforts of many scholars, which made it a new branch of applied mathematics [2].

There are generally the following methods for solving multi-objective programming: one is the method of transforming multiple objectives into a single objective that is easier to solve, such as main objective method, linear weighting method, ideal point method, etc. The other is called hierarchical sequence method, that is, the target is given a sequence according to its importance, and each time the next target optimal solution is found in the previous target optimal solution set, until the common optimal solution is obtained.

The main target method takes a certain $f_1(x)$ as the main target, and the other p–1 are non-main targets. At this time, it is hoped that the main target will reach the maximum value, and the remaining targets should meet certain conditions; the linear weighting method will assign the same weight to the objective functions $f_1(x)$, $f_2(x)$, ..., $f_p(x)$ respectively. The coefficient ω_j, perform a linear weighted sum to obtain a new evaluation function, $U(x) = \sum_{j=1}^{p} \omega_j f_j(x)$, then the multi-objective problem becomes a single-objective problem, but normalization is required when the dimensions are different; for a linear programming problem with multiple objectives, the decision maker hopes to achieve to, these goals in turn under these constraints by means of minimizing the total deviation from the target values, which is the problem to be solved by goal planning [1].

In practical engineering systems, such as many nonlinear, multi-variable, multi-constraint and multi-objective optimization problems in power systems, the existing mathematical methods have limited ability to optimize these problems, and the obtained solutions are not satisfactory [2].

Above discussions indicate that the normalization and the introductions of subjective factors are indispensable treatment in the above "additive" algorithms to transfer diverse criteria into a "unique criterion", and the final result depends on the normalization process significantly [3]. Different normalization methods could result in complete differences in the consequence. Besides, beneficial performance index and unbeneficial performance index are treated in non-equivalent or inconsistent manners in some algorithms. In addition, the "additive" algorithm in the multi-objective optimization is corresponding to the form of "union" from the viewpoint of set theory. So, above algorithms could be seen as a semi-quantitative approach in some sense.

Recently, a probability-based method for multi-

objective optimization (PMOO) was proposed to solve the intrinsic problems of subjective factors in previous multi–objective optimizations [3-5]. A brand new idea of favorable probability was proposed to reflect the favorable degree of a performance index in the optimization in the PMOO. The PMOO aims to treat the simultaneous optimization of multiple objectives in the viewpoint of probability theory. In the novel methodology of PMOO, all performance utility indicators of alternatives are preliminarily divided into two types, i.e., beneficial or unbeneficial types according to their functions and preference in the optimization; each performance utility indicator of an alternative contributes to a partial favorable probability quantitatively. Moreover, the product of all partial favorable probabilities produces the total favorable probability of an alternative, which thus transfers the multi-objective optimization problem into a single-objective optimization one rationally.

In this paper, it regulates a rational approach of multi-objective programming by means of probability theory, discrete uniform experimental design, and sequential algorithm for optimization. Furthermore, examples for illumination of this approach are given.

2 NEW APPROACH OF SOLVING MULTI-OBJECTIVE PROGRAMMING PROBLEM

The rational approach of multi-objective programming is conducted by the combination of probability theory, discrete uniform experimental design, and sequential algorithm for optimization integrally.

The probability-based method for multi-objective optimization is used to conduct conversion of the multi-objective optimization problem into a single-objective optimization one in the viewpoint of probability theory.

The discrete uniform experimental design is used to supply an efficient sampling to simplify the conversion, which is especially important for the goal functions in multi-objective programming problem being continuous functions. Sequential algorithm for optimization is employed to carry out further optimization.

2.1 Probability Theory Based Treatment

In the viewpoint of probability theory, the entire event of appearance of "simultaneous optimization of multi-objective" is corresponding to the product of the each individual objective (event). Therefore, the usual term "the higher the better" for the utility index of performance indicator needs to be expressed quantitatively in term of probability theory, which stimulates us to seek a proper expression for the term "the higher the better" in probability theory quantitatively. A brand new idea of "favorable probability" was proposed in [3-5] to interpret the preference degree of the candidate in the selection, i.e., it uses the term "favorable probability" to characterize the preference degree of the utility index of a performance indicator quantitatively in the optimization.

As to the multi-objective programming problem, each goal is indeed an objective of the PMOO. All performance utility indicators of alternatives are preliminarily divided into two types, i.e., beneficial or unbeneficial types according to their functions and preference in the optimization; thus, the subsequent process of PMOO can be employed rationally.

The assessment of the partial favorable probability P_{ij} for the beneficial index is written as [3-5]

$$P_{ij} = \alpha_j \cdot X_{ij};\ i = 1, 2, ..., n;\ j = 1, 2, ..., m;\ \alpha_j = \frac{1}{n \cdot \overline{X}_j}. \quad (1)$$

While the assessment of the partial favorable probability P_{ij} for the unbeneficial index is expressed as

$$P_{ij} = \beta_j \cdot (X_{j\max} + X_{j\min} - X_{ij});\ i = 1, 2, ..., n;$$
$$j = 1, 2, ..., m;\ \beta_j = \frac{1}{n \cdot \left[(X_{j\max} + X_{j\min}) - \overline{X}_j \right]}. \quad (2)$$

In Eqs. (1) and (2), X_{ij} is the value of utility index of performance indicator; n expresses the number of the performance indicator; m expresses the number of the alternative in the evaluation. $\overline{X}_j$ represents the arithmetic average of the value of utility index of performance indicator X_{ij} over index i for specific j; X_{jmax} and X_{jmin} indicate the maximum and minimum values of X_{ij} over index i for specific j, respectively.

Moreover, the total / overall favorable probability of an alternative is written as

$$P_i = P_{i1} \cdot P_{i2} \cdot ... \cdot P_{im} = \prod_{j=1}^{m} P_{ij}\ . \quad (3)$$

The total / overall favorable probability of an alternative, i.e. Eq. (3), thus transfers the multi-objective optimization problem into a single-objective optimization one in viewpoint of probability theory for the simultaneous optimization of multiple objectives rationally.

2.2 Discrete Uniform Experimental Design and Sequential Algorithm for Optimization

Since the goal functions in multi-objective programming problem are usually continuous ones, discretization can be used to conduct the simplified treatment for the simplicity.

As was stated in [6], the methodologies of good lattice point (GLP) and uniform experimental design (UED) make the discretization possible and practical. The methodologies of GLP and UED are based on number theory, which could supply effective assessment for a definite integral with finite sampling points [6, 7]. The finite sampling points are uniformly distributed within the integral domain with low-discrepancy [8, 9]. The characteristic of the uniformly distributed point set makes the convergence much faster than Monte Carlo sampling [8, 9], which thus has been promising a very good algorithm in approximate calculations with a

surname – "quasi – Monte Carlo Method". Fang specially developed uniform design and uniform design table for the proper using of UED [10]. Sequential uniform design or sequential algorithm for optimization (SNTO) can be used to conduct further optimization for the multi-objective programming problem due to its similarity to problem of multi-objective optimization [6, 8].

Finally, the multi-objective programming problem is conducted by means of the probability - based multi-objective optimization and discrete uniform experimental design straightforward.

3 APPLICATIONS

In this section, two examples are given to illuminate the applications of the regulated approach in solving multi-objective programming problem by means of probability theory and discrete uniform experimental design.

3.1 Production with Maximum Profits and Least Pollutions

A factory produces two kinds of products α and β during the planning period. Each product consumes three different resources, A, B, and C [1]. The unit consumption of resources for each product, the limit of various resources, the unit price, unit profit and unit pollution caused by each product are shown in Tab. 1 [1]. Assume that all products can be sold. Now, the problem is how to arrange production which can maximize profit and output value, and cause the least pollution.

Table 1 Unit consumption of resources, unit profit and pollution for each product

Content	Product		Limit of resource (ton)
	α	β	
Unit consumption of resource A (ton)	9	4	240
Unit consumption of resource B (ton)	4	5	200
Unit consumption of resource C (ton)	3	10	300
Unit price (¥RMB /ton)	400	600	
Unit profit (¥RMB /ton)	70	120	
Unit pollution (CO_2, kg /ton)	3	2	

Solution:

Assume the output of products α and β are x_1 and x_2, respectively, the mathematical model of the problem is as following with s. t. (restraint) conditions,

$\text{Max } f_1(x) = 70x_1 + 120x_2,$

$\text{Max } f_2(x) = 400x_1 + 600x_2,$

$\text{Min } f_3(x) = 3x_1 + 2x_2,$

s. t.

$9x_1 + 4x_2 \le 240,$

$4x_1 + 5x_2 \le 200,$

$3x_1 + 10x_2 \le 300,$

$x_1, x_2 \ge 0.$

Since this problem is with two input variables, says, x_1 and x_2, according to literatures [6] and [10], at least 17 uniformly distributed sampling points are needed to conduct the discretization with uniform experimental design within the working domain. Here we try to employ the uniform table U*24(24^9) to perform the discretization, the consequences are shown in Tab. 2.

From Tab. 2, it can be seen that 5 sampling points are excluded due to the restraint of the s. t. conditions, and 19 sampling points are within the working domain of the s. t. conditions, which meets the requirement of at least 17 uniformly distributed sampling points within the domain of the s. t. conditions. In this problem, both the goal functions $f_1(x)$ and $f_2(x)$ are beneficial indexes, while the goal functions $f_3(x)$ is an unbeneficial index.

Table 2 Consequences of discretization with U*24(24^9)

No.	Input variable		Value of goal function			Notes
	x_1	x_2	f_1	f_2	f_3	
1	0.5556	13.125	1613.889	8097.222	27.9167	
2	1.6667	26.875	3341.667	16791.67	58.75	
3	2.7778	9.375	1319.444	6736.111	27.0833	
4	3.8889	23.125	3047.222	15430.56	57.9167	
5	5	5.625	1025	5375	26.25	
6	6.1111	19.375	2752.778	14069.44	57.0833	
7	7.2222	1.875	730.5556	4013.889	25.4167	
8	8.3333	15.625	2458.333	12708.33	56.25	
9	9.4444	29.375				Excl.
10	10.5556	11.875	2163.889	11347.22	55.4167	
11	11.6667	25.625	3891.667	20041.67	86.25	
12	12.7778	8.125	1869.444	9986.111	54.5833	
13	13.8889	21.875	3597.222	18680.56	85.4167	
14	15	4.375	1575	8625	53.75	
15	16.1111	18.125	3302.778	17319.44	84.5833	
16	17.2222	0.625	1280.556	7263.889	52.9167	
17	18.3333	14.375	3008.333	15958.33	83.75	
18	19.4444	28.125				Excl.
19	20.5556	10.625	2713.889	14597.22	82.9167	
20	21.6667	24.375				Excl.
21	22.7778	6.875	2419.444	13236.11	82.0833	
22	23.8889	20.625				Excl.
23	25	3.125	2125	11875	81.25	
24	26.1111	16.875				Excl.

Table 3 Results of the assessments with PMOO with U*24(24^9)

No.	Favorable probability			
	Partial favorable probability			Total
	P_{f1}	P_{f2}	P_{f3}	$P_t \times 10^5$
1	0.0365	0.0349	0.0858	10.9197
2	0.0755	0.0723	0.0539	29.452
3	0.0298	0.0290	0.0867	7.5014
4	0.0689	0.0665	0.0548	25.0748
5	0.0232	0.0232	0.0875	4.6962
6	0.0622	0.0606	0.0556	20.9791
7	0.0165	0.0173	0.0884	2.5242
8	0.0556	0.0547	0.0565	17.1850
9	0	0	0	0
10	0.0489	0.0489	0.0574	13.7128
11	0.0880	0.0863	0.0254	19.3228
12	0.0423	0.0430	0.0582	10.5826
13	0.0813	0.0805	0.0263	17.2122
14	0.0356	0.0372	0.0591	7.81463
15	0.0747	0.0746	0.0272	15.1322
16	0.0289	0.0313	0.0510	5.4291
17	0.0680	0.0687	0.0280	13.1031
18	0	0	0	0
19	0.06135	0.0629	0.0289	11.145
20	0	0	0	0
21	0.0547	0.0570	0.0298	9.2784
22	0	0	0	0
23	0.0480	0.0512	0.0306	7.5231
24	0	0	0	0

Tab. 3 shows the results of the assessments with PMOO, P_{f1}, P_{f2} and P_{f3} represent the partial favorable probabilities of functions f_1, f_2 and f_3 at the corresponding discretized sampling points, respectively; P_t expresses the total / overall favorable probability of each alternative. From Tab. 3, it can be seen that the sampling point No. 2 exhibits the maximum value of total favorable probability. Therefore, further optimization by using sequential uniform design is conducted around the sampling pint No. 2 of the Tab. 2.

Tab. 4 shows the results of the assessments by using sequential uniform design for further optimization, in which $c^{(t)}$ = (Max $P_t(i–1)$ – Max $P_t(i)$)/Max $P_t(i–1)$ expresses the relative error of the maximum total favorable probability at i^{th} sequential step. If we assume a pre-assigned value $\delta = 2\%$ for $c^{(t)}$, then the final optimal consequences for this multi-objective optimization problem are $f_{1\text{Opt.}} = 3591.927$, $f_{2\text{Opt.}} = 17962.24$ and $f_{3\text{Opt.}} = 59.9609$ at the 5th step with "coordinates" $x_1^* = 0.0521$ and $x_2^* = 29.9023$. Obviously, x_1^* and x_2^* approach to 0 and 30 at ultimate limit, respectively, which corresponds to optimum values of $f_{1\text{Opt.}} = 3600$, $f_{2\text{Opt.}} = 18000$ and $f_{3\text{Opt.}} = 60$, individually.

Table 4 Results of the assessments by using sequential uniform design with U*24(24^9)

Step	Domain	Optimum "coordinates"		Value of goal			Max. total favorable probability $P_t \times 10^5$	$c^{(t)}$
		x_1^*	x_2^*	$f_{1\text{Opt.}}$	$f_{2\text{Opt.}}$	$f_{3\text{Opt.}}$		
0	[0, 26.6667] × [0, 30]	1.6667	26.8750	3341.6670	16791.6700	58.7500	29.4520	
1	[0, 13.3333] × [15, 30]	0.8333	28.4375	3470.8330	17395.8300	59.3750	14.4243	0.5102
2	[0, 6.6667] × [22.5, 30]	0.4167	29.2188	3535.4170	17697.9200	59.6875	12.4265	0.1385
3	[0, 3.3333] × [26.25, 30]	0.2083	29.6094	3567.7080	17848.9600	59.8438	11.5726	0.0687
4	[0, 1.6667] × [28.125, 30]	0.1042	29.8047	3585.8540	17924.4800	59.9219	11.1760	0.0343
5	[0, 0.8333] × [29.0625, 30]	0.0521	29.9023	3591.9270	17962.2400	59.9609	10.9848	0.0171

3.2 Production with Maximum Profits and One Output

A factory produces two products: *A* and *B*. The profit of producing each piece of *A* is 4 ¥RMB, and the profit of producing each piece of *B* is 3 ¥RMB. The processing time of each piece of *A* is twice as long as that of each piece of *B*. If the whole time is used to process *B*, 500 pieces of *B* can be produced for per day. The factory's daily supply of raw materials is only enough to produce a total of 400 pieces of *A* and *B*. Besides, the product *A* is a tight-fitting product that sells very well.

Now, the problem is how to arrange the daily outputs of *A* and *B* so that the factory can obtain the maximum profit under the existing conditions.

Solution:

Let's first set x_1 = daily output of product *A*, x_2 = daily output of product *B* [11]. Then, it gets following mathematical model,

$$\text{Max } f_1(x) = 4x_1 + 3x_2,$$
$$\text{Max } f_2(x) = x_1,$$
s. t.
$$x_1 + x_2 \le 400,$$
$$2x_1 + x_2 \le 500,$$
$$x_1, x_2 \ge 0.$$

Since this problem is with two input variables, says, x_1 and x_2, again at least 17 uniformly distributed sampling points could be used to conduct the discretization with uniform experimental design within the working domain [6, 10]. Here we try to use the uniform table U*31(31^{10}) to perform the discretization, the consequences are shown in Tab. 5. From Tab. 5, it can be seen that 14 sampling points are excluded due to the restraint of the s. t. conditions, and 17 sampling points luckily are within the domain of the s. t. conditions, which satisfies the requirement of at least 17 uniformly distributed sampling points within the domain of the s. t. conditions. In this problem, both the goal functions $f_1(x)$ and $f_2(x)$ are beneficial indexes.

Table 5 Consequences of discretization with U*31(31^{10})

No.	Input variable		Value of goal function		Notes
	x_1	x_2	f_1	f_2	
1	4.0323	109.6774	345.1613	4.0323	
2	12.0968	225.8065	725.8065	12.0968	
3	20.1613	341.9355	1106.452	20.1613	
4	28.2258	45.1613	248.3871	28.2258	
5	36.2903	161.2903	629.0323	36.2903	
6	44.3548	277.4194	1009.677	44.3548	
7	52.4194	393.5484			Excl.
8	60.4839	96.7742	532.2581	60.4839	
9	68.5484	212.9032	912.9032	68.5484	
10	76.6129	329.0323			Excl.
11	84.6774	32.2581	435.4839	84.6774	
12	92.7419	148.3871	816.129	92.7419	
13	100.8065	264.5161	1196.774	100.8065	
14	108.871	380.6452			Excl.
15	116.9355	83.8710	719.3548	116.9355	
16	125	200	1100	125	
17	133.0645	316.129			Excl.
18	141.129	19.3548	622.5806	141.129	
19	149.1935	135.4839	1003.226	149.1935	
20	157.2581	251.6129			Excl.
21	165.3226	367.7419			Excl.
22	173.3871	70.9677	906.4516	173.3871	
23	181.4516	187.0968			Excl.
24	189.5161	303.2258			Excl.
25	**197.581**	**6.4516**	**809.677**	**197.581**	
26	205.6452	122.5806			Excl.
27	213.7097	238.7097			Excl.
28	221.7742	354.8387			Excl.
29	229.8387	58.06452			Excl.
30	237.9032	174.1935			Excl.
31	245.9677	290.3226			Excl.

Tab. 6 shows the results of the assessments with PMOO, P_{f1} and P_{f2} indicate the partial favorable probabilities of functions f_1 and f_2 at the corresponding discretized sampling points, individually; P_t represents the total / overall favorable

probability of each alternative. From Tab. 6, it can be seen that the sampling point *No. 25* exhibits the maximum value of total favorable probability. Therefore, further optimization by using sequential uniform design is conducted around the sampling point No. 25 of the Tab. 5.

Table 6 Results of the assessments with PMOO with U*31(31^{10})

No.	Favorable probability		
	Partial favorable probability		Total
	P_{f1}	P_{f2}	$P_t\times10^3$
1	0.0263	0.0028	0.0729
2	0.0553	0.0083	0.4598
3	0.0843	0.0139	1.1681
4	0.0189	0.01934	0.3671
5	0.0479	0.0249	1.1954
6	0.0770	0.0305	2.3451
7	0	0	0
8	0.0406	0.0416	1.6858
9	0.0696	0.0471	3.2768
10	0	0	0
11	0.0332	0.0582	1.9310
12	0.0622	0.0637	3.9634
13	0.0912	0.0693	6.3173
14	0	0	0
15	0.0548	0.0803	4.4048
16	0.0838	0.0859	7.2000
17	0	0	0
18	0.0475	0.0970	4.6009
19	0.0765	0.1025	7.8376
20	0	0	0
21	0	0	0
22	0.0691	0.1191	8.2299
23	0	0	0
24	0	0	0
25	**0.0617**	**0.1357**	**8.377**
26	0	0	0
27	0	0	0
28	0	0	0
29	0	0	0
30	0	0	0
31	0	0	0

Tab. 7 shows the results of the assessments by using sequential uniform design for further optimization. Again, set a pre-assigned value $\delta = 2\%$ for $c^{(t)}$, then the final optimal consequences for this multi-objective optimization problem are $f_{1\text{Opt.}} = 1000.56$ and $f_{2\text{Opt.}} = 249.597$ at the 6^{th} step with "coordinates" $x_1^* = 249.597$ and $x_2^* = 0.7258$. Analogically, the tendencies of x_1^* and x_2^* are 250 and 0 at ultimate limit, respectively, which leads to optimum values of $f_{1\text{Opt.}} = 1000$ and $f_{2\text{Opt.}} = 250$, separately.

Table 7 Results of the assessments by using sequential uniform design with U*31(31^{10})

Step	Domain	Optimum coordinates		Value of goal		Max. total favorable probability $P_t\times10^3$	$c^{(t)}$
		x_1^*	x_2^*	$f_{1\text{Opt.}}$	$f_{2\text{Opt.}}$		
0	[0, 250] × [0, 400]	197.5810	6.4516	809.6770	197.5810	8.3770	
1	[100, 250] × [0, 200]	218.5484	3.2258	883.8710	218.5484	4.1994	0.4987
2	[170, 250] × [0, 100]	233.2260	1.6129	937.7420	233.2260	3.3800	0.1951
3	[210, 250] × [0, 50]	241.6130	0.8065	968.8710	241.6130	3.0381	0.1011
4	[230, 250] × [0, 20]	248.3870	2.9032	1002.2600	248.3870	1.7052	0.4387
5	[240, 250] × [0, 10]	249.1935	1.4516	1001.1290	249.1935	1.6512	0.0317
6	[245, 250] × [0, 5]	249.5970	0.7258	1000.5600	249.5970	1.6253	0.0157

4 DISCUSSION

In the past, the multi-objective programming problem was solved usually by using "linear weighting method" [1, 2], i.e., "additive algorithm" in the previous approaches to transfer the multiple objectives into a single objective one, which is the intrinsic problem in principle in the viewpoint of probability theory with "union set" in essence [3]. Or some approaches even took certain objectives as restraint conditions to solve the multi-objective programming problem [1, 2], which obviously deviates from the original intention of multi-objective programming problem in the spirit of the "simultaneous optimization of multiple objectives" essentially.

While, the probability-based method for multi – objective optimization attempts to treat the simultaneous optimization of multiple objectives in the viewpoint of probability theory, which is the proper methodology for multi-objective optimization [3-5]. Therefore, the consequences of the previous approaches are incomparable to the results of the probability-based method for multi-objective optimization due to their intrinsic problem.

5 CONCLUSION

By using probability-based multi-objective optimization for the simultaneous optimization of multiple objectives, discrete uniform experimental design for performing simplification, and the sequential algorithm for conducting further optimization, the multi-objective programming problem can be conducted rationally. The approach properly takes the simultaneous optimization of each objective of multi-objective programming problem into account, which reflects the essence of the multi-objective programming naturally, and creates a new way.

Conflict Statement

There is no conflict of interest.

6 REFERENCES

[1] Chang, H., Wu, Q., & Liu, X. (2019). *Operations Research Tutorial*. Tianjin Science and Technology Press, Tianjin, China.

[2] Liu, S. (2014). *Theoretical Method and Application of Multi-objective Programming*. Shanghai Jiaotong University Press,

Shanghai, China.
[3] Zheng, M., Wang, Y., & Teng, H. (2021). A new "Intersection" method for multi-objective optimization in material selection. *Tehnicki Glasnik, 15*(4), 562-568. https://doi.org/10.31803/tg-202109 01142449
[4] Zheng, M., Wang, Y., & Teng, H. (2022). A novel method based on probability theory for simultaneous optimization of multi-object orthogonal test design in material engineering. *Kovove Materialy, 60*(1), 45-53.
[5] Zheng, M., Wang, Y., & Teng, H. (2022). A novel approach based on probability theory for material selection. *Materialwissenschaft und Werkstofftechnik, 53*(6), 666-674. https://doi.org/10.1002/ mawe.202100226
[6] Zheng, M., Teng, H., Yu, J., Cui, Y., & Wang, Y. (2022). *Probability-based multi-objective optimization for material selection*, Springer, Singapore.
[7] Hua, L.-K. & Wang, Y. (1981). *Applications of Number Theory to Numerical Analysis*. Science Press, Springer, Berlin; New York. I
[8] Fang, K-T. & Wang, Y. (1994). *Number-theoretic Methods in Statistics*, Chapman & Hall, London, UK. https://doi.org/10.1007/978-1-4899-3095-8
[9] Fang, K.-T., Liu, M.-Q., Qin, H., & Zhou, Y-D. (2018). *Theory and Application of Uniform Experimental Designs*, Science Press & Singapore: Springer Nature, Beijing. https://doi.org/10.1007/978-981-13-2041-5
[10] Fang, K.T. (1994). *Uniform Design and Uniform Design Table*. Science Press, Beijing, China.
[11] Ying, J. (1988). *Multi-objective planning*. People's Education Press, Beijing, China.

Authors' contacts:

Maosheng Zheng
(Corresponding author)
School of Chemical Engineering, Northwest University,
Xi'an, 710069, China
E-mail: mszhengok@aliyun.com

Haipeng Teng
School of Chemical Engineering, Northwest University,
Xi'an, 710069, China

Yi Wang
School of Chemical Engineering, Northwest University,
Xi'an, 710069, China

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230513115809

Original scientific paper
Received: 2023-05-13, Accepted: 2023-05-17

Numerical Simulation of Conveying Fine Powders in a Screw Conveyor Using the Discrete Element Method

Marko Motaln*, Tone Lerher

Abstract: Due to their high efficiency and spatial utilization, screw conveyors are widely used in pharmacy, agriculture, and industry. Recently, this has made it a popular research subject in the numerical modelling of the transport of bulk solids. Modelling of granular systems at the level of individual particles is mainly possible due to the use of discrete numerical methods. The most common is the use of the Discrete Element Method (DEM), which is still limited from the point of view of simulations on an industrial scale, as increasing the size of the system also increases the cost of simulation. Certain powders with low density, large angles of repose, poor fluidity, and bad flowability can accumulate during transportation, causing inaccurate and non-uniform movement. Additionally, the friction and impact between the particles can cause wear. To address these issues, the present study utilizes the discrete element method to simulate and analyse powder transportation in an inclined screw conveyor using the commercial software ANSYS-ROCKY. Numerous phenomena that arise while transporting and feeding small-sized or irregularly shaped particles, often present in industrial processes, remain insufficiently investigated. This paper aims to analyse the transportation process of adhesive powders in a screw conveyor, with a focus on evaluating the impact of different screw blade speeds on transport. Multiple simulations were conducted, along with the implementation of an additional wear model, to better understand the transport phenomena and wear. An example was used to demonstrate the impact of screw speed on the wear of the transporter due to the interaction between the material and the structure of the conveyor, power consumption, and performance.

Keywords: abrasive wear; bulk handling; coarse graining; conveying equipment; discrete element method; screw conveyor

1 INTRODUCTION

In many industrial processes, it is crucial to maintain accurate control over the transportation of solid materials. Modern approaches to the numerical modelling of transport systems are based on advanced numerical methods. To adequately describe the transport of bulk materials and powders, it is necessary to understand and describe various phenomena that occur during the transportation of different bulk solids [1-3] in order to transport materials to the desired location, production plant, or machine. During transport, we deal with mixing, agglomeration of transported material, the flow behaviour of granular material, recirculation, wear, and many other phenomena that need to be utilized, suppressed, or applied for desired use [4-6].

One of the most commonly used methods for transportation of granular material and powder is the screw conveyor, which is widely utilized in the food, pharmaceutical, chemical, agricultural, and other industries due to its high efficiency, low cost, compact structure, and ability to provide uniform and precise feeding [7-9]. Screw conveyors are primarily utilized for transporting and lifting bulk materials and powders over short distances. The structure consists of an outer cylinder and an inner screw, but despite their seemingly simple mechanical design, the behaviour of granular solids and powders during transport is quite complex. Simulations of granular flow and describing phenomena that occurs during transport, such as wear, mixing, flow, pouring, or particle compaction, have already been included in the simulation in the past. It has been found that we can influence these phenomena with an appropriate design and construction of a transporter [3, 2]. There has been significant interest among the scientific community in studying the screw conveyor's conveying process. Therefore, substantial efforts have been made to describe the conveyors' transportation parameters through conventional methods such as theoretical studies and experiments [7, 9]. Despite considerable efforts in academic research and traditional experimental techniques, a comprehensive understanding of the mechanisms underlying granular flows and their impact on the conveying performance of screw conveyors due to the lack of particle-level information remains unknown. Gaining deep insights into the system is challenging due to the difficulty in obtaining essential details such as particle collisions. With the rapid development of hardware, numerical simulation has emerged as an attractive tool for investigating and simulating complex particle systems. Modified geometry has been considered in many studies that have been conducted using spherical elements to approximate particles. If we highlight only a few, we can see that the first application of numerical modelling on a screw conveyor was already implemented back in 2001 by Shimizu and Cundall [10]. Since then, studies have been conducted on mixing granular media during transportation [11], granular flow, and performance with varying boundary conditions. Recently, we have witnessed a notable increase in the use of graphical processing units (GPU). The GPU-based algorithms in the discrete element method (DEM) have become increasingly popular for simulating complex and full-scale industrial processes, thanks to advancements in computer hardware. Several authors in their research have demonstrated the capabilities of multi-GPU computation and set benchmarks for successful simulations of large-scale industrial processes. These simulations accurately capture the behaviour of complex, non-spherical particles [12, 13]. As computer hardware, software, and numerical algorithms advance, we are gaining deeper insights into powder behaviour through complex geometries (e.g., polyhedral, elliptical elements, or multi-spheres). Cohesion and adhesion can influence the behaviour and flow of powder particles,

varying from one type of powder to another. Moreover, powders can often display characteristics of both liquids and solids, making them difficult to handle [14, 15]. The studies we have mentioned are based solely on spherical particles, which are not typically present in natural systems. A much more accurate approach is to capture the actual geometry of particles, which has been shown to have a significant impact on the granular flow of material [16]. Due to its exceptional importance and economic benefits, special attention has been given in the past to describing the wear during transport. Roughly a quarter of the world's energy consumption is derived from tribological contacts. Of this amount, around 20% is employed to conquer friction, while the remaining 3% is dedicated to the refurbishment of worn-out components and auxiliary equipment due to wear [17]. The economic impact of abrasive wear extends beyond replacement costs, encompassing expenses related to machine downtime and reduced productivity. Since the friction and wear caused by the movement of solid particles and powders during transportation can affect both the surface and the particles, accurate prediction of the location and magnitude of erosion in the equipment is essential to prevent the failure of devices. During the transportation of granules and powders, the primary types of wear mechanisms are impact wear and abrasion wear, which are also commonly observed in mining chutes [18]. Researchers have conducted many experimental and numerical studies to anticipate the wear of transport systems [19-21]. A new, exciting model called SIEM (Shear Impact Energy Model) has been recently proposed [22]. It is suitable for analysing sparsely or densely packed domains with particles in numerical simulations using the DEM method. It is based on the relationship between the normal and shear impact energy of a particle. The results show that almost one-fourth of the shear impact energy is converted into erosion and that the highest material removal rate is achieved at 30 degrees of impact. Furthermore, the lifespan of a chute heavily depends on the chosen material. Materials that exhibit higher resistance to abrasion wear, such as high chromium alloy and ceramics, are commonly chosen for this application[23-26]. Despite the numerous theoretical research and traditional experiments conducted in the past to optimize screw conveyor performance, they have their limitations in providing detailed particle-level information. This information is crucial in understanding the underlying mechanisms of granular and powder flows, which are essential in comprehending the phenomena and performance of screw conveyors. Additionally, powder flow is particularly poorly described in the literature, making it challenging to gain in-depth insights into the behaviour of powders. While simplified cohesion, complicated particle shape, and size distribution of particles are increasingly being considered, our understanding of these factors is still far from comprehensive. A study on screw auger is being conducted with the intention of evaluating powder flow in different screw auger speed regimes. A study is being conducted on screw auger to evaluate powder flow in different speed regimes. Coarse-graining techniques are being employed to analyse the effects of speed on power consumption, mass flow, and wear of the screw auger.

2 MATERIALS AND METHODS

Since the modelling approach is based on the Discrete Element Method (DEM), it is appropriate to provide a brief introduction to the DEM method in the following sections, as well as the corresponding models and modern approaches that were used in the numerical model.

2.1 Model Description

DEM is a widely used method for simulating the behaviour of granular flow. It is based on the Lagrangian description of individual particles, and the description of kinematics based on the principle of Newton's second law. The governing equations for individual particles can be expressed as follows:

$$m_i\,\vec{a}_i = m_i\,\frac{\mathrm{d}}{\mathrm{d}t}\vec{v}_i = \sum_{j=1}^{nc}\vec{F}_{ij} + \vec{F}_i^{\mathrm{g}} + \vec{F}_i^{\mathrm{ext}} \tag{1}$$

and

$$I_i\;\vec{\vartheta}_i = I_i\;\;\frac{\mathrm{d}}{\mathrm{d}t}\vec{\omega}_i = \sum_{j=1}^{nc}\vec{T}_{ij} + \vec{T}_i^{\mathrm{ang}} + \vec{T}_i^{\mathrm{ext}} \tag{2}$$

Since the method employs the Lagrangian numerical approach, each particle is represented as a separate entity with unique physical characteristics, including radius R_i, position $\vec{x}_i$, angle $\vec{\varphi}_i$, translational velocity $\vec{v}_i$, angular velocity $\vec{\omega}_i$, translational acceleration $\vec{a}_i$, and angular acceleration $\vec{\vartheta}_i$. The terms m_i and I_i are the mass and the inertia tensor of the particle i. The inertia term in the equation can be a scalar for spherical particles since they are symmetric, meaning the moment of inertia is equal in all directions. Due to the contacts with neighbouring particles, additional forces $\vec{F}_{ij}$ and moments $\vec{T}_{ij}$ arise in the particle, which is the sum of all neighbouring particle contacts. Due to the gravitational field and other external factors, there are also forces $\vec{F}_i^{\mathrm{g}}$, $\vec{F}_i^{\mathrm{ext}}$, and moments $\vec{T}_i^{\mathrm{ang}}$, $\vec{T}_i^{\mathrm{ext}}$, being a result of rolling friction between particles or other external moments between particles and geometry such as electro-static, Van der Waals forces, or cohesive forces.

In the "soft sphere" approach model, we are dealing with rigid particles where deformation is simulated using overlap (superposition). We can introduce variables δ_{ij}^n and δ_{ij}^t, which denote normal and tangential overlap, respectively. An incremental approach is required to calculate the tangential overlap where $\vec{v}_{\mathrm{rel}}^{\,t}$ denotes relative velocity in the tangential direction.

$$\delta_{ij}^n = R_i + R_j - \left\|\vec{x}_j - \vec{x}_i\right\| \tag{3}$$

$$\vec{\delta}_{ij}^t = \int_{t_0}^{t}\vec{v}_{\mathrm{rel}}^{\,t}\,\mathrm{d}t \tag{4}$$

Modelling a nonlinear viscoelastic model is currently one of the most widely used constitutive models in many software programs because it provides precise and efficient descriptions of contact loads for DEM analysis. This model incorporates nonlinear elasticity and enables the modelling of particle collisions in the tangential direction. Contact in

the normal direction is based on the work of physicist H. Hertz while the credit for describing tangential contact is attributed to the duo of Mindlin-Deresiewicz [27]. The used model is schematically represented in Fig. 1.

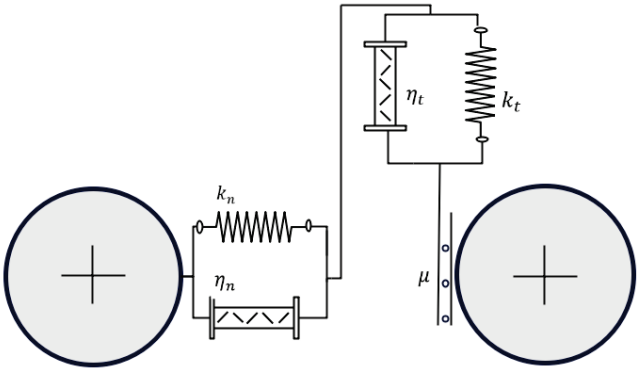


Figure 1 Hertz Mindlin contact scheme

The contact in the normal direction can be expressed as:

$$F_{ij}^{n} = k_n \delta^{n\,\frac{3}{2}} + \eta^n\, \delta^{n\,\frac{1}{4}}\, \vec{v}_{\mathrm{rel}}^{\,n} \tag{5}$$

$$k_n = \frac{4}{3} E^* \sqrt{R^*} \tag{6}$$

$$\eta_n = 2\, \eta_H \sqrt{k_n\, m^*} \tag{7}$$

Where E^*, R^*, and m^* are effective Young's modulus, effective radius, and effective mass respectively. The stiffness coefficient k_n and damping coefficient η_n are represented in equations 5 and 6 [28].Where η_H is described as the damping ratio and is for interested readers detailed described in [29].

Similar follows for contact in the tangential direction where the model in Rocky can describe friction coefficient μ where its value is set to static coefficient μ_s when no sliding occurs and similarly value is set to the value of dynamic coefficient μ_d when sliding occurs. The resulting force can be described by the following equations [28]:

$$\vec{F}_{ij}^{\,t} = -\mu\, F_{ij}^{n} \left(1 - \zeta^{\frac{3}{2}}\right) \frac{\vec{\delta}^{\,t}}{\left|\vec{\delta}^{\,t}\right|} + \beta \sqrt{\frac{6\,\mu\, m^*\, F_{ij}^{n}}{\delta_{\mathrm{MAX}}^{t}}}\; \zeta^{\frac{1}{4}}\, \vec{v}_{\mathrm{rel}}^{\,t} \tag{8}$$

$$\zeta = 1 - \frac{\min\left(\left|\vec{\delta}_{ij}^{\,t}\right|, \delta_{\mathrm{MAX}}^{t}\right)}{\delta_{\mathrm{MAX}}^{t}} \tag{9}$$

Where β is the tangential damping ratio and can be calculated from the coefficient of restitution as follows:

$$\beta = \frac{\ln(c_r)}{\sqrt{\ln^2(c_r) + \pi^2}} \tag{10}$$

The variables $\vec{v}_{\mathrm{rel}}^{\,t}$ and $\delta_{\mathrm{MAX}}^{t}$ are representing relative tangential velocity and maximal allowed tangential overlap, respectively. The reader can quickly determine that the variable ζ in the equation becomes equal to 0 as soon as the tangential overlap exceeds the allowed limit. At that point, the model becomes equivalent to the Coulomb model, and the force is equal to the product of μ and F_{ij}^{n} [28].

The behaviour and flow of powder particles can be influenced by cohesion and adhesion. Several options are available for handling cohesive particles in DEM simulations. The most popular methods used in the overviewed literature [4, 30] are the van der Waals model and the JKR model. As a result of obtaining a material model (refer to section 2.4), which was calibrated using a simplified constant model, we have decided to adopt a similar model. This model applies a constant force $\vec{F}_i^{\mathrm{adh}}$ when separated normal overlap $-s_n$ is smaller than prescribed adhesive distance δ_{adh} . This model is the most basic one and can be expressed using the following equation [28]:

$$\vec{F}_i^{\mathrm{adh}} = \begin{cases} 0 & if - s_n \geq \delta_{\mathrm{adh}} \\ f_{\mathrm{adh}}\; g\; \min(m_1, m_2) & if - s_n \geq \delta_{\mathrm{adh}} \end{cases} \tag{11}$$

We obtained a system of three vector equations usually integrated in time using explicit time integration schemes. The most used procedures include the explicit Euler and velocity Verlet schemes because of their simplicity, stability, and low memory usage.

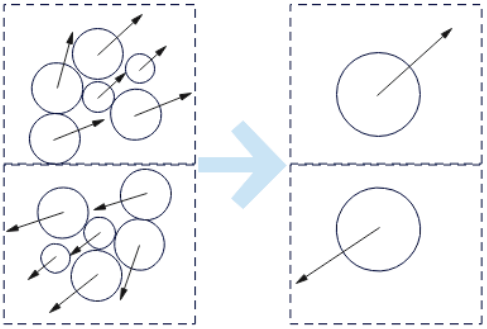


Figure 2 Coarse graining scheme

2.2 Coarse Graining

Despite significant advancements in accelerating discrete numerical simulations through GPU technology, precise simulations of real-world industrial systems are often unfeasible due to increasingly complex geometrical and physical models that address contact. This engineering challenge is particularly evident when using computationally expensive methods like the DEM, where obtaining sufficiently accurate results for a system at acceptable computational costs is a common problem. This involves introducing large pseudo-particles to represent a group of original particles, significantly reducing the number of particles needed for simulation [31]. The used method of particle merging is well presented in the article [32] and relies on correcting contact forces accordingly. The objective is to develop a medium that can scale particle sizes and maintain the same energy density and evolution of energy density as the basic model. If the density and volume fraction of the particle remains constant, the density of gravitational potential energy is not affected by the grain radius. Thus, it is necessary to maintain a constant density. In order to preserve kinetic energy, the scaling process must not

influence particle velocities. Similar studies have been conducted in articles [31, 33-36].

2.3 Wear Model

In DEM most frequently used are Archard or Finnie's model. Finnie developed a mathematical model based on mass, speed, and impact angle [18]. Archard's model, on the other hand, is based on the ratio between the shearing work and the overlapping volume and is also included in our simulation model [5]. Archard's phenomenological principle that correlates the loss of material volume to the energy expended by frictional forces acting on the material's surface. Typically, Archard's law is mathematically represented by the following equation [28]:

$$V = \frac{k\,W\tau}{H} \tag{12}$$

Where V, $W\tau$, H, and k respectively represent the total volume of material worn from the surface, the work done by tangential forces, the hardness of the material, and a dimensionless empirical constant. The fact that precise treatment of wear is a highly complex problem is confirmed by the dependence of wear on the size, shape, and chemical composition of particles. Increasing throughput performance and particle velocity typically result in higher wear rates. Based on the existing literature, it can be argued that the influence of particle characteristics such as shape, size (especially for powders), and adhesion on flow behaviours, conveying performance, and wear patterns in screw conveyors has not been thoroughly investigated. Consequently, this field presents several open challenges and research questions for future studies [16].

2.4 Simulation Model

The most reliable and verifiable models can be produced by calibrating materials using experimental results. Proper material calibration is essential, especially when there is insufficient computing power to solve full scale problems. For simulation previously described Coarse graining with ratio of 10 is employed, and calibration is required to produce the most reliable simulation data. For our numerical analysis, we extracted data from work [37], where full material calibration was performed with the inclusion of the coarse grain method and particle size distribution grouping.

Table 1 Simulation parameters

Value	Property
$r = 0{,}000724$	Particle radius (m)
$\rho_p = 712$	Particle density (kg/m^3)
$\rho_g = 7850$	Geometry density (kg/m^3)
$E_p = 18$, $E_g = 210$	Young's modulus (GPa)
$\nu_p = 0.3$, $\nu_g = 0.3$	Poisson ratio (-)
$\mu_p = 0{,}7$, $\mu_g = 0.35$	Sliding friction (-)
$\mu_p = 0{,}7$, $\mu_g = 0.35$	Roling friction (-)
$c_{rp} = 0{,}83$, $c_{rg} = 0{,}8$ $n = 43000$ $d = 140$	Rest. Coefficient (-) Particle count (-) Screw diameter (mm)

Additionally, we repeated the calibration of the static angle of repose containing a rigid plane and lifting cylinder and used a constant adhesive model that is suficient for most cases where the adhesive force is present which is true for most substances such as fine pharmaceutical powder or wet rocks. The parameters of the adhesive model were adjusted based on repeated calibration.

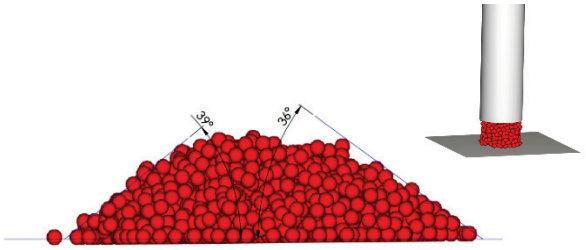


Figure 3 Repetition of calibration with enlarged particles

A three-dimensional model of a typical powder steel screw conveyor has been modelled. A full geometry model can be shown on Fig. 4 and is assembled from a housing, screw, flanges, inlet, outlet chutes, and other standard elements like motor which are not shown. The inlet and outlet can typically be customized to meet the specific requirements of the work site. For the purposes of this simulation, a fixed inclination angle of 45° has been set. Properties or characteristics not visible in Fig. 4 can be found in Tab. 1.

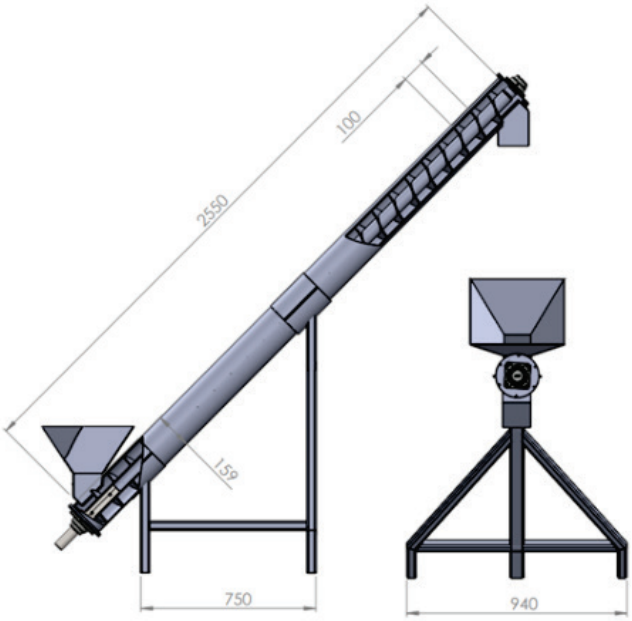


Figure 4 Full inclined screw conveyor geometry

In addition to increasing the number of particles, an additional limitation was imposed for the purpose of numerical simulation due to limited computational power. To describe the behaviour of particles during transport, it was not necessary to model the entire transporter at full scale. Therefore, a simplified model was created which includes a feed port, a helical body, and a housing. To minimize computational power while ensuring that the flow resulting

from scattering or disposal sites has already stabilized in the observed area, the model was designed to reduce the length of the screw conveyor by an additional 2/3, while still maintaining an equivalent description of the original system.

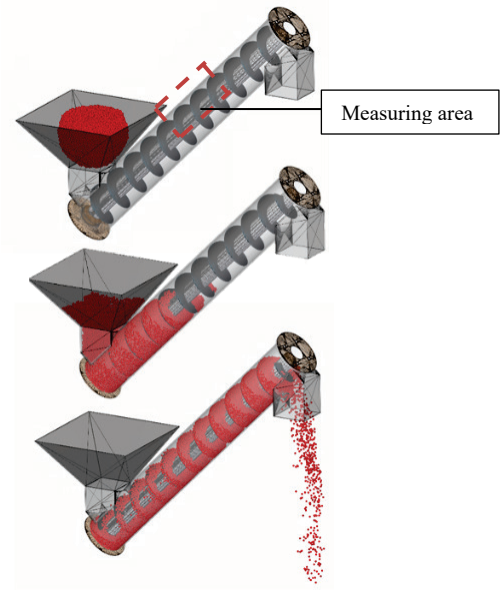


Figure 5 Simplified screw conveyor geometry

The simulation model for this research was created using Ansys-ROCKY 2022/R2. In a simplified model using a volumetric formula for particle creation at the point of the scattering site, an initial quantity of 4 kg of particles is generated. This quantity represents approximately 30% filling of the model container. This was followed by the activation of the screw rotation with variation of three rotational speeds (100 rpm, 300 rpm, and 500 rpm). Once the screw starts rotating, there is an initial transient state where the particles settle, and the stresses stabilize. The flow is allowed to stabilize before any measurements are taken. The simulation is repeated multiple times in different configurations of rotation speed and with or without an adhesion model. Simulation is continued until the container is emptied. Additional measuring points were created to provide a more detailed analysis of the geometric components and particles in different phases of the particle flow. Furthermore, measuring points based on Euler statistics were installed at the midpoint of the transporter tube (Fig. 5).

3 RESULTS AND DISCUSSION

The results are presented in two sections, with the first section focusing on mass flow and power consumption, and the second section describing the results of wear analysis.

3.1 Mass Flow Rate

Fig. 6 shows the results of an analysis on the relationship between the rotational speed of the screw and the measuring area void fraction. The analysis reveals that increasing the rotational speed of the screw results in a higher void fraction. This is because particles have less time to fill the void in the pitch of the screw at higher rotational speeds. Theoretically, the proportion of voids could be reduced by using an extended charging inlet in the hopper, but this is often not feasible in practice. Similarly, from Fig. 7, we can observe the void fraction as function of time. The peaks demonstrate a stable flow at maximum filling. Furthermore, as the rotational speed increases, we can observe a decrease in the height of the stable flow peak. This may be attributed to the decreasing weight of particles in the feed hopper over time. Therefore, the weight or amount of material in the hopper also affects the void fraction.

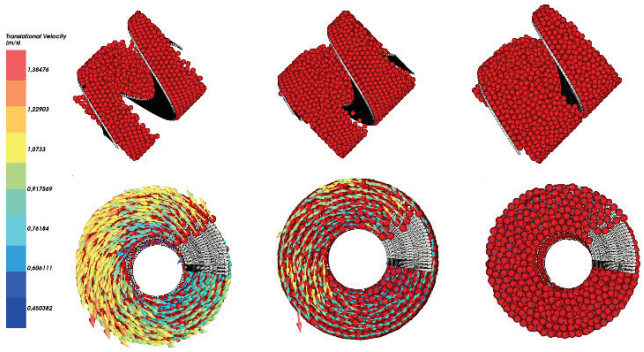


Figure 6 Flow of particles within a screw conveyor and the direction of particle flow when viewed along the axis of the conveyor.

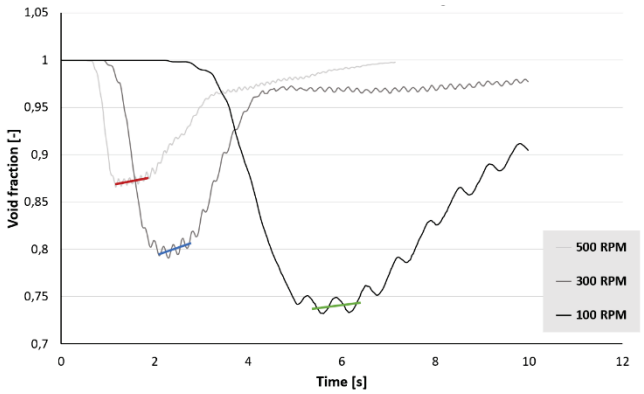


Figure 7 The void fraction of the control volume as a function of time

Despite the highest void fraction at the highest rotational speed, the results in Fig. 8 show that the mass flow rate of material is still the highest at the highest screw speed. The increase in mass flow rate seems to be nonlinear, as there is very little increase in mass flow rate between 300 and 500 revolutions per minute compared to the increase at lower angular speeds. The mass flow rate is calculated based on the calculation of the mass of particles that leave the control volume in a time step. The calculation must take into account the fact that some particles may return to the control volume due to the inclination of the conveyor.

The influence of cohesion can also be seen in a limited way in Fig. 8 which is denoted by the 'coh' sign on the graph. Similar to previous studies results show that cohesion have a significant effect on material flow. At the highest rotational speed, it is visible that the powder with cohesion has a lower mass flow rate, which is result of lower filling of the pitch. Since our material has relatively low adhesive force, these

phenomena are not particularly pronounced. It would be interesting to increase the adhesion and monitor the influence on porosity, filling, or other phenomena, such as arching.

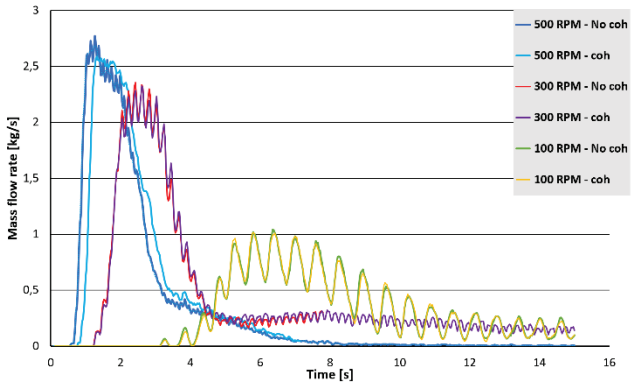


Figure 8 The Particle Mass flow rate in the control volume as a function of time

Through the application of DEM, one can conduct a thorough analysis of the way energy is distributed within a system. Specifically, the screw acts as a supplier of energy to the particles, whereby a certain proportion of this energy is dissipated, while the remaining portion is converted into mechanical energy. Fig. 9 demonstrates the power consumption of the screw auger at three different angular speeds. It can be clearly observed that the power consumption tends to increase with increasing screw rotation speed. The graph was analysed to calculate the ratio of power consumption to mass flow rate for all three angular velocities. Based on the limited data, it can be concluded that beyond a certain speed, the energy utilization changes less and tends to a constant value. This graph also provides insights into selecting the optimal angular velocity based on energy consumption.

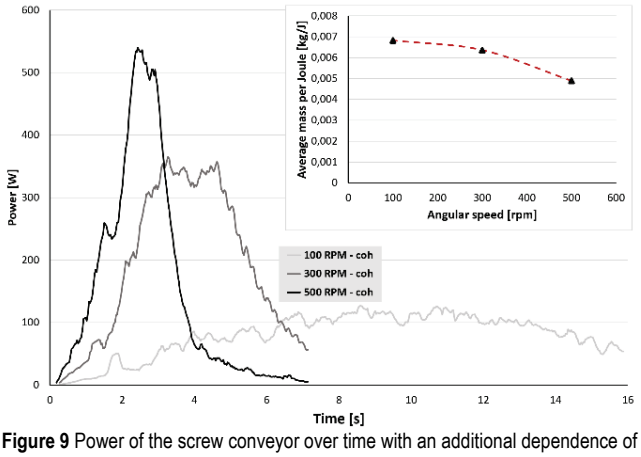


Figure 9 Power of the screw conveyor over time with an additional dependence of mass normalized to energy.

3.2 Wear Rate

Detailed information about the energy dissipated in each collision during the simulated period can be provided. Shear intensity refers to the amount of power that is transferred per unit area, and it is calculated based on the work done by the contact forces during a collision. This kind of energy is used in abrasive wear models as the model is useful for determining shear wear on geometric surfaces. The model is applicable in determining shear wear on geometric surfaces and can be directly linked to wear. Fig. 10 illustrates a similar regime of increasing throughput performance and particle velocity, resulting in higher shear intensity of the screw over time.

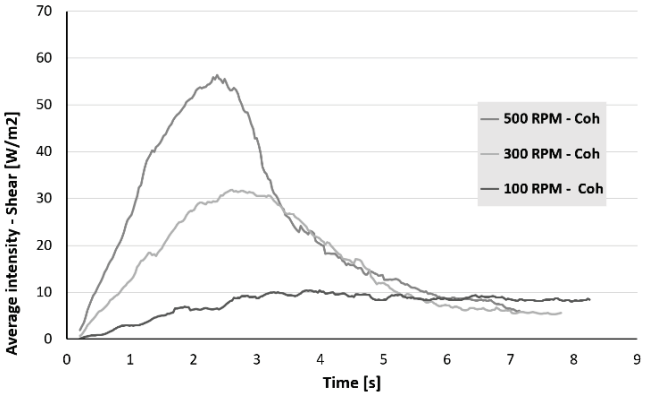


Figure 10 Shear intensity of the screw over time

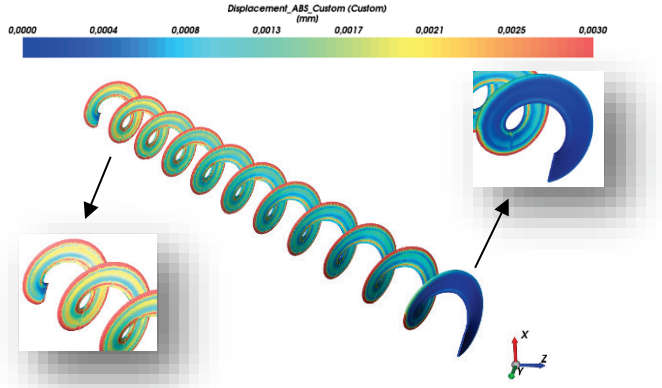


Figure 11 Screw wear rate distribution at 300 rpm

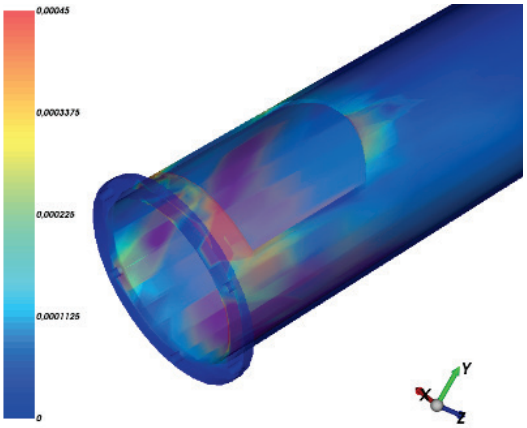


Figure 12 Housing wear rate distribution at 300 rpm

Wear was also described using the Archard wear model presented in section 2.4. It is particularly important to emphasize that in order to obtain an accurate quantitative description of wear, as shown in the images, it is necessary to determine the empirical wear constant correctly. As we did

not have access to this data within the scope of this study, the default value was used, and the description of wear is presented solely for the purpose of obtaining qualitative results, resulting in numerical results that may not be entirely accurate but are sufficient for an overview of the condition of the housing geometry, as the areas most exposed to damage. Wear can be presented in Figs. 11 and 12 for the screw and housing, respectively.

As shown in Fig. 11, the wear rate of the screw blade increases from the screw shaft toward the edge. The maximum wear occurs at the inlet where additional mass acting on the screw, results in greater wear over the entire surface. At the exit of the screw conveyor, only a few contacts are detected, resulting in minimal wear. Reverse side of the screw is due to its negligible amount of wear not considered.

4 CONCLUSIONS

To gain a better understanding of transport phenomena and wear of adhesive powders in screw conveyor, multiple simulations were conducted, incorporating additional wear models. A brief overview of the literature used is presented, followed by a demonstration of the numerical modelling.

The void fraction and mass flow rate in the screw feeder were examined and expressed as a function of time. When varying the screw blade speeds, it was found that the void fraction in the screw conveyor increases as the speed increases. This phenomenon was attributed to the time constraints as particles are unable to sort in time at higher speeds. However, the mass flow rate increases with an increase in the screw blade speed. Limited results also demonstrate the impact of adhesion, resulting in lower mass flow rates. Power consumption of the screw auger was also evaluated for varying angular speeds, revealing that power consumption tends to increase as the screw rotation speed increases. The shear intensity and Archard wear model were used to demonstrate qualitative wear on the screw conveyor and housing. It was demonstrated that the system exhibits a tendency towards increased energy consumption and wear at higher particle speeds.

The study utilizes previously calibrated materials, which may limit the accuracy and reliability of the simulations. Further validation with real materials and simulation data is necessary to support the simulations. It would be worthwhile to explore the effects of increased adhesion on porosity, filling, and arching.

Reliable models require calibration using experimental results. With more precise input data and results, we can expect better solutions for the design of screw conveyors, reduced wear, and the selection of optimal rotational speeds. In the future, we aim to investigate the impact of particle wear without using coarse graining, using real particle shapes, which would require more accurate input data.

Acknowledgments

This research work was supported by the Slovenian Research Agency ARRS in the framework of the applied research project "Warehousing 4.0 – Integration model of robotics and warehouse order-picking systems", grant number L5-2626.

5 REFERENCES

[1] Xia, R., Wang, X., Li, B., Wei, X., & Yang, Z. (2019). Discrete Element Method- (DEM-) Based Study on the Wear Mechanism and Wear Regularity in Scraper Conveyor Chutes. *Mathematical Problems in Engineering*, Article ID 4191570, 1-12. https://doi.org/10.1155/2019/4191570

[2] Gelnar, D. & Zegzulka, J. (2019). *Discrete Element Method in the Design of Transport Systems: Verification and Validation of 3D Models*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-05713-8

[3] Žídek, M, et al. (2021). Effective use of DEM to design chain conveyor geometry. *Chemical Engineering Research and Design, 167*, 25-36. https://doi.org/10.1016/j.cherd.2020.12.017

[4] Shi, Q. & Sakai, M. (2022). Recent progress on the discrete element method simulations for powder transport systems: A review. *Advanced Powder Technology, 33*(8), 103664. https://doi.org/10.1016/j.apt.2022.103664

[5] Lee, S.-J., Lee, J.-H., & Hwang, S.-Y. (2022). Suggestion of Practical Application of Discrete Element Method for Long-Term Wear of Metallic Materials. *Applied Sciences, 12*(20), 10423. https://doi.org/10.3390/app122010423

[6] Gelnar, D., Prokeš, R., Jezerska, L., & Zegzulka, J. (2021). Wood pellets transport with vibrating conveyor: experimental for DEM simulations analysis. *Scientific Reports, 11*(1), 16606. https://doi.org/10.1038/s41598-021-96111-2

[7] Yu, Y. & Arnold, P. C. (1996). The influence of screw feeders on bin flow patterns. *Powder Technology, 88*(1), 81-87. https://doi.org/10.1016/0032-5910(96)03107-5

[8] Owen, P. J. & Cleary, P. W. (2009). Prediction of screw conveyor performance using the Discrete Element Method (DEM). *Powder Technology, 193*(3), 274-288. https://doi.org/10.1016/j.powtec.2009.03.012

[9] Roberts, A. W. (1999). The influence of granular vortex motion on the volumetric performance of enclosed screw conveyors. *Powder Technology, 104*(1), 56-67. https://doi.org/10.1016/S0032-5910(99)00039-X

[10] Shimizu, Y. & Cundall, P. A. (2001). Three-Dimensional DEM Simulations of Bulk Handling by Screw Conveyors. *J. Eng. Mech., 127*(9), 864-872. https://doi.org/10.1061/(ASCE)0733-9399(2001)127:9(864)

[11] Pezo, L., Jovanović, A., Pezo, M., Čolović, R. & Lončar, B. (2015). Modified screw conveyor-mixers – Discrete element modeling approach. *Advanced Powder Technology, 26*(5), 1391-1399. https://doi.org/10.1016/j.apt.2015.07.016

[12] Dong, Y., Yan, D., & Cui, L. (2022). An Efficient Parallel Framework for the Discrete Element Method Using GPU. *Applied Sciences, 12*(6), 3107. https://doi.org/10.3390/app12063107

[13] Govender, N., Cleary, P. W., Wilke, D. N., & Khinast, J. (2021). The influence of faceted particle shapes on material dynamics in screw conveying. *Chemical Engineering Science, 243*, 116654. https://doi.org/10.1016/j.ces.2021.116654

[14] Widartiningsih, P. M., et al. (2020). Coarse graining DEM simulations of a powder die-filling system. *Powder Technology, 371*, 83-95. https://doi.org/10.1016/j.powtec.2020.05.063

[15] Thornton, C. & Yin, K. K. (1991). Impact of elastic spheres with and without adhesion. *Powder Technology, 65*(1-3), 153-166. https://doi.org/10.1016/0032-5910(91)80178-L

[16] Sun, H., Ma, H., & Zhao, Y. (2017). DEM investigation on conveying of non-spherical particles in a screw conveyor. *Particuology, 65*, 17-31. https://doi.org/10.1016/j.partic.2021.06.009
[17] Holmberg, K. & Erdemir, A. (2917). Influence of tribology on global energy consumption, costs and emissions. *Friction, 5*, 263-284. https://doi.org/10.1007/s40544-017-0183-5
[18] Ou, T. & Chen, W. (2022). On accurate prediction of transfer chute wear using a digital wear sensor and discrete element modelling. *Powder Technology, 407*, 117680. https://doi.org/10.1016/j.powtec.2022.117680
[19] Chen, W., Biswas, S., Roberts, A., O'Shea, J., & Williams, K. (2017). Abrasion wear resistance of wall lining materials in bins and chutes during iron ore mining. *International Journal of Mineral Processing, 167*, 42-48. https://doi.org/10.1016/j.minpro.2017.08.002
[20] Markov, D. P. (2022). Adhesion at friction and wear. *Friction, 10*, 1859-1878. https://doi.org/10.1007/s40544-021-0564-7
[21] Archard, J. F. (1958). A crossed-cylinders friction machine. *Wear, 2*(1), 21-27. https://doi.org/10.1016/0043-1648(58)90337-5
[22] Zhao, Y., Ma, H., Xu, L., & Zheng, J. (2017). An erosion model for the discrete element method. *Particuology, 34*, 81-88. https://doi.org/10.1016/j.partic.2016.12.005
[23] Yoganandh, J., Natarajan, S., & Babu, S. P. K. (2013). Erosive Wear Behavior of Nickel-Based High Alloy White Cast Iron Under Mining Conditions Using Orthogonal Array. *J. of Materi Eng and Perform, 22*(9), 2534-2541. https://doi.org/10.1007/s11665-013-0539-6
[24] Wu, L., Guo, X., & Zhang, J. (2014). Abrasive Resistant Coatings—A Review. *Lubricants, 2*(2), 66-89. https://doi.org/10.3390/lubricants2020066
[25] Xie, Y., Jiang, J. (Jimmy), & Islam, M. A. (2021). Applications of elastomers in slurry transport. *Wear, 477*, p. 203773. https://doi.org/10.1016/j.wear.2021.203773
[26] Walker, C. I. & Robbie, P. (2013). Comparison of some laboratory wear tests and field wear in slurry pumps. *Wear, 302*(1-2), 1026-1034. https://doi.org/10.1016/j.wear.2012.11.053
[27] Wang, Q. J. & Zhu, D. (2013). Hertz Theory: Contact of Spherical Surfaces. Wang, Q. J. & Chung, D. (Eds.). Encyclopedia of Tribology, Boston, MA: Springer US, 1654-1662. https://doi.org/10.1007/978-0-387-92897-5_492
[28] Rocky. (2022). ESSS Rocky, Release 2022 R2, DEM Technical Manual, ESSS Rocky DEM, S. R. L.
[29] Antypov, D. & Elliott, J. A. (2011). On an analytical solution for the damped Hertzian spring. *EPL, 94*(5), p. 50004. https://doi.org/10.1209/0295-5075/94/50004
[30] Sebastian Escotet-Espinoza, M., Foster, C. J., & Ierapetritou, M. (2018). Discrete Element Modeling (DEM) for mixing of cohesive solids in rotating cylinders. *Powder Technology, 335*, 124-136. https://doi.org/10.1016/j.powtec.2018.05.024
[31] Mori, Y., Wu, C.-Y., & Sakai, M. (2019). Validation study on a scaling law model of the DEM in industrial gas-solid flows. *Powder Technology, 343*, 101-112. https://doi.org/10.1016/j.powtec.2018.11.015
[32] Shimizu, Y. & Cundall, P. (2001). Three-Dimensional DEM Simulations of Bulk Handling by Screw Conveyors. *Journal of Engineering Mechanics, 127*(9). https://doi.org/10.1061/(ASCE)0733-9399(2001)127:9(864)
[33] Xu, L., Bao, S., & Zhao, Y. (2020). Multi-level DEM study on liner wear in tumbling mills for an engineering level approach. *Powder Technology, 364*, 332-342. https://doi.org/10.1016/j.powtec.2020.02.004
[34] Cai, R. & Zhao, Y. (2020). An experimentally validated coarse-grain DEM study of monodisperse granular mixing. Powder *Technology, 361*, 99-111. https://doi.org/10.1016/j.powtec.2019.10.023
[35] Radl, S., Radeke, C., Khinast, J. G., & Sundaresan, S. (2011). Parcel-Based Approach for the Simulation of Gas-Particle Flows.
[36] Queteschiner, D., Lichtenegger, T., Pirker, S., & Schneiderbauer, S. (2018). Multi-level coarse-grain model of the DEM. *Powder Technology, 338*, 614-624. https://doi.org/10.1016/j.powtec.2018.07.033
[37] Martinčević, N. (2017). Numerična analiza transporta sipkega materiala v procesu suhe granulacije. *MSc thesis*, Univerza v Mariboru, Fakulteta za strojništvo. (in Slovenian) https://dk.um.si/IzpisGradiva.php?id=68896

Authors' contacts:

Marko Motaln, MSc
(Corresponding author)
University of Maribor, Faculty of Mechanical Engineering,
Smetanova ul. 17, 2000 Maribor, Slovenia
marko.motaln2@um.si

Tone Lerher, Professor PhD
University of Maribor, Faculty of Mechanical Engineering,
Smetanova ul. 17, 2000 Maribor, Slovenia
tone.lerher@um.si

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220321232047

Preliminary communication
Received: 2022-03-21, Accepted: 2022-10-17

Face Detection and Recognition Using Raspberry PI Computer

Mario Dubovečak, Emil Dumić, Andrija Bernik*

Abstract: This paper presents a face detection and recognition system utilizing a Raspberry Pi computer that is built on a predefined framework. The theoretical section of this article shows several techniques that can be used for face detection, including Haar cascades, Histograms of Oriented Gradients, Support Vector Machine and Deep Learning Methods. The paper also provides examples of some commonly used face recognition techniques, including Fisherfaces, Eigenfaces, Histogram of Local Binary Patterns, SIFT and SURF descriptor-based methods and Deep Learning Methods. The practical aspect of this paper demonstrates use of a Raspberry Pi computer, along with supplementary tools and software, to detect and recognize faces using a pre-defined dataset.

Keywords: face detection; face recognition; haar cascades; LBPH; Raspberry PI

1 INTRODUCTION

As new technologies emerged, so did the desire and need to create artificial intelligence, which would improve and simplify our daily lives [1]. All fields of business employ computer technologies to improve processes while also advancing those activities through the introduction of unique, new solutions. Artificial intelligence, of which computer vision is an essential factor, seeks to automate specific operations as closely as possible to human nature [2]. It has been used for some time to identify qualities that are both unique to each individual and part of the group, such as papillary lines, speech patterns, and other physical features, as well as to collect additional information for data analysis, such age, sex, personal habits, and the like. It is a well-known fact that these data are already employed in virtualization for market research, machine learning, creating target products, and surveillance with the aim of enhancing safety and data protection. One of the most common examples is unlocking smart phones using biometric information (fingerprint or facial recognition unlocking) [3]. It all began with this specific goal in mind and then moved to other mobile applications in mobile banking, security and data storage, the application procedure for e-services, and other related areas.

A variety of options are provided with computer vision, with a continuing focus on new software advancements that make use of the hardware at hand. The majority of those are freely accessible on the Internet and have the potential to be upgraded, expanded, and implemented. One example that offers a variety of functions from the computer vision section and supports several programming languages such as Python, Java, and C++, is the open source computer vision library (OpenCV) [4]. Additionally, it is designed to be compatible with a variety of operating systems, including Windows, Linux, iOS, and Android. This paper demonstrates the interaction of software and hardware in a shared unit, as well as publicly accessible tools and services that do not fall under the definition of "personal computers" but rather are modular devices that enable and support all essential recognition functions.

2 ALGORITHMS FOR FACE DETECTION

The most important face detection techniques, including Haar cascades [5], histogram of oriented gradients (HOG) [6], support vector machine (SVM) [7] and deep learning methods [8] (which can be used for both object detection and recognition), will be discussed in the sections below.

2.1 Haar Cascades

Using the automatic object detection technology known as Haar cascades, objects can be instantly recognized in a video or image. The algorithm was developed by Michael Jones and Paul Viola in 2001. Although being an older algorithm, the Viola-Jones (or Haar cascades) technique proved to be an effective tool for real-time face detection (or generally object detection) [4]. The technique is based on four basic steps: calculating Haar features, creating integral image [9], AdaBoost training [10] and cascade classifiers [11].

2.1.1 Calculating Haar Features (Viola-Jones)

Human faces have inherent characteristics that distinguish one face from another. These include the cheeks, nose, lips, and eyes. The regions that can be utilized to train algorithms to find faces using rectangular regions differ significantly from one another. The calculation requires computing the differences between the sums after adding the pixel intensities in each location.

Three different features are used by Viola-Jones: The two-rectangle feature is a difference in the sum of the pixels contained in two rectangular sections (Figs. 1a, 1b). The three-rectangle feature (Fig. 1c) calculates the sum in a centre rectangle minus the sum in two outer rectangles. The diagonal difference between pairs of rectangles is computed using the four-rectangle feature (Fig. 1d).

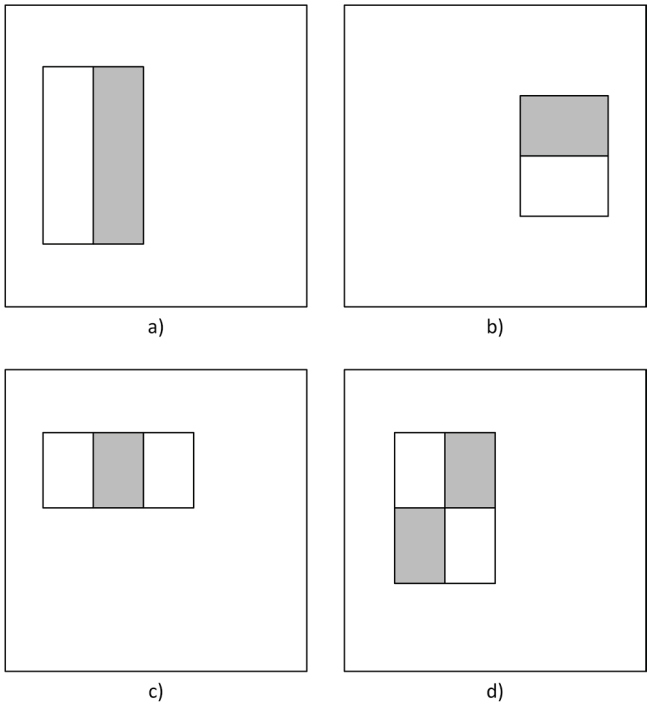


Figure 1 Four different types of features used in the frame

2.1.2 Creating Integral Image

Using an indirect image representation known as an integral image [9] or a summed-area table [12], rectangle features can be quickly determined. In fact, the integral image is a double integral of the image (first along the rows and then along the columns). Fig. 2a shows an integral picture at position x, y, where $ii(x, y)$ is an integral image and $i(x, y)$ is the original image. Eq. (1) states that the integral image at position x, y is the sum of the pixel values above and to the left of x, y.

$$ii(x, y) = \sum_{x' \le x,\, y' \le y} i(x', y') \tag{1}$$

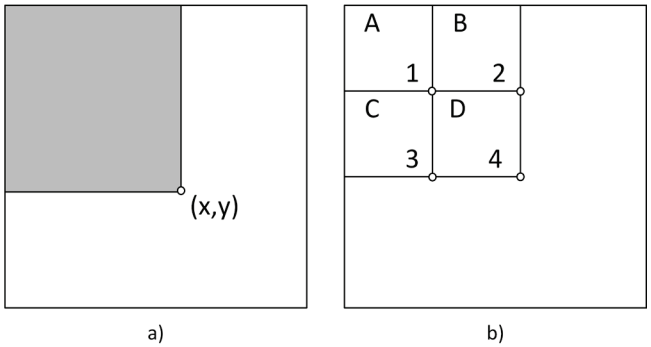


Figure 2 a) Integral image value at point (x, y); b) Example of four rectangle features

As an example, integral image calculation at position x, y is shown in Fig. 2a, as a sum of all pixels in the gray rectangle. The total value of rectangle D can be calculated based on the locations shown in Fig. 2b. Area A can be calculated by summing the pixels in image up to point 1 ($Sum_A = 1$ in integral image). Area B can be calculated by subtracting all the pixels in image within the space of point 2 and point 1 ($Sum_B = 2 - 1$ in integral image). Area C can be calculated by subtracting all the pixels in image within the space of point 3 and point 1 ($Sum_C = 3 - 1$ in integral image). Area D can be calculated by subtracting all pixels in image within the space of points 4 and 1, 2 and 3 ($Sum_D = 4 - Sum_A - Sum_B - Sum_C = 4 - 1 - (2 - 1) - (3 - 1) = 4 - 2 - 3 + 1$ in integral image).

Concluding, each rectangular sum may be computed using the integral images with four numbers, and as a result, the difference between two rectangular sums can be computed using eight numbers. It requires six numbers to calculate the difference between two adjacent rectangle features (Figs. 1a, 1b), eight numbers to compute three rectangle feature example (Fig. 1c) and nine numbers to compute the four-rectangle feature example (Fig. 1d).

2.1.3 Adaptive Boosting (AdaBoost)

Yoav Freund and Robert Schapiro [10] created the machine learning meta-algorithm AdaBoost, which may be used with a variety of other learning algorithms to increase performance. The results of the other learning algorithms are merged to create a weighted sum, which represents the boosted classifier's final results.

The algorithm becomes more accurate by learning from the images it receives as input, which might define false positives and true negatives. A very accurate model might be created by looking at every possible position and combination. Due to the numerous possibilities and combinations that must be tested for each scene or image, such training can be time-consuming.

2.1.4 Cascade Classifiers

Cascade classifiers are made of stages, each of which is an ensemble of weak learners [11]. Each stage is trained with the help of the boosting method. By using a weighted average of the decisions made by the weak learners, boosting makes it possible to train a classifier that is highly accurate.

Each classifier stage assigns a positive or negative label to the area that is specified by the sliding window's current position. Positive indicates an object was detected in a window, whereas negative means the object was not detected in the window. If the label is negative, the classification of that region is complete, and the detector then moves on to the first location afterwards. If not, the classifier with a positive label passes the scanned area for additional processing. The detector analyzes the selected area and provides information about the discovered object. It takes a lot of time to find genuine positive results because the great majority of windows do not contain the object of interest.

The more samples that help the recognition process successful, the better the overall result for indicating and rejecting false positives will be, which will be used as the basis for the judgment. For this, a tool for image labeling that identifies an object of interest and generates a table of positive samples can be used. To obtain acceptable detector accuracy, it is also necessary to offer a set of negative images

from which the function can automatically generate negative samples.

When calculating performance metrics for object detection, usually three sample patterns are used: A true positive, TP – when a positive sample is correctly classified; A false positive, FP – when a negative sample is mistakenly classified as positive; A alse negative, FN – when a positive sample is mistakenly classified as negative. A true negative, TN – when a negative sample is correctly classified as negative, is usually discarded in performance metrics for object detection due to the higher number of TN samples.

2.2 Histogram of Oriented Gradients (HOG)

In computer vision and image processing, a feature descriptor HOG can be used for object detection. The inventor of the HOG concept was Robert K. McConnell. Navneet Dalal and Bill Triggs presented in 2005 their supplementary work on HOG descriptors concentrated on pedestrian detection in static images as well as person detection in videos [6].

Given the enormous variety of appearances and poses that humans might take, it is necessary to develop a collection of features that would make it possible to distinguish the human figure from the background. The HOG approach is comparable to Edge Orientation Histograms and SIFT descriptors, but it varies in that it is computed on a dense grid of evenly spaced cells and uses overlapping local contrast normalization for increased accuracy.

The basic classifier uses a linear SVM which results in simplicity and speed. The essential idea behind the HOG descriptor is that the object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The image is divided into small connected regions called cells, and for the pixels within each cell a HOG is compiled. This type of histogram concatenation is called a descriptor. For improved accuracy, the local histograms can be contrast-normalized by computing a measure of the intensity across a larger region of the image (a block) which is then used to normalize all cells within the block. HOG operates on local cells, it is invariant to geometric and photometric transformations, with the exception of object orientation.

A linear SVM is used in the basic classifier, which results in speed and simplicity. The central part of the HOG descriptor is that the distribution of intensity gradients or edge directions can be used to describe the appearance and shape of objects inside an image. A HOG is compiled for each pixel contained in each of the image's small, interconnected cells. A descriptor is a histogram concatenation from all cells. By obtaining a measure of the intensity across a larger area of the image (a block), which is then used to normalize all cells inside the block, the local histograms can be contrast-normalized for increased accuracy. With the exception of object orientation, HOG operates on local cells and is invariant to geometric and photometric transformations.

2.3 Support Vector Machines (SVM)

SVM is used as a classification method to create the class border [7]. One-class SVM is used to discover anomalies in the normal class, and all points outside of margins are considered anomalous. SVM belongs to the class of linked algorithms for supervised machine learning that combine data analysis and pattern recognition to analyze data for classification or regression. The classifier's structure and properties determine the methods.

Liner SVM is the best-known classifier which predicts each input's member class between two possible classifications. SMV builds a hyperplane or set of hyperplanes to classify all inputs in a high-dimensional or infinite space. The closest values to the classification margin are known as support vectors whose goal is to maximize the margin between the hyperplane and the support vectors. Optimal separation can be achieved when the hyperplane is at the maximum distance from any class data point (functional margin). The following rule applies: the larger the distance the smaller is the classifier's generalization error. Support vectors are in fact data points that lie at the edge of the plane closest to the hyperplane that separates them.

The most well-known classifier, liner SVM, predicts the member class of each input between two alternative classes. For the purpose of categorizing all inputs in an infinite or high-dimensional space, SVM constructs a hyperplane or group of hyperplanes. Support vectors are the values that are closest to the classification margin and whose objective is to maximize the margin between the hyperplane and the support vectors. The hyperplane can be at its farthest position from any class data point to ensure optimal separation (functional margin). The following rule applies: the generalization error of the classifier decreases with increasing distance. In fact, the data points at the edge of the plane closest to the hyperplane separating them are the support vectors.

A training dataset of n points in the form $(\vec{x}_i, Y_1), \ldots, (\vec{x}_n, Y_n)$ is given, where the Y_i are either 1 or –1 depending on the class to which $\vec{x}_i$ belongs. Each $\vec{x}_i$ is a p-dimensional vector. The objective of a linear SVM is to find the maximum-margin hyperplane that separates the group of points $\vec{x}_i$ for $Y_i = 1$ from the group of points for $Y_i = -1$, where the distance between the hyperplane and the closest point $\vec{x}_i$ is maximal.

3 DEEP LEARNING MODELS

The detection of objects, faces, human activity, human poses, and data sets can all be done using deep learning algorithms [8]. Deep learning techniques could be divided into: Convolutional neural networks, Deep Belief Networks (DBN), Deep Boltzmann Machines (DBM), as well as stacked autoencoders (SAE).

3.1 Convolutional Neural Network (CNN)

The convolutional neural network (CNN) typically has three types of layers: the convolutional layer, the pooling layer, and the fully connected layer. Each layer has its own role in detecting input data and converting it into neural activation. In the end it leads to fully connected layers which

results in copying the input data into the vector feature (output). CNN is widely used in the areas of face detection and recognition, machine navigation and pattern recognition, the automotive industry, and the like.

3.2 Deep Belief Network (DBN) and Deep Boltzmann Machines (DBM)

DBN as well as DBM are both deep neural networks. They use the restricted Boltzmann machine (RBM), a generative stochastic neural network, as a learning module. Restricted Boltzmann Machine is s variant of Boltzmann machine, with the restriction that the neurons must form a bipartite graph. DBNs have undirected connections at the top two layers made of restricted Boltzmann machines and directed connections to the lower layers (Fig. 3a). DBMs connect all network layers using undirected connections (Fig. 3b).

By stacking the RBMs, DBNs are models that guarantee joint probability distribution on the observed data and labels. At the beginning, DBNs use training layer-by-layer to calculate initial weights, and afterwards fine-tuning is being performed over all weights to calculate the correct outputs. DBNs are graphical models that can be trained with the aim of extracting a deep hierarchical representation of the training data. However, there is a significant disadvantage associated with DBN, namely that the two-dimensional structures of the input image are not taken into account, which can be related with the speed and applicability of the algorithm itself.

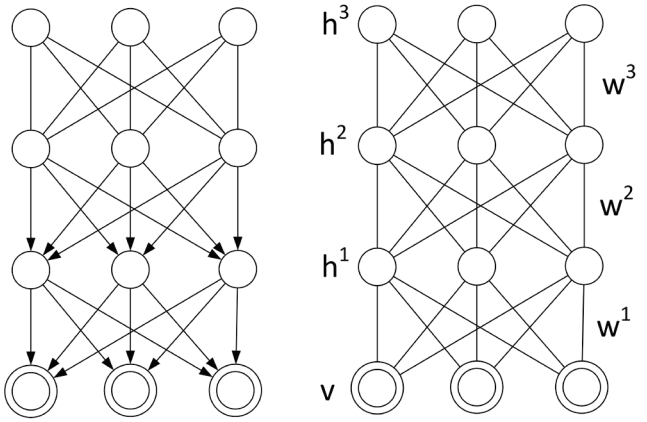


Figure 3 Schematic of the: a) DBN and b) DBM

3.3 Stacked Autoencoders (SAE)

Autoencoder is the the primary component of stacked autoencoders (SAE). The objective of an autoencoder is to encode the input $\boldsymbol{x}$ into reproduction $\boldsymbol{r}(\boldsymbol{x})$, which implies that the input is actually the target output. The reconstruction error is minimized in this process.

When training the network with one hidden linear layer using mean squared error (MSE) as a loss function, k hidden units project the input to first k principal components of the input data, similarly to the principal component analysis (PCA). The autoencoder does not operate as PCA if the hidden layer is nonlinear.

For example, in denoising autoencoders the input is "randomly" corrupted while the original input is used for target output reconstruction. The learning process begins by setting some inputs to zero after which the autoencoder tries to predict the corrupted values in relation to the uncorrupted ones. Two important steps in this process are:

- Autoencoder aims to encode the input by remembering the input data;
- Autoencoder aims to cancel the influence of random corruption applied to the input.

A deep network can be formed as a SAE by feeding the lower layer output features as an input to the next layer. Unsupervised pretraining is performed layer-by-layer, where each layer is trained as an autoencoder that minimizes the MSE by input reconstruction. Afterwards a logistic regression layer is added after each layer and then fine-tuning of the stacked encoding parts is performed, in a supervised manner.

A comparison of features of the mentioned models is presented in Tab. 1 [8].

Table 1 Comparison between different deep learning-based methods

Model features	CNN	DBN/DBM	SAE
Unsupervised learning	-	+	+
Training effectiveness	-	-	+
Learning features	+	-	-
Scaling, rotation, translation	+	-	-
Generalization	+	+	+

4 ALGORITHMS FOR FACE RECOGNITION

The output of facial recognition system is identification or verification of a subject or subjects that appear on an image or in a video.

In the Eigenfaces method, image is used as an input and the name of the individual is the output [13]. Eigenfaces method can be slower for larger datasets and larger dimensions of input images. Reducing the large images is required to accomplish effective performance. An image of dimension $m \times n$ is in fact $(m \times n) \times 1$ vector. Principal component analysis is one method for reducing dimensionality (PCA), which is based on determining the hyper-surface onto which all maximally scattered points are projected. When using PCA, the $m \times n$ image is reduced to a smaller space, allowing for quick implementation and resistance to noise and change. Since the face in the input image is not always centered, a cascade classifier is used to separate the face from the background of the image.

Fisherfaces method teaches a specific class of matrix transformation insensitive to illumination changes and different facial expressions [14]. Performance of this method depends on the input data, and the reconstruction of the projected image is made similarly as in Eigenfaces method.

A very straightforward but efficient visual descriptor, which can be used for face recognition tasks, is called local binary patterns (LBP) [15]. It marks a boundary between

neighboring pixels, assigning a binary number as a result. The procedure involves dividing the test window into cells (for instance, 16×16 pixels for each cell), comparing each pixel with its 8 neighbors, and then moving in a circle around the cells (clockwise or counterclockwise). If the value of the center pixel is higher than the neighboring pixel's value, the value 0 is written; otherwise, the value 1 is written. This produces an 8-digit binary number, which is usually converted to a decimal value. The cell histogram computation of each number's frequency is shown as a 256-dimensional vector. To obtain LBPH (LBP Histograms), LBP image is divided into multiple grids. Histogram is calculated from each grid (each having the size of 256, representing intensity of 0-255 after LBP calculation), and finally all histograms are concatenated to represent the characteristics of the whole image. In the face recognition process, LBPH from tested face is compared with the trained LBPH dataset and face (or a person identity) is returned with the closest histogram.

SIFT (Scale-Invariant Feature Transform) approach provides a solution regarding image rotation, scaling, intensity, changes in viewpoint, and changes in lightning [16]. The SIFT algorithm has four steps. The scale-space peak selection is used in the first step to determine the position and keypoints' ratio, by using difference of Gaussians (DoG) filter. In the second stage low contrast features are discarded. The third step is to assign an orientation, to establish keypoint orientation on a gradient image. The final step computes a descriptor for the local image region for each keypoint based on the size and orientation of the image gradient in each image pattern point in the region centered on a keypoint. This produces a 3D histogram of the location and orientation of the gradient with a 4×4 pixel neighborhood and an 8-bin cell for each pattern, producing the 128-dimension keypoint descriptor, which can be used to find matching pairs in different images. SIFT descriptors can be also used for face recognition, for example as in the paper [17].

SURF (Speed Up Robust Features) method [18], built on a multi-scaled space, consists of three basic components: detection, description and matching. Feature detector is based on the Hessian matrix. The SURF method can operate in parallel for different scalings and approximates LoG with the box filter, using integral image (described earlier). In the description part, SURF uses wavelet responses in both the x- and y-directions to assign orientation. Dominant orientation is estimated by adding up all responses within a $\pi/3$ sliding orientation window. Most applications do not require orientation, which further speeds up the procedure. This version of SURF method is called U-SURF ("upright" SURF). Finally, descriptors from different images can be compared to find matching pairs. Matching can be done only based on the contrast of the feature attributes, which additionaly speeds up the process. When compared with SIFT, SURF method has the same performance, but with the reduced computational complexity. SURF descriptors can be also used for face recognition, for example as in the paper [19].

For facial recognition, the FaceNet method can be also used [20]. In essence, it is a neural network that teaches how to map facial images into a compact space where distances represent the similarity of faces. The so-called triplet loss method is used to calculate the loss function. On the one hand, it increases the distance between an anchor and a negative (different identity), while minimising the distance between an anchor and a positive (same identity). Other deep learning models include DeepFace [21], VGGFace/VGGFace2 [22], FaceID-GAN [23] etc.

5 SETUP, TRAINING AND TESTING ENVIRONMENT

This paper uses the latest version of Raspberry Pi 4B device with all the required interconnected parts. VNC (Virtual Network Computing) remote control is used for visualization. A personal computer connected to joint network can be used for this purpose. Once a successful configuration of the device is performed, the installation of all necessary software such as OpenCV and other packages follows. The Viola-Jones algorithm was used as the algorithm for face detection, while the LBPH algorithm was used for face recognition [24]. Those algorithms were used to be able to run the algorithm on the Raspberry Pi device, which has limited hardware recources.

The main features of this program are placed in two files, one for testing and the other for recognition. The test file takes images located in Datasets directory's maps, indexes folder names (person names), and creates a *training.yml* file with all the necessary values that is used for comparison using the face recognition program. The file named *haarcascade_frontalface_default.xml* contains cascading classifiers associated with both files.

Program creates the file *oznaketrening.yml* from the images saved in the Datasets directory, based on which the data from the webcam video is later compared.

After launching a training program used to create and load values from all images from the corresponding folders and after comparing those same images with the video submitted from the webcam, the person's face in the frame is recognized and if it matches the one in the database then the name is printed in the recognition frame. It also sends a signal to the output of the Raspberry Pi device by turning off the red light on the LED module and turning on the green in a cycle time of 0.75 seconds (or 1.34 Hz). For example, this can be used as a signal to open the door.

Image dataset for 5 different people was created for the purpose of this paper. Characteristics of Raspberry Pi 4B devices were taken into account and dataset of 100 images (RGB format, dimensions 640×480 pixels with *.png* extension and in real/daily conditions) was made for each person, a total of 500 images. PNG format was used as a lossless compression, to be able to further process all images without adding additional artefacts. Higher-dimension images might be also used, but this resulted in additional heating of the Raspberry Pi device (up to 79° C without using additional coolers).

Half images per person were used for the training and the rest for testing purpose. Training was performed by running

Face_Trener.py file from the prepared dataset in order to train our program with the created dataset. The next step was to copy the entire dataset of images (250 in total) from which a slideshow was made (video file with *.avi* extension) for each of the 5 people. The settings for each image lasted 1 second, with an image editing transition of half a second, which resulted in processed 0.67 fps. Accordingly, separate video files do not overload the device. Each of the corresponding files is then loaded separately into the face recognition program using *Face_Prepoz.py*. In order to change the input to a video file it is necessary to change according to *cam=cv2.VideoCapture('Name_video_file.avi')* where the name of the file we want to load with the extension is specified. Loading these video files through the recognition system resulted in the expected result where faces were recognized in each of the images, although there were occasional inaccuracies in the comparison of the test images with the predefined ones.

Performance metrics for face recognition task can be calculated according to the following formulas, similarly as for the classification, using true positive (*TP*), false positives (*FP*), false negatives (*FN*) and true negative (*TN*) samples as defined in section 2.1.4.

Overall accuracy, Eq. (2):

$$ACC_{\text{overall}} = \frac{\sum(correct)}{\sum(overall)} = \frac{\sum(main\_diagonal)}{\sum(overall)} \tag{2}$$

Error rate, Eq. (3):

$$ERR_{\text{overall}} = 1 - ACC_{\text{overall}} = \frac{\sum(false)}{\sum(overall)} \tag{3}$$

Precision (Positive predictive value, *PPV*), Eq. (4):

$$PPV = \frac{TP}{TP + FP} \tag{4}$$

False discovery rate, *FDR*, Eq. (5):

$$FDR = 1 - PPV = \frac{FP}{TP + FP} \tag{5}$$

Recall (sensitivity or True positive rate, *TPR*), Eq. (6):

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

False negative rate, *FNR*, Eq. (7):

$$FNR = 1 - TPR = \frac{FN}{TP + FN} \tag{7}$$

From the Fig. 4, we can see that the overall accuracy of the proposed algorithm is 95.2% (last row on the right, green number). Precision is given for each person in the last column on the right (green number), while recall is in the last row (green number). Opposite results (1-green number) are given as red numbers: Error rate, 1-(overall accuracy) is last row on the right, red number. False discovery rate (*FDR*) is given for each person in the last column on the right (red number), while false negative rate (*FNR*) is in the last row (red number).

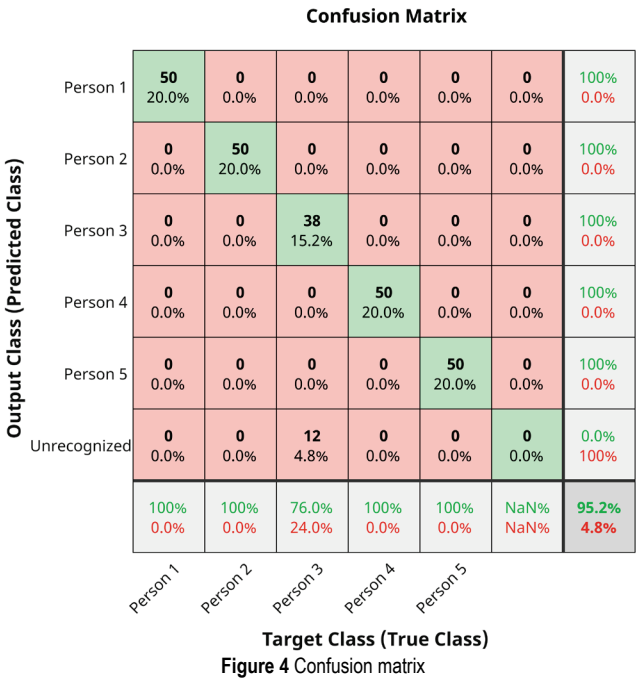


Figure 4 Confusion matrix

The results indicate that for the marked "Person 3" there is a total of 12 images who are not recognized, lowering its recall to 76%. In other words, the faces are detected on those images using Haar cascades, but they are not recognized compared to persons previously trained in the database – there were no matches. As a reason, it can be said that is was a woman with longer hair who in the unrecognized images had her head tilted to the side at an angle of approximately 45°.

6 CONCLUSIONS

In this paper, different algorithms for face detection and recognition are presented and discussed. In addition to the theoretical part, the procedure of preparing, installing and commissioning Raspberry Pi computer is described, together with all components, including software. The Haar cascades (Viola-Jones) were used as the algorithm for face detection, while LBPH algorithm was used for face recognition. Concerning the accuracy, it was noticed that the accuracy of face recognition can differ due to various parameters such as lighting, face rotation in all directions, distance from the camera as well as camera quality. System was tested to work with 0.67 fps processing time per 1 recognized identity.

Although higher-dimension images are more recognizable, they are also more demanding to handle. In

practice, this resulted in additional heating of the Raspberry Pi 4B device to an occasional 79° C without using additional coolers.

Due to its modularity, the Raspberry Pi device offers a foundation for further process in terms of controlling the entrance of selected persons through the door. The idea is to use the existing platform and upgrade it. In the future experiments, in addition to traffic control, network data collection (time, place, person) and remote control can be also enabled.

7 REFERENCES

[1] Lee, R. S. T. (2020). Artificial Intelligence in Daily Life. 1st Edition. Springer. https://doi.org/10.1007/978-981-15-7695-9

[2] Szeliski, R. (2022). *Computer Vision: Algorithms and Applications*. 2nd Edition. Springer. https://doi.org/10.1007/978-3-030-34372-9

[3] Unar, J. A., Seng, W. C., & Abbasi, A. (2014). A review of biometric technology along with trends and prospects. *Pattern recognition, 47*(8), 2673-2688. https://doi.org/10.1016/j.patcog.2014.01.016

[4] Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools, 120*, 122-125.

[5] Viola, P. & Jones, M. (2001). Robust real-time object detection. *International journal of computer vision, 57*, 34-47. Springer.

[6] Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. *IEEE computer society conference on computer vision and pattern recognition (CVPR'05), 1*, 886-893. https://doi.org/10.1109/CVPR.2005.177

[7] Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine learning, 20*(3), 273-297. https://doi.org/10.1007/BF00994018

[8] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience, 2018*, Article ID 7068349, 13 pages. https://doi.org/10.1155/2018/7068349

[9] Lewis, J. P. (1995). Fast template matching. *Vision Interface 95*, 120-123.

[10] Freund, Y. & Schapire, R. E. (1995). A desicion-theoretic generalization of on-line learning and an application to boosting. *Lecture Notes in Computer Science*, 23-37. https://doi.org/10.1007/3-540-59119-2_166

[11] Gama, J. & Brazdil, P. (2000). Cascade Generalization. *Machine Learning, 41*(3), 315-343. https://doi.org/10.1023/A:1007652114878

[12] Crow, F. (1984). Summed-area tables for texture mapping. *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer Graphics and Interactive Techniques*, 207-212. https://doi.org/10.1145/800031.808600

[13] Sirovich, L. & Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A, 4*(3), 519-524. https://doi.org/10.1364/JOSAA.4.000519

[14] Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence, 19*(7), 711-720. https://doi.org/10.1109/34.598228

[15] Ojala, T., Pietikäinen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition, 29*(1), 51-59. https://doi.org/10.1016/0031-3203(95)00067-4

[16] Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on computer vision, 2*, 1150-1157. https://doi.org/10.1109/ICCV.1999.790410

[17] Geng, C. & Jiang, X. (2009). Face recognition using sift features. *The 16th IEEE International Conference on Image Processing (ICIP)*, 3313-3316. https://doi.org/10.1109/ICIP.2009.5413956

[18] Bay, H., Tuytelaars, T., & Gool, L. V. (2006). Surf: Speeded up robust features. *European conference on computer vision*, 404-417. https://doi.org/10.1007/11744023_32

[19] Du, G., Su, F., & Cai, A. (2009). Face recognition using SURF features. Proc. *SPIE 7496, MIPPR 2009: Pattern Recognition and Computer Vision*, 749628. https://doi.org/10.1117/12.832636

[20] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 815-823. https://doi.org/10.1109/CVPR.2015.7298682

[21] Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. *IEEE Conference on Computer Vision and Pattern Recognition*, 1701-1708. https://doi.org/10.1109/CVPR.2014.220

[22] Cao, Q., Shen, L., Xie, W., Omka Parkhi, M., & Zisserman, A. (2018). VGGFace2: A Dataset for Recognising Faces across Pose and Age. *The 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 67-74. https://doi.org/10.1109/FG.2018.00020

[23] Shen, Y., Luo, P., Luo, P., Yan, J., Wang, X., & Tang, X., (2018). FaceID-GAN: Learning a Symmetry Three-Player GAN for Identity-Preserving Face Synthesis. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 821-830. https://doi.org/10.1109/CVPR.2018.00092

[24] Dubovecak, M. Face detection and recognition using Raspberry PI. (2020). *Master's thesis*. https://zir.nsk.hr/en/islandora/object/unin%3A3197, Accesed: 6.10.2022

Authors' contacts:

Mario Dubovečak
University North,
104. brigade 1, 42000 Varaždin, Croatia
madubovecak@unin.hr

Emil Dumić, PhD, Associate Professor
University North,
104. brigade 1, 42000 Varaždin, Croatia
edumic@unin.hr

Andrija Bernik, PhD, Assistant Professor
(Corresponding author)
University North,
104. brigade 1, 42000 Varaždin, Croatia
andrija.bernik@unin.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221227094126

Preliminary communication
Received: 2022-12-27, Accepted: 2022-12-31

A Study on Verification of CCTV Image Data through Unsupervised Learning Model of Deep Learning

Yangsun Lee

Abstract: Abnormal behavior is called an abnormal behavior that deviates from the same normal standard as the average. The installation of public CCTVs to prevent crimes is increasing, but the crime rate is rather increasing recently. In line with this situation, artificial intelligence research using deep learning that automatically finds abnormal behavior in CCTV is increasing. Deep learning is a type of artificial intelligence designed based on artificial neural networks, and the quality of learning data is important for high accuracy in the development of artificial intelligence through deep learning. This paper verifies whether learning data for abnormal behavior detection is suitable as learning data which is being constructed using an MPED-RNN model for binary classification to determine whether there is an abnormal behavior by frame using skeleton data of a person based on an autoencoder. As a result of the experiment, the unsupervised learning-based MPED-RNN model used in this paper is not suitable for verifying images with a similar number of frames with and without abnormal behavior, such as the corresponding data, and it is judged that appropriate results can be derived only when verified with a supervised learning-based model.

Keywords: abnormal behavior; abnormal behavior detection; artificial neural network; deep learning; MPED-RNN; unsupervised learning model

1 INTRODUCTION

Abnormal behavior is called an abnormal behavior that deviates from the same normal standard as the average. The installation of public CCTVs (CCTV - close-circuit television) to prevent crimes is increasing, but the crime rate is rather increasing recently. In line with this situation, artificial intelligence research using deep learning that automatically finds abnormal behavior in CCTV is increasing. Deep learning is a type of artificial intelligence designed based on artificial neural networks, and the quality of learning data is important for high accuracy in the development of artificial intelligence through deep learning [4, 5, 12, 15].

This paper verifies whether learning data for abnormal behavior detection is suitable as learning data which is being constructed using an MPED-RNN model for binary classification to determine whether there is an abnormal behavior by frame using skeleton data of a person based on an autoencoder. As a result of the experiment, the unsupervised learning-based MPED-RNN model used in this paper is not suitable for verifying images with a similar number of frames with and without abnormal behavior, such as the corresponding data, and it is judged that appropriate results can be derived only when verified with a supervised learning-based model [8, 10, 11].

2 RELATED STUDIES

2.1 Autoencoder

The autoencoder is an unsupervised deep learning, consisting of two structures: encoder and decoder. The autoencoder learns through the process of minimizing the difference between the original data and the restored data after encoding the input data and restoring it again through the decoder [1, 2, 13]. Fig. 1 shows the structure of the autoencoder.

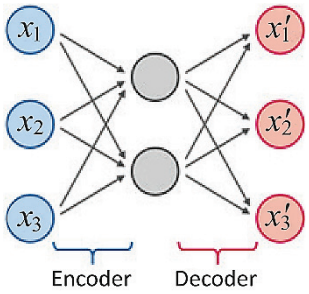


Figure 1 Autoencoder

2.2 RNN (Recurrent Neural Network)

RNN is a neural network that continuously uses the information of the previous step while repeating itself. We use historical information as a loop structure to improve the performance of neural networks on current inputs [6, 7, 9]. Fig. 2 shows the RNN structure.

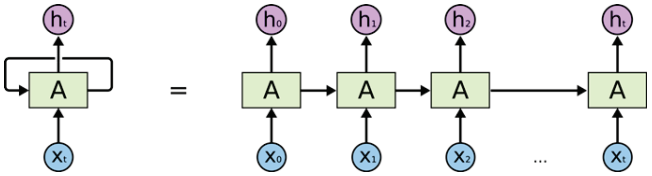


Figure 2 RNN structure

2.3 MPED-RNN Model

The MPED-RNN model is an autoencoder-based anomaly detection model with skeleton data input. The encoder-decoder has a repeated structure and features a temporal and spatial pattern of the skeleton trajectory [14, 16, 19, 20, 21]. Fig. 3 shows the structure of the MPED-RNN model.

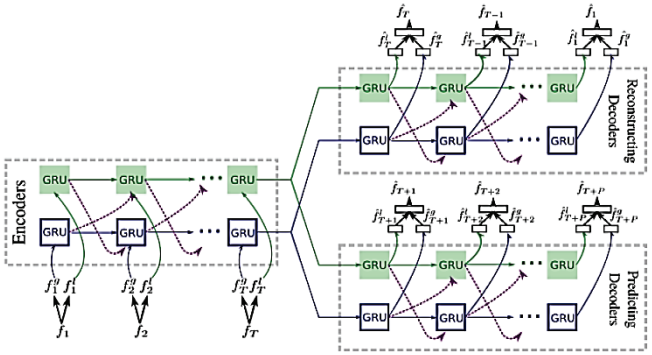


Figure 3 MPED-RNN model structure

The MPED-RNN Model learns by dividing the model's input, skeleton data, into global body movement which is information about large movements with little shape, size, and deformation, and local body posture which is information about fine movements such as internal deformation of skeleton movements. When an irregular pattern occurs during learning, the frame in which the pattern occurs is classified as an abnormal behavior.

2.4 AUROC (Area under the ROC Curve)

In the MPED-RNN model, the default output evaluation index is AUROC. AUROC represents the area under the ROC curve, a graph that corresponds to the vertical and horizontal axes of the True Positive Rate (TPR), which is the ratio that accurately predicted the normal, and false positive rate (FPR), which is the ratio that incorrectly predicted the normal. Fig. 4 shows the ROC curve. [3, 6, 7].

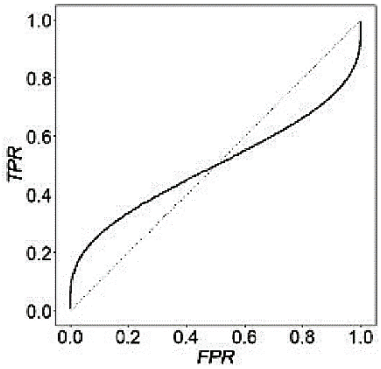


Figure 4 ROC curve

2.5 Unsupervised Learning

As a kind of machine learning, it falls into the category of problems that determine how data is composed. Unlike supervised learning or reinforcement learning, this method is not given a target value for the input [2, 11, 18, 22, 24]. Fig. 5 shows the unsupervised learning process.

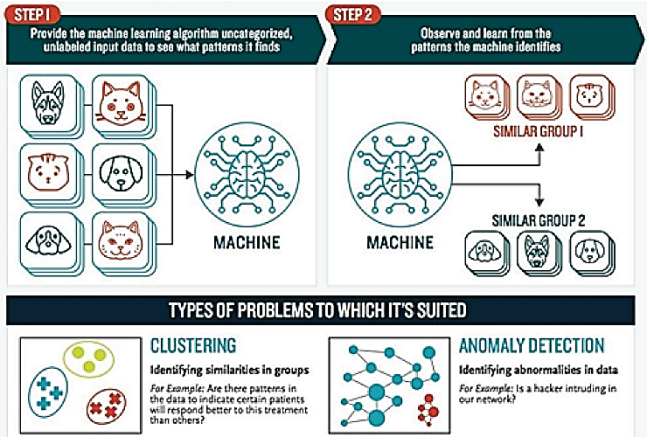


Figure 5 Unsupervised learning process

3 VERIFICATION OF LEARNING DATA WITH MPED-RNN

MPED-RNN, an anomaly detection model, uses skeleton data for each person in the video as learning data, and evaluation is conducted using skeleton data and a frame-level mask that expresses which frame the anomaly behavior occurred. Fig. 6 is a data verification system for learning conducted in this paper.

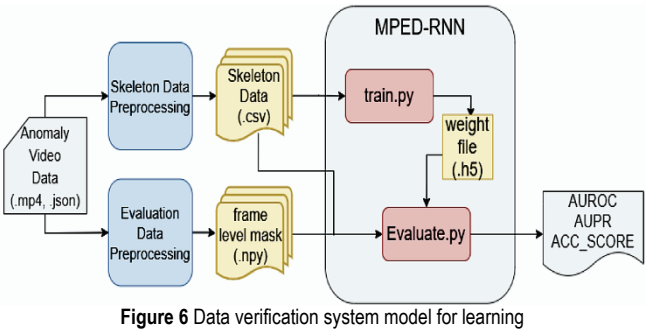


Figure 6 Data verification system model for learning

3.1 Preprocessing of Skeleton Data

In order to preprocess image data among learning data as skeleton information used as learning data in the MPED-RNN model, skeleton data was first extracted from the image. The extracted skeleton data is a JSON (JavaScript Object Notation) file with a frame number, a person number, and a person's joint coordinates as shown in Fig. 7. Fig. 7 shows the extracted skeleton data.


```
"frame_index": "0",
"objects": [],
"objects_changed": "no",
"persons": [
    {
        "index": "0",
        "keypoints": [
            "0.000000,0.000000",
            "877.502808,380.112000",
            "948.150146,380.220673",
            "974.620850,503.878204",
```


Figure 7 Extracted skeleton data

The input data of the MPED-RNN model is a csv file representing the trajectory of skeleton data for each person. Therefore, the extracted skeleton data were used to divide the files by person, and the frame in which the person appeared in each file and the coordinates of the 17 joints observed in the frame were stored in the form of a csv file. Fig. 8 shows a preprocessed skeleton file.

framenumber	point1_x	point1_y	point2_x	point2_y	point3_x	point3_y
2	1278	277	1366	318	1366	318
3	1272	277	1366	318	1366	318
4	1266	274	1366	318	1366	318
5	1260	271	1366	319	1366	319
6	1260	271	1363	318	1363	318
...	...	...	...	...	...	...
3597	1166	763	1186	798	1186	798
3598	1166	763	1184	798	1184	798
3599	1166	763	1186	798	1186	798
3600	1166	763	1186	798	1186	798
3601	1166	763	1186	798	1186	798

Figure 8 Preprocessed skeleton data

3.2 Preprocessing Evaluation Data

The frame_level_mask file used to evaluate abnormal behavior classification in the MPED-RNN model is a binary file that expresses 0 and 1 with and without abnormal behavior by frame. In order to produce a frame_level_mask of learning data, a start_frame_index in which abnormal behavior begins and an ends_frame_index in which the abnormal behavior ends was extracted from the annotation file provided with the learning image data to produce a binary file with the information. Fig. 9 shows a data annotation file for learning, and Fig. 10 shows a generated frame-level mask.


```
"block_detail": "A21",
"start_time": "00:00:59.400",
"main_object": "Ob0",
"block_type": "action",
"end_time": "00:01:53.157",
"block_index": "2",
"start_frame_index": "1800",
"end_frame_index": "3429",
"num_persons": "1"
```


Figure 9 Data annotation file for learning

3.3 Learning

Learning was conducted using the generated skeleton data as an input to an unsupervised learning model. 454 images out of a total of 572 images were used as learning data. Fig. 11 shows part of the learning data.

Fig. 12 shows the learning settings. Epoch proceeded to 20. One epoch refers to the forward pass/backward pass process for the entire data in an artificial neural network. In other words, in the model, a total of 20 learning are conducted on the entire data.

In the learning process, if the video is put in, it is separated for each frame and clustering is performed by grouping frames with similar skeleton values. If most frames have similar skeleton values, but there are frames with different skeleton values than other frames, we classify the frames as abnormal behavior and proceed with learning.

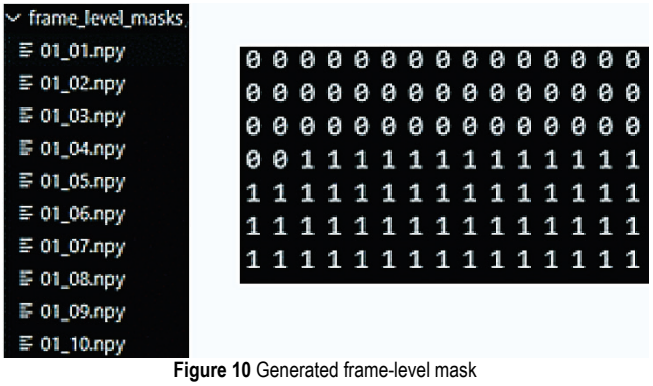


Figure 10 Generated frame-level mask

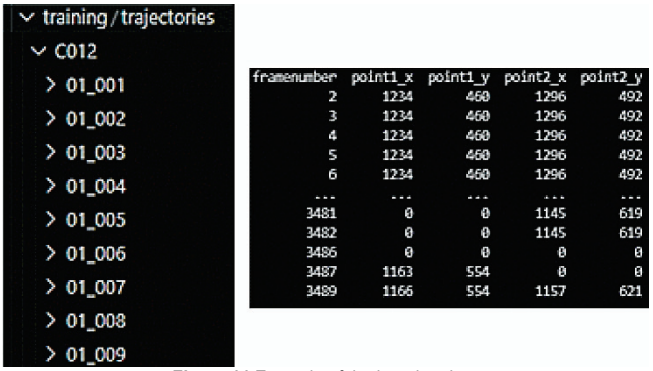


Figure 11 Example of the learning data

```
input_length 12
global_input_dim 4
local_input_dim 34
reconstruction_length 12
prediction_length 6
global_hidden_dims 8
local_hidden_dims 16
extra_hidden_dims
output_activation linear
reconstruct_reverse True
reconstruct_original_data True
multiple_outputs True
multiple_outputs_before_concatenation True
cell_type gru
optimiser adam
learning_rate 0.001
loss mse
```

Figure 12 Learning settings

3.4 Evaluation

Since the learning model of this paper is unsupervised learning, the intermediate result before AUROC output is predicted by frame. TPR(true positive rate) and FPR(false positive rate) were calculated based on an arbitrary classification point with the predicted value, and AUROC, the lower area of the green ROC curve, was output.

AUPR, the lower area of the precision-replay graph, was output with precision, which is the actual normal ratio among frames predicted to be normal, and reproduction, which is the

normal ratio among frames predicted to be normal. Fig. 13 shows an example of a predicted value for each frame and an output result.

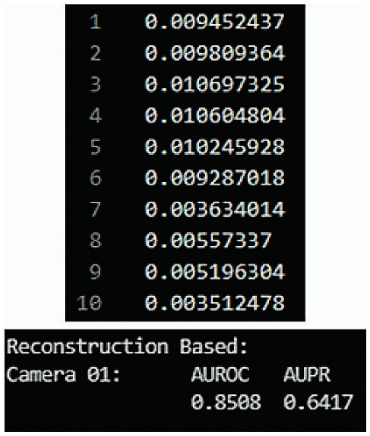


Figure 13 Predicted value for each frame and an output result

The model's evaluation method uses learned weights to quantify abnormal behavior for each frame, and then classify abnormal behavior using clustered values based on arbitrarily determined values. It is a method of extracting accuracy by comparing the classified binary file with the frame level mask, which is an answer binary file input by the user.

4 EXPERIMENTAL RESULTS AND ANALYSIS

The verification of learning data was conducted in a Geforce RTX 2080 environment with about 11 GB of memory. Fig. 14 shows a data source image for learning.

Figure 14 Data source image for learning

4.1 AUROC Results

Fig. 15 shows the verification results of learning data. The model's input data uses 17 joint coordinates, but 13 of the joint coordinates extracted from learning data were available. Therefore, as if the four uninputed joint coordinates were not observed, it is the result of entering the joint coordinates at the location most similar to the result of the learning by entering '0'.

Looking at the AUROC value, it was about 0.6. The value of AUROC in a binary classifier is from 0.5 to 1.0, and the binary classifier must have at least an AUROC value of 0.8 in order for it to be useful.

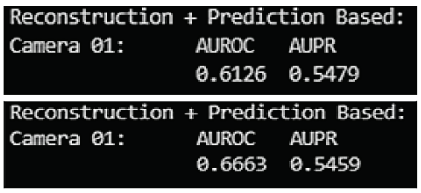


Figure 15 Verification results of learning data

4.2 Added Evaluation Index

In addition to AUROC and AUPR, which are essentially provided evaluation indicators in the MPED-RNN model, the numerical values of classification points were changed to find the optimal classification points for the specific section with the highest accuracy, and output the maximum accuracy. As can be seen from Fig. 16, it can be seen that the AUROC of the learning data is about 0.66 and the optimal classification point accuracy is about 0.65. The learning data shows lower accuracy than the HR-Avenue data used as the performance evaluation of the model.

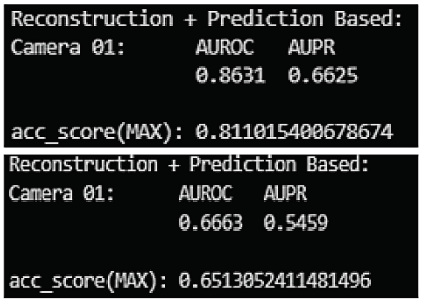


Figure 16 HR-Avenue and E2ON data verification results

However, this is not the low quality of the image, but in the case of the MPED-RNN model, which finds irregular patterns in the image and classifies them as abnormal behavior, since it is based on an autoencoder that performs unsupervised learning, if the number of frames in which abnormal behavior occurs in the learning data is similar to the number of frames in which abnormal behavior does not occur, the accuracy is lowered, and the accuracy of the learning data is lowered.

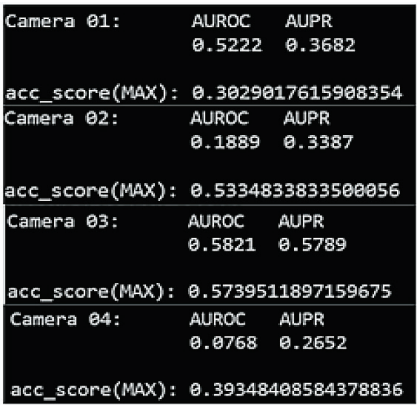


Figure 17 Abnormal behavior evaluation results

4.3 Abnormal Behavior Evaluation Results

Fig. 17 shows the evaluation results of child abuse, home invasion, theft, and vehicle theft learning data, respectively.

It was confirmed that the evaluation results were very low in the case of residential intrusion and vehicle theft abnormal behavior with more than 50% of the total number of frames of the learning video. Therefore, it is judged that the learning data is not suitable because the MPED-RNN model shows very low accuracy.

5 CONCLUSION AND FUTURE RESEARCH

This paper verified whether learning data for abnormal behavior detection is suitable as learning data through the MPED-RNN model. Due to the nature of the data, the accuracy was not high in the unsupervised learning-based MPED-RNN model, but it is judged as valid learning data in supervised learning-based models because the frame of precursor and abnormal behavior is clear and skeleton data extraction is accurate.

Currently, artificial intelligence technology is a technology that attracts attention among the 4th industrial revolution, and active research is being conducted, and many companies are trying to use it in industrial sites. However, since high-quality learning data for artificial intelligence development is difficult and difficult to build, it is believed that more learning data led by highly reliable national institutions can promote the development of artificial intelligence technology and popularization of artificial intelligence technology.

In the future, we will continue to conduct research on technology that verifies the data with other models based on supervised learning and applies abnormal behavior detection technology to public CCTVs.

Acknowledgements

This research was supported by Seokyeong University in 2022.

6 REFERENCES

[1] Bae, H.-J. et al. (2021). LSTM (long short-term memory)-based abnormal behavior recognition using AlphaPose. *KIPS Transaction on Software and Data Engineering, 10*(5), 187-194. https://doi.org/10.3745/KTSDE.2021.10.5.187

[2] Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures, workshop on unsupervised and transfer learning. *Proceedings of ICML Workshop on Unsupervised and Transfer Learning, PMLR, 27*, 37-50.

[3] Bradley, A. (1997). The use of the area under the ROC Curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

[4] Chalapathy, R. & Chawla, S. (2019). Deep learning for anomaly detection: A Survey. arXiv preprint arXiv: 1901.03407. https://doi.org/10.48550/arXiv.1901.03407

[5] Chandola, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41*(3), 1-58. https://doi.org/10.1145/1541880.1541882

[6] Fawcett, T. (2006). Introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

[7] Feng, K. (2019). Decision making with machine learning and ROC curves. arXiv: 1905.02810, 1-52. https://doi.org/10.2139/ssrn.3382962

[8] Haroon, U. (2022). A novel 3D-convolution neural network for human interaction recognition in videos. *The Journal of KING Computing, 18*(1), 19-28. https://doi.org/10.23019/kingpc.18.1.202202.002

[9] Hong, C. & Choi, S. (2020). ROC curve generalization and AUC. *Journal of the Korean Data & Information Science Society, 31*(4), 477-488. https://doi.org/10.7465/jkdi.2020.31.4.477

[10] Ji, S. (2013). 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(1), 221-231. https://doi.org/10.1109/TPAMI.2012.59

[11] Kiran, B. (2018). An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. *Journal of Imaging, 4*(36), 1-25. https://doi.org/10.3390/jimaging4020036

[12] Lee, J. (2020). A study on the implementation of intelligent abnormal behavior monitoring system using deep learning. *PhD Thesis*, Hanse Univ.

[13] Malhotra, P. (2016). LSTM-based encoder-decoder for multi-sensor anomaly detection. *ICML 2016 Anomaly Detection Workshop*, arXiv: 1607.00148. https://doi.org/10.48550/arXiv.1607.00148

[14] Morais, R. (2019). Learning regularity in skeleton trajectories for anomaly detection in videos. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 11996-12004. https://doi.org/10.1109/CVPR.2019.01227

[15] Park, H. (2013). A study on monitoring system for abnormal behaviors by object's tracking. *Journal of Digital Contents Society*, 14(4), 589-596. https://doi.org/10.9728/dcs.2013.14.4.589

[16] Park, S. Anomaly detection by a surveillance system through the combination of C3D and object-centric motion information, *Journal of KIISE, 48*(1), 91-99.

[17] RNN Structure, https://velog.io/@sksmslhy/RNN

[18] Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., & Langs, G. (2017). Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. *Information Processing in Medical Imaging, IPMI 2017. Lecture Notes in Computer Science, 10265*. Springer, Cham. https://doi.org/10.1007/978-3-319-59050-9_12

[19] Tran, D. (2015). Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 4489-4497. https://doi.org/10.1109/ICCV.2015.510

[20] Wang, P. (2018). 3D shape segmentation via shape fully convolutional networks, computers & amp. *Graphics, Elsevier BV*, 76, 182-192. https://doi.org/10.1016/j.cag.2018.07.011

[21] Wu, T. & Lee, E. (2019). Human action recognition based on 3D convolutional neural network from hybrid feature. *Journal of KMMS*, 22(12), 1457-1465. https://doi.org/10.9717/kmms.2019.22.12.1457

[22] Xu, H. (2018). Unsupervised anomaly detection via variational auto-encoder for seasonal KPIs in web applications. *Proceedings of the 2018 WWW Conference*, 187-106. https://doi.org/10.1145/3178876.3185996

[23] Osama, M. (2021). Behavior Recognition based on signal processing technology. *International Journal of Hybrid Innovation Technologies, 1*(2), 75-90.

https://doi.org/10.21742/ijhit.2653-309X.2021.1.2.05
[24] Rashid, E. (2016). Software Fault Prediction Using Unsupervised Learning Technique: A Practical Approach. *International Journal of u - and e - Service, Science and Technology, NADIA, 9*(11), 275-288, https://doi.org/10.14257/ijunnesst.2016.9.11.24.

Author's contacts:

Yangsun Lee, Professor
Seokyeong University,
124 Seogyeong-ro Seongbuk-gu,
Seoul, 02173, Korea
yslee@skuniv.ac.kr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221228154330

Preliminary communication
Received: 2022-12-28, Accepted: 2022-12-31

Text Classification Based on Neural Network Fusion

Deageon Kim

Abstract: The goal of text classification is to identify the category to which the text belongs. Text categorization is widely used in email detection, sentiment analysis, topic marking and other fields. However, good text representation is the point to improve the capability of NLP tasks. Traditional text representation adopts bag-of-words model or vector space model, which loses the context information of the text and faces the problems of high latitude and high sparsity,. In recent years, with the increase of data and the improvement of computing performance, the use of deep learning technology to represent and classify texts has attracted great attention. Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and RNN with attention mechanism are used to represent the text, and then to classify the text and other NLP tasks, all of which have better performance than the traditional methods. In this paper, we design two sentence-level models based on the deep network and the details are as follows: (1) Text representation and classification model based on bidirectional RNN and CNN (BRCNN). BRCNN's input is the word vector corresponding to each word in the sentence; after using RNN to extract word order information in sentences, CNN is used to extract higher-level features of sentences. After convolution, the maximum pool operation is used to obtain sentence vectors. At last, softmax classifier is used for classification. RNN can capture the word order information in sentences, while CNN can extract useful features. Experiments on eight text classification tasks show that BRCNN model can get better text feature representation, and the classification accuracy rate is equal to or higher than that of the prior art. (2) Attention mechanism and CNN (ACNN) model uses the RNN with attention mechanism to obtain the context vector; Then CNN is used to extract more advanced feature information. The maximum pool operation is adopted to obtain a sentence vector; At last, the softmax classifier is used to classify the text. Experiments on eight text classification benchmark data sets show that ACNN improves the stability of model convergence, and can converge to an optimal or local optimal solution better than BRCNN.

Keywords: attention mechanism; deep learning; neural network; presentation; text classification

1 INTRODUCTION

1.2 Background and Significance

Today, people are more and more accustomed to frequently expressing their views and sharing their lives in the online world, involving a wide range of contents. In order to satisfy the users' expression needs, Tik Tok, Weibo, Tieba, etc. seized the opportunity, accumulated a mass of users, and produced a wealth of data with rich contents [1-3]. This kind of data grows and updates very fast and there are many ways to obtain it. The problem that people have to solve is no longer the acquisition of information, but the extraction of useful information. How to extract information quickly and effectively will become an important research topic. Compared with sound and image as carriers of information, text uses less network resources and is easy to transmit [4-6]. Effective understanding and analysis of the profound meaning of a text is called text representation. The traditional text representation method is to manually select features to label large-scale original text sets, which is tedious and difficult to achieve the desired results, and even more difficult to meet the information processing needs of users. How to make users get the required information accurately has become the focus of research. Text classification can not only truly reflect the text information by finding a suitable structured text representation method, but also have corresponding distinguishing ability for different texts [7-10]. Text classification automatically marks texts according to some classification systems or a standard, which is common in the fields of text subject, emails detection, public opinion analysis and so on, and has important practical significance in the efficient management and effective utilization of text information. The traditional text representation method has irreparable shortcomings, which makes it difficult to achieve the expected performance in practice. Therefore, the research of text representation and text classification is a subject with great theoretical value and practical needs. Since 1990s, statistical machine learning has made great progress in the research of part-of-speech tagging, syntactic analysis, named entity recognition and other topics. Traditional text representation methods have gradually become the mainstream text representation methods. However, BOW's transformation of words into one-hot vectors ignores word order information and cannot reflect the internal relationship between words, which leads to semantic loss. TF-IDF suppresses the weight of high-frequency words in order to highlight low-frequency words, but exaggerates the importance of uncommon words. Common words are not equal to meaningless words. Text Rank's graph-based sorting algorithm extracts keywords to represent the text by voting, which is better for long texts than short texts, and the semantic loss of the text representation results is serious. LDA uses the topic distribution of the text and words in the topic to extract keywords to represent the text, which will also cause serious information loss, and has higher requirements for the topic dictionary. In order to solve the problems such as the lack of text information, specific keywords are often artificially selected to supplement the text information according to the specific situation. Although it has a certain effect, it makes the application scope of the model cramped and difficult to popularize, and it is even more difficult to face massive data. Common machine learning classification methods are: Logistic Regression (LR), which is prone to under-fitting or over-fitting, and the classification accuracy, is not ideal; Support Vector Machine (SVM)'s order is equal to the number of training samples. When there are many samples, it will be very difficult to calculate and store the matrix; KNN algorithm, the value of super parameter K has a great influence on the classification results. KNN has a high demand for data distribution, and it

wants the data to be clustered, so the classification effect is very poor when dealing with spiral data.

As for sentence-level text representation, with the increase of data and the improvement of computing performance, the use of deep learning techniques to represent and classify texts has attracted much attention. For example, NLP tasks such as using CNN [11-14], RNN [15-18, 26] and attention mechanism [19, 20] to represent documents and then classify texts have better performance than traditional methods. These neural network-based text representation and classification methods have been proved to be very effective. This is very important to deal with massive data in text form. It reduces the labor cost and improves ext classification accuracy and speed. However, compared with the achievements of deep neural network model in the area of computer vision, the application of deep neural network in the area of NLP is just beginning. In image classification, compared with the traditional machine learning method, the deep neural network model can reduce the error by more than 10%. In the competition of ImageNet Large Scale Recognition Association, the error of the deep neural network model has been reduced to 3.57%, which has exceeded the recognition ability of people. However, in natural language processing, such achievements can't be achieved at present. In recent years, experts and scholars in the field of natural language processing are also looking for ways to use deep neural network model to improve the performance of natural language processing tasks. Therefore, it is still a worthy research direction to use deep neural network for NLP.

1.3 Related Work

1.3.1 Research Status of Text Representation

In the long history of the development of human characters, characters are gradually different from sounds and images. Characters are no longer a simple signal, but a very abstract concept. To understand human words, we can't rely on logic alone, but also need a very strong knowledge base. The essence of text is a collection of a large number of words, which is accompanied by information such as spoken language, popular words, abbreviations, spelling mistakes and even emojis. It is extremely difficult for the existing text classification algorithms to directly identify this set, so it is necessary to convert it into a unified representation that can be easily recognized by the text classifier, that is, text representation. In 2005, Zhou proposed a graph-based text representation method. The text is converted into text features according to the established rules, and these features are synthesized to define the similarity used to calculate the measurement chart. In this method, the word order relationship in the text is taken into account, but many manually set parameters are included in the configuration process. In 1998, Salton put forward the Vector Space Model (VSM), which uses vectors to represent text features. Related research issues mainly focus on feature selection and weight calculation. Set a threshold value for each feature, remove the features whose support is less than the threshold value, and regard the rest features as valid features. Common feature selection methods are: chi-square statistics, text frequency, mutual information, expected cross entropy, etc. Usually, the corresponding weights are calculated based on the frequency of the selected features. VSM ignores the relationship between words and text contexts, which leads to the loss of text information. In recent years, although the Bag of Word (BOW) model has been widely used in text representation, BOW defaults that each word is independent, ignoring the relationship between words and the information between text contexts. Moreover, the expression vector of words is generally in the form of one-hot coding vector, and the vector dimension is the size of the bag of words, which makes the result of text feature representation face the problem of dimension disaster and high sparseness. Word embedding model gradually replaces word bag model, which is used for vector representation of words. The essence of word vector is the first full connection layer parameter of deep neural network, in which unique hot coding is used as input. The closer the word meaning is, the higher the similarity of the corresponding word vectors in feature space, and vice versa. Although the word embedding model solves the problem of high sparseness and reduces the vector dimension compared with the bag of words model, the dimension is still very high, which makes it difficult to show its advantages in traditional classification algorithms. In 2002, Lai proposed the RCNN model, using the advantages of RNN and CNN. Because RNN itself is a sequence model, for the text feature sequences represented by word vectors, RNN processing can extract the context information of the text well, and CNN processing can obtain a number of local information and greatly reduce the amount of calculation, so RCNN has achieved good results in text classification; In 2014, Shen embedded the text representation based on word vectors, and combined with CNN to mine the high-level semantics of the text; In 2014, Santos proposed DCNN model for sentiment classification based on word vectors of characters, words and sentences. The model achieved good results, but it was difficult to build a depth model when the text length was not fixed. To sum up, appropriate text representation and appropriate depth model are the basis of text classification.

1.3.2 Research Status of Text Classification

Text classification is a process of labeling each sample with appropriate category according to relevant features. Text classification methods are mainly based on pattern system or machine learning [27]. The classification method based on pattern system uses knowledge engineering technology and professional assistance to design appropriate classification rules for each category. If the text conforms to the corresponding rules, it is considered that the text belongs to that category. The text features of knowledge engineering refer to the related attributes of the established rules in the text. Due to the role of artificial judgment, a good accuracy rate is obtained, but the defects of the pattern system are obvious. The classification rules determine the quality of the classification results, and the formulation of rules requires a lot of research and demonstration by experts. However, these rules cannot be used across fields, or even transferred to

different tasks in the same field, which leads to the basic lack of popularization of knowledge engineering and restricts knowledge. Classification models based on statistical machine learning, including Naive Bayes (NB), KNN, SVM, neural network, decision tree, etc. [21, 22]. The classification performance of NB, SVM, decision tree and KNN is poor, so some people improve it. Zhou Zhihua et al. put forward the selective integration theory in 2002, and proved the superiority of integrated learning system, and achieved good results in text classification. However, the above models are all shallow machine learning methods, which can deal with simple classification problems. Its generalization ability is often not strong, and the phenomenon of over-fitting or under-fitting often occurs in the face of high-dimensional feature data. Therefore, it is a feasible direction to explore the application of deep model in classification.

1.3.3 Research Status of Deep Neural Networks

DL (Deep learning) combines low-level features with nonlinear transformation to form more abstract high-level features, so that the model can better learn the distribution law of data [23]. In 1986, Rumelhart and Hinton proposed the back propagation algorithm, which made the neural network change from a simple model to a complex model, making great contributions to the development of deep learning. In 1998, Lecun et al. used convolution neural network to reduce parameters and calculations and improve the performance of model training by using local perception and weight sharing. In 2000, Hinton put forward the learning algorithm of contrast divergence, and in 2006, he put forward the Restricted Boltzman Machine (RBM). Through layer-by-layer training, the problem of deep learning model optimization is solved, which makes DL develop rapidly. In 2007, Alex Graves et al., based on Long Short-Term Memory (LSTM), recognized handwritten characters, and Sutskever et al., in 2014, proposed to build a machine translation framework by using two multilayer LSTM network structures. In 2008, Vincent proposed Auto-Encoder (AE), which is a neural network to reconstruct the input information, hoping that the input is equal to the output. Its basic structure consists of input layer, hidden layer and output layer. The input data is the original data. After the change of the hidden layer, the output results are as consistent as possible with the input data. Then the parameter matrix of the hidden layer can be used as the characteristics of the original data to achieve better results. Some researchers have made some improvements to AE. The denoising self-coding improves the anti-interference ability of data, and the variational self-coding changes the sample distribution to generate new samples. In 2011, Socher and others proposed a Recursive Neural Network to predict the tree structure. In 2012, when DL was used in Imag Net tasks, the error rate dropped from 26% to 15%. DL developed from the field of speech to the field of image recognition in academic circles. In 2013, Mikolov proposed sequence-based deep neural network RNN [24, 25]. The model ingeniously adds self-connection and interconnection in the hidden layer, and has certain memory ability. In 2015, Hinton et al. said on Nature: "In the next few years, DL will have a huge impact on the field of natural language understanding". In 2017, Young et al. compared DL models in various NLP fields and analyzed the possible trends in the future.

1.3.4 Research Status of Deep Learning in NLP

NLP is another important application field of deep learning. In the field of NLP, the use of neural network model for word embedding, RNN and CNN for text classification and translation tasks have made great progress. In 2000, Xu first put forward the idea of using neural network to train language model. A language modeling method using three-layer neural network is introduced in detail in the literature. Literature puts forward the idea of hierarchy to replace the matrix multiplication from hidden layer to output layer in the method of literature, which reduces the amount of calculation when the effect is equal to that of literature. Colobert and Weston introduced their word vector calculation method in, and systematically introduced their work in. Huang et al. improved the model proposed by Colobert et al. in, using full-text information to assist local information. Another innovation of Huang et al. is learning multiple word vectors to represent polysemous words. Mikolov proposed CB0W and Skip-gram models, and in proposed negative sampling and hierarchical softmax to improve the calculation speed of the models. Rong analyzed the learning process of the parameters of Mikolov's word vector model in detail in. Jeffrey et al. put forward a method of word vector representation by decomposing the global "word-word" co-occurrence matrix in Glove model. These methods are all based on the distribution representation of distribution hypothesis, which can map words into low-dimensional vector space without losing the correlation between words. They are the most advanced word embedding methods at present. The method of text representation and classification at sentence level or document level is also a research hotspot. Hill et al. proposed to learn the distributed representation of sentences from unlabeled data. Conneau et al. proposed using supervised learning to learn general sentence representation from natural language reasoning data. And lemikolov in proposed learning paragraph vectors for classification. Illinois-LH is a distribution table system for learning semantics. Kim proposed a word-level CNN text classification model. Sochei et al. proposed using deep RNN model for text classification. Dai and Lip came up with using semi-supervised learning to improve sequence learning. Kalchbrenner et al. used CNN to model sentences. Skipthouht proposed by Kiros et al. and F-Dropout proposed by Wang and Manning have achieved good performance. Machine translation is one of the best tasks combining deep learning with natural language processing, and many machine translation models based on deep neural networks have achieved good performance. Cho et al. used RNN encoder-decoder to learn the phrase representation for statistical machine translation. Cho et al. in introduced the properties of the encoder-decoder method of neural machine translation. Devlin et al. applied neural network language model to the decoding process of machine translation

framework. Sutskever et al. used the SEQ 2 SEQ model of a single neural network. On the basis of sequence-to-sequence machine translation model, Bahau et al. added attention mechanism to automatically learn the aligned words of the target words in the source language. Dong et al. used the multi-task learning method to jointly learn the translation model from one source language to multiple target languages. Meng et al. used neural turing machine to build a deep memory neural network. Kalchbrenner et al. introduced a neural machine translation model with linear time complexity. These translation models based on deep neural networks have achieved better performance than traditional translation models. To sum up, in general, the achievements of DL in NLP are far from those in image and speech recognition, and there is still much room for research.

2 RELATED WORK

2.1 Text Classification Method

Text classification refers to the process of classifying a large number of texts into one or more categories. Text classification is not much different from general classification problems. It is based on the characteristics of the samples to be classified to judge and choose the best classification result. The earlier text classification method is the matching method, which focuses on whether there are the same words or words with the same meaning in the text [28]. This factor is used to analyze the category of the text. This method is too simple and narrow to get satisfactory classification results. At present, the mainstream text classification method is machine learning method based on statistics, which takes the data set of known categories as the training set. According to the effective training classifier in the training set, the text of unknown category is classified and predicted. Compared with knowledge engineering technology, statistical learning method has been used as a common method for classification problems, because there are a large number of technologies with detailed theoretical basis, and there are many subjective factors of knowledge engineering experts. Traditional classification methods include KNN, decision tree, ensemble learning, etc.

2.2 Overview of Deep Neural Networks

Deep learning is a method of learning data distribution, which is mainly used in interpreting and analyzing image, sound and text data. In recent years, deep learning has made great achievements, and then it has been widely studied and applied. Here are some common deep learning models.

2.2.1 RNN

Recurrent Neural Networks consists of 3 layers, input layer (x), hidden layer (s) and output layer (o). As shown in Fig. 1, self-connection and interconnection are added to the hidden layer of its model, which enables RNN to acquire short-term memory properties and be widely used to process sequence features.

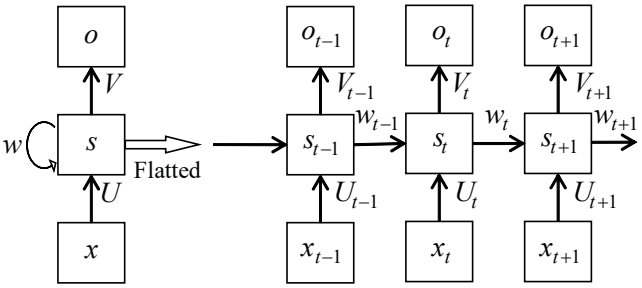


Figure 1 Basic structure of rnn network

Each node in Fig. 1 represents a unit. RN is a sequence model, x_{t-1}, x_t, x_{t+1} is an input sequence, o_{t-1}, o_t, o_{t+1} is an output sequence and s_{t-1}, s_t, s_{t+1} is a memory sequence of the hidden layer.

$$S_t = f\left(U * X_t + U * X_{t-1}\right) \tag{1}$$

$$O_t = soft\max\left(VS_t\right) \tag{2}$$

For the node at time t of RNN:

(1) S_t is in a hidden state, capturing the information on the node at the previous moment.

(2) o_t is obtained from the memory of the current time node and all previous time nodes.

(3) However, s_t can't capture the information of all nodes after T time.

(4) Each memory cell shares a set of parameters (U, V, w), which greatly reduces the amount of calculation.

(5) In many cases, o_t is not output, only the result of the last moment of the sequence is output.

However, there is not only one common network structure of RNN. Several common RNN structures are shown in Fig. 2.

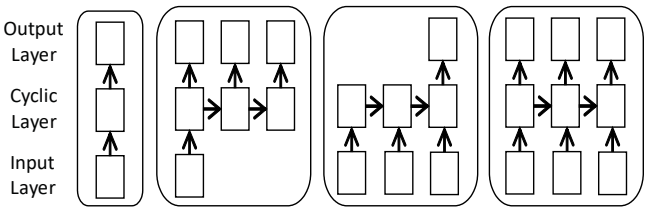


Figure 2 Common structure of cyclic neural network

In addition, RNN can be divided into two types: static and dynamic, depending on whether the number of cells needs to be set in advance. The static circular neural network requires that the number of hidden units of the preset circular neural network should not be changed, and the length of the input text sequence should be determined strictly according to the number of hidden units, that is to say, the length of the input text sequence must be consistent. Dynamic neural network doesn't need to set the number of hidden units of cyclic neural network in advance, but only automatically adjusts hidden units' amount according to the length of input sequence, that is to say, it doesn't need the same text length. However, dynamic RNN is not very effective in dealing with the problems of long text and text coding and decoding.

2.2.2 CNN

CNN has achieved great success in computer vision, and it has been deeply studied and widely used. From Lenet, Alex Net, VGG Net, Inception (Google Net) and later Rest Net, the network has become deeper and deeper, but its performance has become more and more prominent. In recent years, CNN has become popular in the field of text processing. The essence of convolution is an integral transformation. Generally, multiple convolution kernels are often used in convolution operation, which makes the number of channels of the obtained feature matrix more than that of the original feature matrix. After finishing, the dimension of the feature matrix will become larger and larger, which not only does not simplify the problem, but also requires more parameters and more memory in the subsequent calculation. Therefore, pooling operation is widely used to solve this problem. Pooling is used in the non-overlapping areas of the matrix, which is equivalent to an abstract process, filtering out unnecessary information, and generally, there are three kinds of pool operations:

(1) Maximum pool: the maximum value in the pool window is regarded as the value of this area;
(2) Average pooling: the average value in the pool window is regarded as the value of this area.
(3) Random pool: randomly select the values in the pool window.

Among them, the largest pool operation is the most widely used in text processing.

3 EXPERIMENTAL PROCESS

3.1 Experimental Principle

Usually, the CNN is composed of multiple convolutional layers, and each convolutional layer usually performs the following operations: Pool the output result of the activation function (usually using the maximum pool operation) to get the most significant features,

These steps constitute a common convolution layer, and sometimes local response normalization (local re-spone normalization (LRN) operation. The superposition of multiple convolution layers can obtain higher-level features. Brcnn model firstly combines loop operation, convolution operation and pool operation to extract the features of the text and express the text as feature vectors; then, the full connection layer is used to transform the feature space, and then the Softmax classifier is used for classification. The model structure diagram is shown in Fig. 3.

The input of the text classification part is the sentence vector M obtained from the representation part, which is mainly used to predict the category of the sentence.

(1) Full connection layer feature space conversion

Here, a fully connected layer is used to transform the feature space.

$$O_f = relu\left(W_f^{\mathrm{T}} m + b_f\right) \tag{3}$$

The input of the whole connection layer is the sentence vector learned at the end of the convolution layer $m \in R^M$, and the output is a transformation with the same dimension as the input $O_f \in R^M$. $W_f \in R^{M\times M}$ and $b_f \in R^M$ are available parameters for learning. The activation function used here is *relu*. To prevent over-fitting, drop operation is added.

(2) Output layer classification result prediction

There is a full connection mapping from the full connection layer to the output layer:

$$y = W_o^T O_f + b_o \tag{4}$$

The input of this layer is the output of the full connection layer, and the output is the predicted value belonging to each category $y \in R^C$, where C is the number of target categories $W_o \in R^{M\times C}$, $b_o \in R^C$ is the weight available for learning. To have an intuitive understanding, use softmax to normalize y to get the predicted probability of each category $p \in R^C$, where the formula of the i^{th} component p_i in p is as follows:

$$P_i = \frac{y_i}{\sum_{c=1}^{C} y_c}. \tag{5}$$

3.2 Data Set

In this experiment, eight benchmark data sets are used, as shown in Tab. 1: C corresponds to the number of target categories in the data set, L corresponds to the average length of sentences in the data set, N corresponds to the number of samples in the data set, V is the vocabulary size of the data set, and Test is the sample number of the test set, and CV means cross-validation with 10% discount.

Table 1 A summary of the datasets

Data Set	C	L	N	V	Test
MR	3	19	10,660	18,765	CV
Subj	2	24	10,000	21,323	CV
SST	6	17	11,860	17,836	2,210
SST2	3	19	9,600	16,185	1,821
IMDB	3	231	50,000	39,200	25,000
TREC	7	11	5,952	9,125	500
CR	2	19	3,741	5,340	CV
MPQA	2	3	10,504	6,246	CV

The specific description of each data set is as follows:

MR: MR is movie review data marked by Pang et al. for emotion classification.

Set, which was first used in reference. Each sentence in the data set corresponds to a comment. This data set is divided into positive/negative reviews, and it is a two-category task data set, in which there are 5331 positive examples and 5331 negative examples, with a total of 10662 reviews. The average length of sentences in this data set is 20, and the vocabulary size is 18,765 words. In this experiment, 10-fold cross-validation is used.

Subj: sub (objectivity) is a sentiment analysis data set containing 5,000 subjective and 5,000 objective sentences, which was used for the first time by Pang et al. in literature. This is also a data set with two target classes (subjective and objective :), that is, a binary data set. Similarly, each sentence corresponds to a sample. The average sentence length is 23, and the number of vocabulary words is 21,323. In the experiment of this paper, the 10-fold cross-validation is also used.

SST: SST (Stanfod Sentimental Treebank) is marked and published by Socher et al., which is an extension of MR. This data set includes a total of 11,855 movie reviews, which are labeled as five categories (very positive, positive, neutral, negative and very negative), that is, it is a five-category data set. This data set provides the segmented training set (8544), calibration set (1101) and test set (2210).

SST: In SST2, remove neutral comments in SST, and combine very positive and positive into positive, and very negative and negative into negative. The final SST2 contains a total of 9163 samples, including 7792 samples in the training set and 1821 samples in the testing set. Similarly, this is a binary classification task with two target classes.

MDB: IMDB data set is a two-class sentiment analysis data set, including 50,000 samples, 25,000 samples in training set and 25,000 samples in testing set, which was marked and published by Maas et al. This is a binary data set, which contains more data than the previous benchmark data set. Additional unlabeled data can also be used.

TREC: TREC is the problem classification task data set T97 marked by Li et al. This data set is divided into training set and testing set and the training set is randomly divided into training sets of 1000, 2000, 3000, 4000 and 5500 samples. This paper uses the training set with 5500 training samples.

CR: CR is customer reviews M1 of 14 products obtained from Amazon annotated by Humin Qing et al. Its task is to classify each customer's review into positive and negative categories.

3.3 Comparison and Analysis

This section introduces the comparison and analysis of experimental results in detail. Including BRCNN and ACNN and existing modules

Performance comparison and analysis of BRCNN, ACNN and their variants, multilayer.

Comparison and analysis of performance of BRCNN and ACNN, comparison and analysis of dropout strength of BRCNN circulating layer. Comparison and analysis of accuracy, recall and F1 value, P-R curve analysis, confusion matrix analysis, ROC curve and AUC analysis, etc. Tab. 2 introduces BRCNN and ACNN, their variants and existing models in eight benchmark data sets.

Tab. 2 shows the comparison of the accuracy of BRCNN and ACNN and their variants with the existing models on MR, SUBJ, TREC, CR and MPQA data sets. BRCNN and ACNN based on RNN have almost the same accuracy rate as the existing methods, while BRCNN and ACNN based on LSTM and GRU have the same or higher accuracy rate as the existing models. Particularly, on the TREC data set, the BRCNN based on LSTM designed in this paper reduces the error rate by 23.4%; GRU-based BRCNN reduces the error by 26.7%; ACNN based on RNN reduces the error rate by 6.3%; ACNN based on LSTM reduces the error by 34.4%; ACNN based on GRU also reduced the error rate by 26.7%. For CR data sets, BRCNN based on LSTM reduces the error rate by 7.3%; ACNN based on LSTM also reduces the error by 8.8%. ACNN based on LSTM reduces the error by 1.8% on MR data sets, while BRCNN based on LSTM also achieves the second highest performance. As for Subj data set and MPQA data set, BRCNN and ACNN also got the second highest accuracy.

What is listed in Tab. 3 is the comparison of the correct rates on SST data set and SST2 data set. Br CNN and ACNN also got the second highest and the third highest accuracy respectively. SST data sets are divided into five categories. On the other hand, SST2 data set is a binary classification task based on SST data set, which removes the neutrality and combines the positive and negative data sets. Therefore, the classification features are more obvious, so the classification results are better than SST data set.

Table 2 Results of our models and the state-of-the-art models on MR> Subj TREC CR and MPQA datasets

Model	MR	Subj	TREC	CR	MPQA
BRCNN(RNN)	81.6	93.1	92.4	80.0	90.3
BRCNN(LSTM)	83.1	94.2	95.1	87.3	90.7
BRCNN(GRU)	82.6	94.1	95.3	85.6	90.8
ACNN(RNN)	81.6	93.3	94.0	86.1	90.3
ACNN(LSTM)	83.4	95.2	95.8	87.5	90.5
ACNN(GRU)	82.7	94.1	95.3	86.9	90.8
FastSent	70.8	89.0	76.8	78.4	80.6
FastSent+AE	71.8	88.8	80.4	76.7	81.5
SkipThought	76.5	93.6	92.2	80.1	87.1
SkipThought-LN	79.4	93.7	88.4	83.1	89.3
ParagraphVec(DBOW)	60.2	76.3	59.4	66.9	70.7
SDAE+embs.	74.6	90.8	78.4	78.0	86.9
Unigram-TFIDF	74.7	90.3	85.0	79.2	82.4
CNN-static	81.0	93.0	92.8	84.7	89.6
CNN-non-static	81.5	93.4	93.6	84.3	89.5
CNN-multichannel	81.1	93.2	92.2	85.0	89.4
Paragraph-Vec	75.8	90.5	91.8	78.1	74.2
MNB	79.0	93.6	-	80.0	86.3
word2vecBOW	77.7	90.9	83.6	79.8	88.3
fastTextBOW	76.5	91.6	81.8	78.9	87.4
GloVeBOW	78.7	91.6	83.6	78.5	87.6
GloVePositionalEncoding	78.3	91.1	83.3	77.4	87.1
BiLSTM-Max(untrained)	77.5	89.6	85.8	81.3	88.7
CaptionRep(bow)	60.9	77.4	72.2	69.3	70.8
DictRepBOW+embs.	76.7	90.7	81.0	78.7	87.2
NMTEn-to-Fr	64.7	84.9	82.8	70.1	81.5
BiLSTM-Max(onSNLI)	79.9	92.1	88.7	84.6	88.8
BiLSTM-Max(onA11NLI)	81.1	92.4	88.2	86.3	90.2
NaiveBayes-SVM	79.4	93.2	87.4	81.7	86.3
RNN	77.2	82.3	90.2	82.3	90.1
BRNN	82.3	94.2	91.0	82.6	90.3
AdaSent	83.1	95.5	92.4	86.3	93.3
G-Dropout	79.0	93.4	93.4	82.1	86.1
F-Dropout	79.1	93.6	89.7	81.8	86.3

What is shown in Tab. 4 is the comparison of the correct rates on IMDB data sets. The accuracy of BR-CNN and ACNN on IMDB data sets is 91.1%, which is higher than

most existing models. Compared with paragraph-Vec, SA-LSTM and SEQ 2-bowen-CNN, the accuracy performance of BRCNN and ACNN is not good enough. As can be seen from Tab. 4, the average sentence length of IMDB data set is 231. And in IMDB data set, a sample corresponds to multiple sentences. BRCNN and ACNN are more effective for single-sentence short texts, while paragraph and other methods are designed to handle long texts, so they have better performance on IMDB data sets than BRCNN and ACNN. BRCNN and ACNN can also get higher accuracy on data sets that they are not good at, so BRCNN and ACNN can still extract sentence features very well. To sum up, the models BRCNN and ACNN designed in this paper can well extract the feature information of sequence data, convert sentences into corresponding vector representations, and get better sentence representations, so as to achieve higher or equivalent accuracy than the existing models in the classification task. This shows the feasibility of BRCNN and ACNN.

Table 3 Accuracy of models on the SST and SST2 datasets

Model	SST	SST2
BRCNN(RNN)	463	86.0
BRCNN(LSTM)	49.0	87.8
BRCNN(GRU)	47.6	87.4
ACNN(RNN)	46.6	86.3
ACNN(LSTM)	49.1	87.8
ACNN(GRU)	48.9	87.5
SkipThought	47.9	82.0
SkipThought-LN	48.3	82.9
CNN-static	45.5	86.8
CNN-non-static	48.0	87.2
CNN-multichannel	47.4	88.1
RCNN	47.2	87.6
RNTN	45.7	85.4
DCNN	48.5	86.8
Paragraph-Vec	48.7	87.8
word2vecBOW	47.6	79.7
fastTextBOW	46.7	78.8
GloVeBOW	45.5	79.8
GloVePositionalEncoding	43.6	80.6
BiLSTM-Max(untrained)	44.9	80.7
BiLSTM-Max(onSNLI)	42.5	83.3
BiLSTM-Max(onA11NLI)	43.3	84.6
NaiveBayes-SVM	45.6	83.1

Table 4 Accuracy of models on the IMDB dataset

Models	Accuracy
BRCNN(RNN)	88.0
BRCNN(LSTM)	91.1
BRCNN(GRU)	91.0
ACNN(RNN)	88.4
ACNN(LSTM)	91.5
ACNN(GRU)	93.1
MNB	86.9
G-Dropout	91.2
F-Dropout	93.1
NBSVM	91.2

Tab. 5 is a comparison of the performance of multi-layer BRCNN and multi-layer ACNN on TREC data sets. With the increase of layers, the accuracy of BRCNN increases, but that of ACNN hardly increases. However, ACNN on the first layer has achieved the same accuracy as that of multi-layer BRCNN. This further shows that ACNN can extract the features of the text better and get a better sentence feature representation (that is, get a better sentence vector), so it can get a better classification result in the classification task.

Table 5 Accuracy of ACNN and BRCNN models on the TREC dataset

Models	Accuracy
BRCNN(1 Bi-LSTM layer)	95.5
BRCNN(2 Bi-LSTM layers)	95.7
BRCNN(3 Bi-LSTM layers)	95.7
ACNN(1 Bi-attention layer)	95.7
ACNN(2 Bi-attention layers)	95.7
ACNN(3 Bi-attention layers)	96.3

4 CONCLUSION

Text classification has always been a hot topic in NLP field, and how to express text into digital features is the key of text processing. The main work of this paper is also devoted to the research of text representation and classification methods. In this paper, two text representation and classification models based on deep neural networks are designed: BRCNN and ACNN. The feasibility of BRCNN and ACNN is verified by experiments. This paper mainly designs a text representation and classification model based on Bidirectional RNN and CNN (BRCNN). BRCNN extracts word order information by using RNN, then uses CNN to extract higher-level features, and then uses maximum pooling operation to get sentence vectors, and embeds sentence features in vector space. Then use softmax classifier to classify. Finally, through experiments on eight benchmark data sets and comparative analysis with existing models, it can be found that BRCNN and ACNN can represent and classify texts well. It can also get good performance for data sets with unbalanced data. Even for long text, you can get good performance. In addition, by analyzing BRCNN, ACNN and their variants, it is concluded that the model based on LSTM and GRU can get better text representation than the model based on RNN, so the classification accuracy is higher. The convergence of LSTM and GRU-based models requires much less iteration rounds than that of RNN-based models. Comparing the accuracy of BRCNN and ACNN models, it is found that there is little difference between them. However, for BRCNN with multi-layer circulation layer and ACNN with multi-layer attention mechanism layer, the latter can stably converge to an optimal or local optimal solution.

Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2021R1F1A1051940).

5 REFERENCES

[1] Kim, J. N. (2019). The mediating effect of fear of missing out between interpersonal relationship orientation and SNS addiction tendency in college students. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 5*(3), 1-9.

https://doi.org/10.21742/apjcri.2019.09.01
[2] Shin, B. & Kim, J. (2020). Chain-based routing model for improving the network lifetime in WSNs. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(11), 173-181. https://doi.org/10.47116/apjcri.2020.11.15
[3] Kim, E. H. & Yilong, C. (2022). The effect of YanBang dishes selection attributes and sns information characteristics in china on brand loyalty: Focusing on the mediating effect of consumer satisfaction. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8*(1), 37-51 https://doi.org/10.47116/apjcri.2022.01.04
[4] Wang, X. (2020). The construction of employment and entrepreneurship service system for university students based on big data. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(10), 203-213. https://doi.org/10.47116/apjcri.2020.10.16
[5] Hong, Y., Shin, S., Kim, N., Kim, J., & Kim, H. (2022). A study on the perception based on big data analysis related to Osiria tourism complex in Busan. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8*(3), 59-68 https://doi.org/10.47116/apjcri.2022.03.06
[6] Yoo, H. (2022). Analysis of research status of early childhood physical education using big data. Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8(5), 97-106 https://doi.org/10.47116/apjcri.2022.05.09
[7] Kim, S. Y. (2018). Urban image formation through evaluation of public design guideline of bus stop in Jeju. Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 4(2), 21-30. https://doi.org/10.14257/apjcri.2018.06.03
[8] Asish Vardhan, K. (2019). Some studies on digital image segmentation techniques. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 5*(1), 77-89. https://doi.org/10.21742/apjcri.2019.03.08
[9] Trikha, V., Fiaidhi, J., & Mohammed, S. (2020). Identifying EEG binary limb motor imagery movements using thick data analytics. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(9), 169-189 https://doi.org/10.47116/apjcri.2020.09.15
[10] Lee, J. Y. & Durbha, M. (2016). Customary broadcast encryption with advanced encryption and short ciphertexts. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 2*(2), 27-33. https://doi.org/10.21742/APJCRI.2016.06.04
[11] Liu, T. & Yang, D. (2021). A three-branch 3D convolutional neural network for EEG-based different hand movement stages classification. *Scientific Reports, 11*(1), 10758. https://doi.org/10.1038/s41598-021-89414-x
[12] Li, G., Liu, F., & Wang, Y. (2021). A convolutional neural network (CNN) based approach for the recognition and evaluation of classroom teaching behaviour. *Scientific Programming, 2021*, Article ID 6336773. https://doi.org/10.1155/2021/6336773
[13] Li, T., Xu, Y., & Luo, J. (2021). A method of amino acid terahertz spectrum recognition based on the convolutional neural network and bidirectional gated recurrent network model. *Scientific Programming, 2021*, Article ID 2097257. https://doi.org/10.1155/2021/2097257
[14] Sang, X., Zhou, R. G., & Li, Y. (2021). One-dimensional deep convolutional neural network for mineral classification from Raman spectroscopy. *Neural Processing Letters, 54*(1), 677-690. https://doi.org/10.1007/s11063-021-10652-1
[15] Lee, S. H., Goau, H., & Bonnet, P. (2020). Attention-based recurrent neural network for plant disease classification. *Frontiers in Plant Science*, 11, Article 601250. https://doi.org/10.3389/fpls.2020.601250
[16] Ren, X., Gu, H., & Wei, W. (2021). Tree-RNN: Tree structural recurrent neural network for network traffic classification. *Expert Systems with Applications, 167*(1), 114363. https://doi.org/10.1016/j.eswa.2020.114363
[17] Jing, Z., Liu, J., & Ibrahim, M. S. (2020). Lifetime prediction of ultraviolet light-emitting diodes using a long short-term memory recurrent neural network. *IEEE Electron Device Letters, 41*(12). https://doi.org/10.1109/LED.2020.3034567
[18] Liu, Y., Al-Shehari, H., & Zhang H. (2022). Attention mechanism enhancement algorithm based on cycle consistent generative adversarial networks for single image dehazing. *Journal of Visual Communication and Image Representation, 83*, 103434. https://doi.org/10.1016/j.jvcir.2021.103434
[19] Zhang, H., Peng, G., & Wu, Z. (2022). MAM: A multipath attention mechanism for image recognition. *IET Image Processing*, 16. https://doi.org/10.1049/ipr2.12370
[20] Xu, X., Gao, T., & Wang, Y. (2022). Event temporal relation extraction with attention mechanism and graph neural network. *Tsinghua Science and Technology, 27*(1), 79-90. https://doi.org/10.26599/TST.2020.9010063
[21] Dehghan-Dehnavi, S., Fotuhi-Firuzabad, M., & Moeini-Aghtaie, M. (2020). Decision-making tree analysis for industrial load classification in demand response programs. *IEEE Transactions on Industry Applications*, (99), 1. https://doi.org/10.1109/TIA.2020.3032932
[22] Amgad, M., Atteya, L. A., Hussein, H., Hosny Mohammed, K., Hafiz, E., Elsebaie, M. A. T., Mobadersany, P., Manthey, D., Gutman, D. A., Elfandy, H., & Cooper, L. A. D. (2022). Explainable nucleus classification using Decision Tree Approximation of Learned Embeddings. Bioinformatics, 38(2), 513-519. https://doi.org/10.1093/bioinformatics/btab670
[23] Baek, M. & Baker, D. (2022). Deep learning and protein structure modelling. *Nature Methods, 19*(1), 13-14. https://doi.org/10.1038/s41592-021-01360-8
[24] Zhang, P., Xue, J., & Lan, C. (2020). EleAtt-RNN: Adding attentiveness to neurons in recurrent neural networks. *IEEE Transactions on Image Processing*, 29. https://doi.org/10.1109/TIP.2019.2937724
[25] Tian, Y., Yao, H., & Cai, M. (2021). Improving RNN transducer modeling for small-footprint keyword spotting. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021)*. https://doi.org/10.1109/ICASSP39728.2021.9414339
[26] Barak, F. & Kaplan, K. (2021). The study of handwriting recognition algorithms based on neural networks. *International Journal of Hybrid Innovation Technologies*, 1(2), 63-74. https://doi.org/10.21742/ijhit.2653-309X.2021.1.2.04
[27] Yang, L., Liguo, H., & Xuesen, C. (2015). A Small Sample Text Classification Algorithm based on Kernel Density Estimation, *International Journal of Signal Processing, Image Processing and Pattern Recognition, NADIA, 8*(6), 235-244. https://doi.org/10.14257/ijsip.2015.8.6.25
[28] Zhong, L. (2017). Semi-supervised Text Classification Using SVM with Exponential Kernel. *International Journal of Database Theory and Application, NADIA, 10*(1), 79-88. https://doi.org/10.14257/ijdta.2017.10.1.08

Author's contacts:

Deageon Kim
Architectural Engineering, Dongseo University,
47 Jurye-ro, Sasang-gu, Busan, 47011, Republic of Korea
gun43@hanmail.net

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221228180808

Preliminary communication
Received: 2022-12-28, Accepted: 2022-12-30

Text Classification of Mixed Model Based on Deep Learning

Sang-Hwa Lee

Abstract: At present, deep learning has been widely used many fields, but the research on text classification is still relatively few. This paper makes full use of the good learning characteristics of deep learning, proposes a hybrid model based on deep learning, and designs a text classifier based on the hybrid model. This hybrid model uses two common deep learning models, sparse automatic encoder and deep confidence network, to mix. The hybrid model is mainly composed of three parts, the first two layers are constructed by sparse automatic encoder, the middle layer is a three-layer depth Convolutional Neural Network (CNN), and finally Softmax regression is used as the classification layer. In order to test the classification performance of the classifier based on deep learning hybrid model, relevant experiments were conducted on English data set 20Newsgroup and Chinese data set Fudan University Chinese Corpus. In the English text classification experiment, the classifier based on deep learning hybrid model is used to classify, and a high classification accuracy rate is obtained. In order to further verify the superiority of its performance, a comparative experiment with naive Bayes classifier, K-Nearest Neighbor (KNN) classifier and Support Vector Machine (SVM) classifier demonstrates that the classification effect of the classifier based on deep learning hybrid model is better than that of naive Bayes classifier, KNN classifier and support vector machine classifier. In the experiment of Chinese text classification, the Chinese corpus of Fudan University is tested, and a good classification effect is obtained. The influence of different parameter settings on the classification accuracy is discussed.

Keywords: classification; deep confidence network; deep learning; sparse automatic encoder; softmax

1 INTRODUCTION

Since the widespread application of Internet technology, people are faced with the severe problem of information explosion. The information on the Internet keeps increasing, and its growth momentum is rapid, with geometric magnitude [1-3]. The Internet can carry amazing information, and the world is submerged in information. The Internet has now become a key tool for most people to search for or acquire information, making it an essential tool for people's daily lives and work. The Internet has provided quite a lot of information, among which how to find valuable information accurately and quickly has become an important issue. At present, text information contains a lot of valuable information [4-6]. Classifying published text is one of the important ways to analyze data, and improving the efficiency and quality of text usage makes it possible to organize or manage text effectively. Text classification refers to the analysis of the content of the text, and it is determined that the text belongs to any of the given categories [7][8]. In the early days, people relied on manual classification of texts. This traditional method was time-consuming and laborious, unable to deal with massive text information, and it was difficult to unify the standard because of the unstable classification results caused by human factors. At present, the main methods of text classification are statistics and machine learning, which has made a lot of progress and entered a stage of rapid development [9][10]. So far, text classification is still a hot research topic of many researchers. The main idea is to apply the text classification algorithm, learn the known samples, classify the unknown texts through the learned rules, and finally get the text categories. Text classification can handle a large number of texts, reduce the consumption of manpower and material resources, and enable users to obtain valuable content quickly and efficiently. It provides convenience for the follow-up research work and makes the text information processing rise to a new height. With the continuous in-depth exploration of the field of text classification, text classification has been well promoted in search engines, digital libraries, and email filtering and other fields.

(a) Search engine

Search engine is an indispensable tool in people's life, which can get information from the Internet. The Internet is composed of a huge number of web pages, and it is difficult for people to get the information they want. The function of search engine is to quickly classify the information on the Internet and screen the relevant information from the categories. It involves the classification of texts, which classifies texts according to their contents and then manages them separately. When a user wants to inquire about information, the search engine can provide retrieval service to the user, retrieve the relevant information that the user wants from the relevant classified information, and provide the information in the form of a page.

(b) Digital library

Because information technology has developed steadily and rapidly, digital library has become the development direction of most libraries. The technique of classifying acquired text has also become one of the important techniques for retrieving information. When the library classifies books, it adopts the text classification technology, which can effectively manage books and reduce the tedious work of librarians. The digitalization of the library is convenient for readers, and enables readers to get all kinds of library information in different places.

(c) Mail filtering

With the development of the Internet, e-mail provided great convenience to communicate with other people. And the existence of spam also adds trouble to life. Text classification technology can filter out junk information, so that users can avoid the interference of this information. Text classification technology is trained according to the characteristics of spam, and a spam classifier is obtained. E-

mail classifier can filter out the junk information and keep only the information that users need, so that users' daily life is free from interference. It can be seen from the above functions of text classification in various fields that the research of text classification has important theoretical and practical significance. Text classification, as a basic task, can provide an effective guarantee for deep mining of valuable information in the text.

2 RELATED WORK

2.1 Research Status of Text Classification

Text classification was first proposed in the 1960s. Manual classification is the earliest method of text classification, which was manually classified by professional researchers. This classification method would waste much manpower and resources and was limited by the number of professional researchers. Specific classification problems must be formulated and implemented by specific researchers. By the 1890s, the number of texts was exploding, and the proposal and development of machine learning attracted many researchers. Firstly, the text classifier trains a large number of data sets to establish a certain mathematical model, and then automatically classifies other new sample data. In the middle of 20th century, the research on text classification has been carried out abroad. In 1957, Luhn put forward the idea of applying word frequency statistics to text classification, which laid the foundation for text classification. Then, Maron et al. put forward probability model and factorization model algorithm successively, which made the text classification technology develop. In 1970, Salton et al. put forward a vector space model which can represent text well. During this period, the text classification technology mainly uses the method of knowledge engineering, and the method of knowledge engineering depends on the rules formulated by experts. However, the formulation of relevant rules will take a lot of time and energy, which makes this method unable to be popularized. In the 1990s, with the rapid development of the Internet, there was an urgent need to classify more and more different kinds of texts. At this time, machine learning methods emerged and were quickly applied to text classification. Text classification based on machine learning doesn't need manual operation to construct a classifier. It finds different features among texts by learning samples, summarizes these features, and automatically generates a text classifier according to relevant rules. Text classification using machine learning is superior to knowledge engineering in accuracy and efficiency, and it gradually replaces the method based on knowledge engineering and becomes the mainstream. Through a certain experimental analysis, it was found that the new text classifier was comparable to professional researchers in classification accuracy, so it became a common way of text classification technology at that time. In 1971, Rocchio proposed a new linear classifier [11]. In 1979, van Rijsbergen put forward some new concepts in the field of information retrieval and applied them to text classification technology, such as evaluation criteria such as accuracy and recall. In 1995, Vipnik proposed the method of Support Vector Machine. Thorsten Joachims applied linear kernel support vector machine to text classification technology for the first time, so up to now, the theory and application of support vector machine still have great influence on text classification technology. After 1995, Yoav Freund and Robert E. Schapire published a paper on Ada Boost. Robert E. Schapire proposed an Ada Boost algorithm framework and carried out relevant experimental verification. Later, some scholars designed many similar algorithms according to this framework, and these algorithms have made great achievements in the research of text classification [12, 26]. Joachims took the lead in proposing a text classification algorithm based on support vector machine in 1997, which started the upsurge of various theories and researches on the application of support vector machine in text classification. Alfons J. et al. studied the smooth Bayesian text classification algorithm in 2002. Hiroshi O., Hiroshi A., et al. have studied the features in unbalanced texts, and come to the conclusion that different feature selections will affect the results of text classification. Tantreev et al. proposed an improved feature selection method of TF-IDF. Chin H W, Lam H L and others put forward a text classification method based on KNN and support vector, which improves the accuracy of classification. Guptal et al. used rough set method to select features, which greatly reduced the training time of the classifier and achieved good classification results. Hirsch et al. used TF-IDF model for feature selection, and used genetic algorithm as text classification algorithm to classify Reuters data. Arunasalam et al. put forward a thresholding-free association rule classification algorithm for the first time. This algorithm uses a new measure to solve the problem of unbalanced distribution of categories in the text. K. Yi et al. selected the features of the medical field and classified the medical texts by using hidden Markov model classification algorithm. In 2008, Zhou Puxiong used ANN algorithm, KNN and SVM algorithm to classify texts. Compared with the traditional classification methods, using the existing natural language processing tools has the problem of error superposition in the processing process [13]. In 2014, Zeng D J et al. put forward a learning method of text semantic features based on deep convolution neural network. According to the degree of correlation between apparent and potential semantics and the categories of documents, this method can handle the classification of irregular texts such as Chinese network short texts well [14]. In 2018, Li H M, et al. proposed a short text classification model based on dense network as direct expression text [15]. In 2019, Wang Gensheng, Huang Xuejian and Chloe Wang optimized the text classification algorithm by modifying the word vector weight and manually building a dictionary, respectively. However, its learning time complexity is much higher than the traditional method, and it needs further improvement [16]. In 2019, Jin W Z proposed a text classification method based on feature fusion model of deep learning [17]. Although machine learning has made extremely important achievements in the field of text classification, the research on text classification was once stagnant before this, but the characteristics of text classification itself put forward a new development direction for machine learning [18-21], so the

research on text classification is still an extremely important direction in the area of NLP at present.

2.2 Overview of Text Classification

Text classification is a kind of supervised learning. It is known that there is a set of training documents $D = \{d_1, d_2, \ldots, d_m\}$, and each document in the set has a category label. The rules between the category label and the attributes in each document are found through supervised learning, and then the category label is obtained by using the rules for new documents.

Text classification can be defined in the following mathematical form: given a set of documents, d_i represents the i^{th} document, and there are m documents in D. Assuming a set of document categories $C = \{c_1, c_2, \ldots, c_m\}$, we can find that there is a certain mapping between the set of documents and the set of categories $f \rightarrow D \times C \rightarrow A$, $A = \{0, 1\}$, the task of text classification is actually to make it equal to F as much as possible. Called a classifier. If $f'\left(\left\langle d_j, c_i \right\rangle\right) = 0$, d_j belongs to the class c_i. If $f'\left(\left\langle d_j, c_i \right\rangle\right) = 1$, it is said that it does not belong to the class c_i.

The text classification process consists of training process and classification process as shown in Fig. 1. In the training process, the training text generally needs to go through the steps shown in Fig. 1, which are the basis of text classification, and then the classifier is continuously trained by the selected classification algorithm. In the process of classification, the test document generally needs to be processed by the steps shown in Fig. 1. After the trained classifier, the classifier will identify the category of the test document.

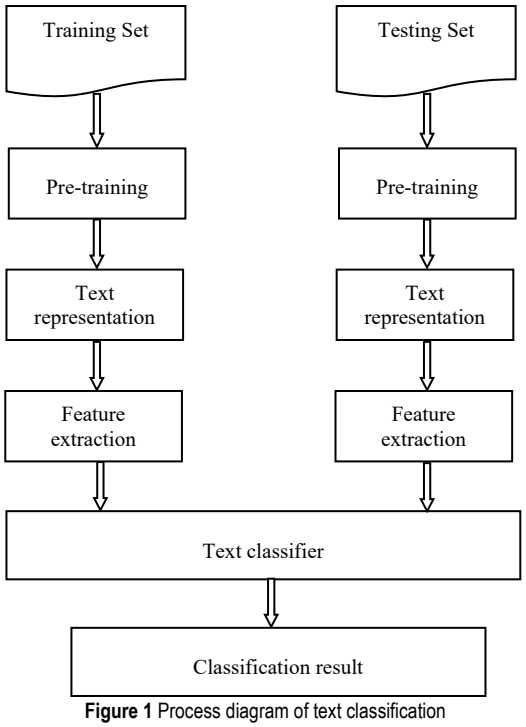


Figure 1 Process diagram of text classification

The text contains a large amount of unstructured or semi-structured information, which is not easily recognized by the classifier. It is necessary to preprocess the text documents to remove these useless information. Text preprocessing refers to processing text information into structured information that can be operated by computer. Text preprocessing is the initial stage of text classification, and the preprocessing results have great influence on the classification results. Text preprocessing includes denoising, word segmentation and stopword removal. In English, spaces and punctuation marks are commonly used for word segmentation. In English, it is necessary to take root, which is to unify words with the same semantics but slightly different forms into one form. It mainly aims at singular and plural forms of nouns, comparative forms of adjectives and adverbs, and various tense forms of verbs in English. To go to stop words is to remove pronouns, prepositions, conjunctions and other features unrelated to classification. These stop words are irrelevant to the meaning of the document.

2.3 Common Models of Deep Learning

2.3.1 Automatic Encoder

Automatic encoder is a kind of unsupervised learning, a new network reconstructed by neural network [22, 23]. The encoder principle of automatically acquiring data features makes input data and output data identical. By constantly adjusting the weight of each layer through training, each hidden layer is another representation of the input data and can be used as the features of the input data. Compared with principal component analysis, automatic encoder relies on the limitation of linear dimension reduction, and it can use nonlinear neural network to reduce the dimension of features. Automatic encoder consists of encoder and decoder. The output of the original data after passing through the encoder is used as the input of the decoder, and finally the output is obtained through the decoder. Then, the original data is printed in another form, as shown in Fig. 2 below.

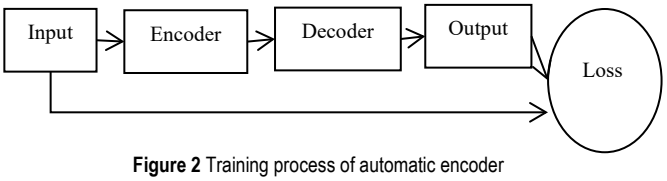


Figure 2 Training process of automatic encoder

2.3.2 Convolutional Neural Network (CNN)

It was put forward by Lecun in 1989, and it is well applied in the field of speech recognition and image recognition. Convolutional neural network is essentially a multilayer perceptron that can recognize images well [27]. Because of its special structure, it can highly perceive other forms of invariance such as translation and scaling of graphics.

CNN [24] is composed of one or more convolution layers and the top fully connected layer, and includes correlation weights and pooling layers. This structure enables CNN to make use of the two-dimensional structure of input data. The

structure diagram of CNN is shown in Fig. 3. First, the input original features are convolved in C1 layer, and then transformed into feature maps after passing through three filters. Then, the feature maps are weighted, and then biased, and finally Sigmoid function is processed to generate S2 layer feature maps. The obtain feature map is processed as above to successively obtain C3 and S4 layer feature map. The feature mapping mentioned above can well realize the feature that the position is not easy to change, and Sigmoid function is used as the activation function. The middle layer C is the feature extraction layer, and the nodes of neurons in each layer are connected with the local nodes in the front layer to extract local features.

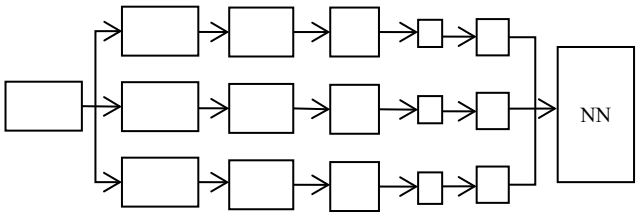


Figure 3 Schematic diagram of convolutional neural network structure

Table 1 20Newsgroups data set

Category	Number of texts
alt.atheism	900
computer.graphics	900
computer.os.ms-windows.misc	900
computer.windows.x	900
misc.forsale	900
computer.system.ibm.pc.hardware	900
computer.system.mac.hardware	900
rec.sport.baseball	900
rec.sport.hockey	900
rec.autos	900
rec.motorcycles	900
science.med	900
soc.religion.christian	900
science.crypt	900
science.electronics	900
talk.politics.guns	900
science.space	900
talk.politics.misc	900
talk.politics.mideast	900
talk.religion.misc	897

3 EXPERIMENTAL RESULT ANALYSIS

3.1 Data Set

The standard foreign language classification database includes: Reuters-21578, 20Newsgroups, OHSUMED, Web KB, etc. Domestic standard Chinese corpus such as Tan Corp, etc. These data sets can be downloaded for free. In this paper, 20Newsgroups data sets such as Tab. 1 are selected for the English text classification experiment. This data set is a text data set compiled by Lang in 1995. It contains the message texts of 20 newsgroups (20 categories) in Usenet, with a total of 1997 articles. Except one newsgroup contains 997 messages, each newsgroup has 1000 message texts. This data set is a typical single-label text classification corpus.

3.2 Text Classification Experiment

A certain number of texts are selected from the random English data set 20Newsgroup to preprocess the texts. Text preprocessing is implemented on Eclipse platform using Java language. For the preprocessed documents, 30% of them are randomly selected as test data sets, and the rest of them are training data sets. The feature dimension of the document is 1500 dimensions. In this paper, the classifier based on the mixed model of deep learning is implemented by MATLAB. Because the original feature dimension is 1500, the number of input nodes in the sparse automatic encoder layer is 1500. After using the sparse automatic encoder with 3000-1500 hidden nodes, the data is compressed by a three-layer deep confidence network with 200-100-20 hidden nodes. Finally, the Softmax layer outputs the test data set with the highest probability that the documents belong to all categories of documents. After text classification, the best accuracy of each category can be obtained as shown in Fig. 5 and Tab. 2.

Table 2 Accuracy of text classification experiment

Category	Number	Accuracy rate
alt.atheism	471	93.49%
computer.graphics	574	96.71%
computer.os.ms-windows.misc	590	80.23%
computer.system.ibm.pc.hardware	590	85.11%
computer.system.mac.hardware	580	88.3%
computer.windows.x	590	93.3%
misc.forsale	585	95.29%
rec.autos	594	94.61%
rec.motorcycles	600	96.37%
rec.sport.baseball	598	94.85%
rec.sport.hockey	600	96.73%
science.crypt	600	97.46%
science.electronics	590	96.45%
science.med	595	89.31%
science.space	594	94.25%
soc.religion.christian	600	95.71%
talk.politics.guns	550	94.68%
talk.politics.mideast	560	95.88%
talk.politics.misc	467	89.77%
talk.religion.misc	380	62.4%
total	11380	91.52%

In order to further verify the performance of text classification based on the hybrid model of deep learning, this paper compares the proposed classifier SDBN with naive Bayes classifier, KNN classifier and support vector machine classifier. In the contrast experiment, the same data set is selected as the training set and the test set, and the text of the training set and the test set is preprocessed. In the experiment of naive Bayes classifier, the naive Bayes classifier [25] of MATLAB is used to get the classification accuracy of the test set. In the experiment of KNN classifier, the KNN-classify classifier built in MATLAB is used to get the classification accuracy of the test set. In the classification experiment of support vector machine, LIBSVM, an open source software package of support vector machine, is used for the experiment, and the classification accuracy of the test set is obtained. Fig. 5 shows the experimental results of comparative experiment, from which it can be seen that the

performance of text classifier SDBN based on deep learning hybrid model is slightly better than other classifiers.

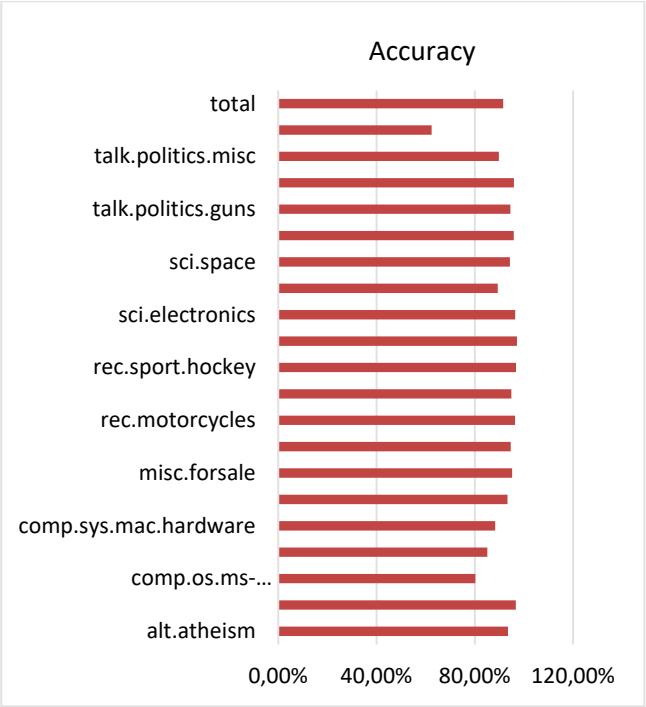


Figure 4 Accuracy rate of text classification experiment

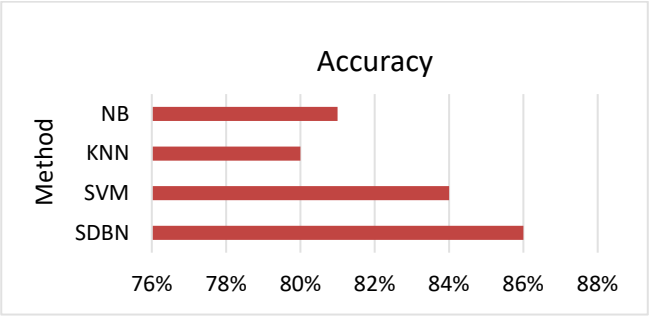


Figure 5 Classification accuracy rate

From Fudan University Chinese Text Classification Corpus, four kinds of documents, namely, economy, sports, computer and agriculture, are selected as the training set and test set of Chinese experiment. In this paper, the classifier based on the mixed model of deep learning is implemented by MATLAB, and 30% of the documents in the preprocessed Chinese data set are randomly selected as the test set. Select features with 1000 dimensions as raw data. Firstly, a sparse automatic encoder with 2000-1000 hidden nodes is used, then a three-layer deep confidence network with 200-100-20 hidden nodes is used to compress the data, and the BP fine-tuning times are 200. Finally, Softmax layer outputs the probability that the documents in the test data set belong to various categories in the documents, and the category with the highest probability is the category to which the documents belong. Tab. 3 is the best correct rate after text classification.

Table 3 Accuracy rate of text classification experiment

Category	Number	Accuracy rate
C34-Economy	1360	86.1%
C39-Sports	1023	87.7%
C19-Computer	1602	92.5%
C32-Agriculture	1255	81.3%

Tab. 4 to Tab. 11 show the comparison of accuracy recall and F1 value of BRCNNs and ACNNNS on MR, Subj, SST, SST2, IMDB, TREC, CR and MPQA data sets respectively. As can be seen from the following accuracy recall and F1 tables, on all data sets used in this paper. There is little difference between the accuracy recall rate and F1 value of BRCNN and ACNN and their variants. This shows that there is no case that a large number of samples of a certain category are predicted as low recall rate caused by other categories, and there is no case that a large number of samples of other categories are predicted as a certain category with low accuracy rate. Therefore, the prediction of the model is relatively average, and even unbalanced data can predict the categories well. For example, the data in the data set TREC is unbalanced, but it can also be predicted well. This shows that the corresponding accuracy recall rate in the table above these models is the value near the "balance point" on the P-R curve. This is mainly because 0.5 is usually used as the threshold for judging categories in the second classification task, and this paper also uses 0.5 as the classification threshold in the experiment. This also shows that BRCNN and ACNN are suitable for text classification tasks.

Table 4 The precision score, recall score and F1 score of ACNNs and BRCNNs models on the MR dataset

Models	macro-P	macro-FI	macro-R	micro-P	micro-FI	micro-R
BRCNN (RNN)	82.7	82.4	82.5	82.4	85.4	82.4
BRCNN (LSTM)	83.9	83.5	83.5	83.5	83.5	83.5
BRCNN (GRU)	83.0	81.6	81.8	81.6	81.6	81.6
ACNN (RNN)	83.5	80.5	80.5	80.5	80.5	80.5
ACNN (LSTM)	83.9	80.7	80.8	80.7	80.7	80.7
ACNN (GRU)	86.8	86.8	86.8	86.8	86.8	86.8

Table 5 The precision score, recall score and F1 score of ACNNs and BRCNNs models on the Subj dataset

Models	macro-P	macro-FI	macro-R	micro-P	micro-FI	micro-R
BRCNN (RNN)	92.4	92.1	92.0	92.2	92.2	92.2
BRCNN (LSTM)	93.5	93.5	93.6	94.5	95.5	93.5
BRCNN (GRU)	92.7	91.3	92.1	93.4	94.4	94.4
ACNN (RNN)	90.8	90.3	90.1	90.4	90.4	90.4
ACNN (LSTM)	92.0	92.4	92.1	92.6	92.1	92.1
ACNN (GRU)	92.9	92.9	92.8	92.9	92.9	92.9

Table 6 The precision score, recall score and F1 score of ACNNs and BRCNNs models on the ST dataset

Models	macro-P	macro-FI	macro-R	micro-P	micro-FI	micro-R
BRCNN (RNN)	46.3	46.2	46.0	46.3	46.3	46.3
BRCNN (LSTM)	48.0	48.1	48.2	48.1	48.1	48.1
BRCNN (GRU)	47.3	47.7	47.5	47.4	47.4	47.4
ACNN (RNN)	46.6	46.6	46.5	46.6	46.6	46.6
ACNN (LSTM)	48.7	46.5	46.0	48.3	48.3	48.4
ACNN (GRU)	47.9	47.8	47.8	47.7	47.9	47.9

Table 7 The precision score, recall score and F1 score of ACNNs and BRCNNs models on SST2 dataset

Models	macro-P	macro-F1	macro-R	micro-P	micro-FI	micro-R
BRCNN (RNN)	83.3	83.3	83.3	83.3	83.3	83.3
BRCNN (LSTM)	83.0	93.3	83.3	83.3	83.3	83.3
BRCNN (GRU)	83.7	83.7	83.7	83.7	83.7	83.7
ACNN (RNN)	83.7	83.7	83.7	83.7	83.7	83.7
ACNN (LSTM)	87.3	85.8	85.9	85.9	85.9	85.9
ACNN (GRU)	85.7	85.7	85.7	85.7	85.7	85.7

Table 8 The precision score, recall score and F1 score of ACNNs and BRCNNs models on the IMDB dataset

Models	macro-P	macro-F1	macro-R	micro-P	micro-FI	micro-R
BRCNN (RNN)	86.0	86.0	86.l	86.l	86.2	86.2
BRCNN (LSTM)	88.5	88.5	88.5	87.5	88.5	88.5
BRCNN (GRU)	86.9	86.9	86.9	86.9	86.9	86.9
ACNN (RNN)	86.2	86.2	86.2	86.2	86.2□	86.2
ACNN (LSTM)	88.6	87.5	88.5	88.5	88.5	87.5
ACNN (GRU)	88.2	88.3	88.3	88.3	88.3	87.3

Table 9 The precision score, recall score and fl score of ACNNs and BRCNNs models on TREC dataset

Models	macro-P	macro-F1	macro-R	micro-P	micro-F1	micro-R
BRCNN (RNN)	81.2	81.8	83.7	88.6	88.6	88.6
BRCNN (LSTM)	93.6	91.0	90.7	94.0	94.0	94.0
BRCNN (GRU)	91.9	89.6	87.4	91.1	91.1	91.1
ACNN (RNN)	91.4	90.8	89.7	91.9	91.9	91.9
ACNN (LSTM)	94.9	91.3	90.4	93.8	93.8	93.8
ACNN (GRU)	94.4	91.2	90.7	93.4	93.4	93.4

Table 10 The precision score, recall score and F1 score of ACNNs and BRCNNs models on CR dataset

Models	macro-P	macro-F1	macro-R	micro-P	micro-F1	micro-R
BRCNN (RNN)	86.7	86.2	81.7	86.7	86.3	86.7
BRCNN (LSTM)	86.2	86.8	88.7	88.4	88.4	88.4
BRCNN (GRU)	84.9	85.3	85.9	86.3	86.3	86.3
ACNN (RNN)	85.6	85.6	85.6	85.6	85.6	85.6
ACNN (LSTM)	84.3	83.9	83.6	85.3	85.3	85.3
ACNN (GRU)	86.7	86.9	88.2	88.7	88.7	88.7

Table 11 The precision score, recall score and F1 score of ACNNs and BRCNNs models on PQA dataset

Models	macro-P	macro-F1	macro-R	micro-P	micro-FI	micro-R
BRCNN (RNN)	87.1	86.1	85.2	88.4	89.4	88.4
BRCNN (LSTM)	88.2	88.3	88.4	91.9	90.9	90.9
BRCNN (GRU)	90.9	87.1	86.2	90.6	90.8	90.6
ACNN (RNN)	91.5	87.3	85.8	90.6	90.6	90.7
ACNN (LSTM)	88.5	87.9	87.4	90.8	90.8	90.8
ACNN (GRU)	90.6	88.1	86.9	92.1	90.1	92.1

In practice, sometimes we pay more attention to the accuracy of classification, and sometimes we may pay more attention to the recall rate of classification. At this time, we can set the classification threshold to different values according to the specific situation to get different accuracy rates or recall rates. For example, reducing the classification threshold will get a higher recall rate, but the accuracy rate will be relatively reduced. On the contrary, increasing the classification threshold will increase the pricision, but the recall rate will decrease. On the whole, the peicision and recall rate of the model are relatively average when it is near the "balance point". In practice, you can draw the P-R curve first, and then set the classification threshold according to the situation analysis.

4 CONCLUSION

To learn the classification performance of classifiers based on deep mixed model, relevant experiments were conducted on English data set 20Newsgroup and Chinese data set Fudan University Chinese Corpus respectively. In the English text experiment, the classifier based on deep learning hybrid model is used for classification, and the classification accuracy rate is 91%. In order to further verify

the superiority of its performance, the comparison experiment with naive Bayes classifier, KNN classifier and support vector machine classifier shows that the classification effect based on deep learning hybrid model is slightly better than that of SVM classifier, KNN classifier and naive Bayes classifier. In the Chinese text experiment, the Chinese corpus of Fudan University is tested and a good classification effect is obtained, and the influence of different parameters on the classification accuracy is discussed.

In the future study, we will further study the work of this paper. We can improve the algorithm in the model of deep learning, and try to use other models in deep learning, such as CNN, to learn the features of text classification. In a word, text classification based on deep learning hybrid model will have a good application prospect in the future development. With the continuous improvement of deep learning theory, more new research results of deep learning will be added to the further research of this paper, which will greatly improve the performance of text classifier based on deep learning hybrid model.

5 REFERENCES

[1] Vaddempudi, S. (2016). Classification of user behaviour in mobile internet. *Asia-pacific Journal of Convergent Research Interchange, SoCoRI, 2*(2), 9-18. https://doi.org/10.21742/APJCRI.2016.06.02

[2] Meng, L., Zhang, S., & Wang, F. (2020). Influence of internet-based social big data on personal credit reporting. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(7), 39-57. https://doi.org/10.47116/apjcri.2020.07.05

[3] Ko, J.-U., Kim, K.-H., & Jeon, J.-H. (2021). The effects of self-control on internet immersion: The moderating effects of parenting attitude. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 7*(9), 47-57. https://doi.org/10.47116/apjcri.2021.09.05

[4] Kim, H.-G. & Ko, J.-U. (2022). The effect of adolescents' perceived parenting attitudes on internet immersion - Focusing on the mediating effect of self-control. *Asia-pacific Journal of Convergent Research Interchange*, FuCoS, 8(3), 187-201. https://doi.org/10.47116/apjcri.2022.03.17

[5] Park, W., Jung, D., Jung, H., Ekouka Elvis, T., & Kim, H. (2022). Effecting system quality factors on purchase intention of internet insurance. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8*(4), 47-56. https://doi.org/10.47116/apjcri.2022.04.05

[6] Park, W. & Kim, H. (2022). Effecting information quality of information system on corporate performance: Focused on internet insurance. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 8*(5), 21-30. https://doi.org/10.47116/apjcri.2022.05.03

[7] Kim, D. (2020). Application of deep neural network model for automated intelligent excavator. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(4), 13-22. https://doi.org/10.21742/apjcri.2020.04.02

[8] Mononteliza, J. (2020). Research on EIoT reservation algorithm based on deep learning. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 6*(9), 191-205 https://doi.org/10.47116/apjcri.2020.09.16

[9] Jang, S.-B. (2021). Deep neural network structure design for equipment failure prediction in smart factory. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 7*(12), 1-10. https://doi.org/10.47116/apjcri.2021.12.01

[10] Lee, Y. G. (2022). Developing an automatic floor plan generation and evaluation technology using deep learning in the architectural design process. *Asia-pacific Journal of Convergent Research Interchange*, FuCoS, 8(5), 1-10. https://doi.org/10.47116/apjcri.2022.05.01

[11] Xie, R., Yuan, X., & Liu, Z. (2017). Lexical sememe prediction via word embeddings and matrix factorization. *The 26th International Joint Conference on Artificial Intelligence*, 4200-4206. https://doi.org/10.24963/ijcai.2017/587

[12] Liu, Z. W., Ding, D., & Li, C. W. (2015). A Chinese short text word segmentation method based on conditional random fields. *Journal of Tsinghua University (Science and Technology), 55*(8), 906-910+915.

[13] Du, L. P., Li, X. G., & Yu, G. (2016). New word detection based on an improved PMI algorithm for enhancing segmentation system. *Journal of Peking University (Natural Science), 52*(1), 35-40.

[14] Hu, J. & Zhang, J. C. (2017). Bidirectional recurrent network for Chinese word segmentation. *Journal of Chinese Computer Systems, 38*(3), 522-526.

[15] Zhang, Y. N., Xu, J. N., & Miao, G. Y. (2018). Improving neural Chinese word segmentation using unlabeled data. *IOP Conference Series: Materials Science and Engineering, 435*(1). https://doi.org/10.1088/1757-899X/435/1/012032

[16] Yang, G. B., Yang, F. H., & Mao, G. J. (2019). Chinese word segmentation technology based on group hash and variable length matching. *Computer Age*, 4, 52-55.

[17] Chen, X. C., Shi, Z. & Qiu, X. P. (2017). Adversarial multi-criteria learning for Chinese word segmentation. ACL2017: *Computation and Language*. https://doi.org/10.18653/v1/P17-1110

[18] Yu, C. H., Wang, S. P., & Guo, J. J. (2019). Learning Chinese word segmentation based on bidirectional GRU-CRF and CNN network model. *International Journal of Technology and Human Interaction (IJTHI), 15*(3). https://doi.org/10.4018/IJTHI.2019070104

[19] Darshan, S. L. & Jaidhar, C. D. (2018). Performance evaluation of filter-based feature selection techniques in classifying portable executable files. *Procedia Computer Science*, (125), 346-356. https://doi.org/10.1016/j.procs.2017.12.046

[20] Hancer, E. (2019). Differential evolution for feature selection: A fuzzy wrapper–filter approach. *Soft Computing, 23*(13), 5233-5248. https://doi.org/10.1007/s00500-018-3545-7

[21] Maldonado, S. & López, J. (2018). Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification. *Applied Soft Computing*, (67), 228-246. https://doi.org/10.1016/j.asoc.2018.02.051

[22] Cheong, Y. G. (2021). Analysis of autoencoders for network intrusion detection. *Sensors*, (21). https://doi.org/10.3390/s21134294

[23] Li, X. Liu, Z., & Huang, Z. (2020). Denoising of radar pulse streams with autoencoders. *IEEE Communications Letters*, (99), 1-1. https://doi.org/10.1109/LCOMM.2020.2967365

[24] Davis, P., Aziz, F. B., & Newaz, T. (2021). The classification of construction waste material using a deep convolutional neural network. *Automation in Construction*, 122. https://doi.org/10.1016/j.autcon.2020.103481

[25] Thong, W. & Snoek, C. G. M. (2021). Feature and label embedding spaces matter in addressing image classifier bias. arXiv: 2110.14336. https://doi.org/10.48550/arXiv.2110.14336

[26] Barak, F. & Kaplan, K. (2021). The study of handwriting recognition algorithms based on neural networks. *International Journal of Hybrid Innovation Technologies, 1*(2), 63-67. https://doi.org/10.21742/ijhit.2653-309X.2021.1.2.04

[27] Chayangkoon, N. & Srivihok, A. (2019). Feature Reduction of Short Text Classification by Using Bag of Words and Word Embedding. *International Journal of Control and Automation, NADIA, 12*(2), 1-16. https://doi.org/10.33832/ijca.2019.12.2.01

Author's contacts:

Sang-Hwa Lee
Department of Webtoon Contents, Seowon University,
377-3 Musimseo-ro, Seowon-gu, Cheongju-si,
Chungcheongbuk-do, 28674, Republic of Korea
gomawooi@naver.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230504144856

Preliminary communication
Received: 2023-05-04, Accepted: 2023-05-08

A Returnable Transport Item to Integrate Logistics 4.0 and Circular Economy in Pharma Supply Chains

Gerald Schneikart*, Walter Mayrhofer, Josef Frysak, Clemens Löffler

Abstract: Recent global events, such as the COVID-19 pandemic, the war in Ukraine and the climate crisis, force the pharma logistics sector to rapidly improve their processes and establish more resilient and sustainable medical supply chains. For this purpose, the pharma logistics sector needs to catch up in Industry 4.0 adoption and establish circular economies. In the context of the applied research project DigiPharmaLogNet, a prototypic returnable transport item (RTI) is enhanced with communication technology and piloted in pharma-specific use-cases. The results will build the base for developing business models and roadmaps towards sustainable pharma logistics networks. This article describes the technological developments and economical evaluations of potential business models.

Keywords: Circular Economy; Industry 4.0; Pharma Logistics; Pharma Supply Chain; Reverse Logistics; Sustainable Development Goals

1 INTRODUCTION

1.1 Circular Economies for the Reduction of Packaging Waste in Pharma Logistics

Recent shortages in medical supply illustrate the importance of resilient pharma logistics networks [1, 2]. The pharmaceutical industry puts high requirements on its supply chains; most notably temperature monitoring and traceability of medicinal products are a priority in order to warrant product functionality or avoid distribution of counterfeit drugs, respectively [3].

In pharma logistics, carboard-based transport items are most commonly in use for delivery of medical products from manufacturers to the point of administration. The life cycle of such one-way transport solutions entails a considerable footprint [4, 5]. Even though paper waste can be recycled to manufacture new transport items, the quality of the recycled fibers limits the number of reuse cycles (Fig. 1A). Besides, transportation of paper waste to package manufacturing plants is necessary, which further contributes to an increase of the environmental footprint.

As a consequence of the objective to reach the global sustainable development goals (SDG), the European Green Deal proposes a policy framework for handling packaging waste that triggers more sustainable business innovations. The introduced directive will also set new standards in logistics, since transport crates rely heavily on one-way solutions for packaging and filling materials [6].

In order to advance the European Green Deal and move towards a circular economy the usage of returnable transport items (RTI) seems very promising. RTIs eliminate a required package manufacturer as intermediary, thus allowing shorter circulation routes. In addition, higher numbers of reuse cycles compared to cardboard-based transport items are expected (Fig. 1B) [5].

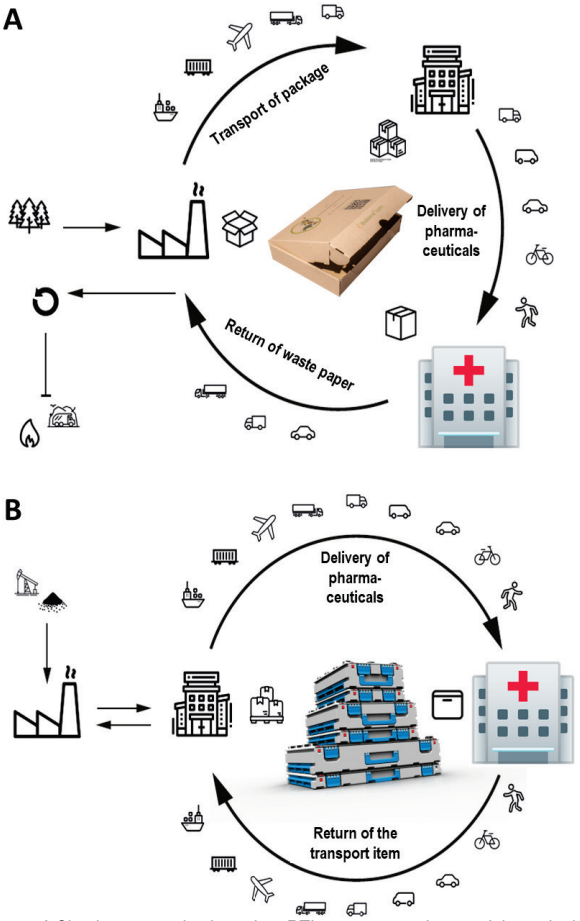


Figure 1 Circular economies based on RTIs are expected to result in ecological and economic advantages over the recycling process of fibers from cardboard-based boxes. (A) Life cycle of fibers devoted for cardboard-based transport items. Packages are fabricated and transported to pharmaceutical manufacturers utilizing cardboard-based boxes for delivery purposes. Once shipped pharmaceuticals reach the customers (e.g., hospitals or pharmacies), fibers of package material are either recycled in package factories or incinerated. (B) Life cycle of material devoted for RTIs. Once fabricated, pharmaceutical manufacturers may repeatedly utilize returned RTIs for delivery purposes. Package material may be eventually recycled after RTIs become damaged or unemployable.

1.2 Organization of the Article

Section 2 summarizes current challenges and technical opportunities in establishing circular economies in pharma logistics. Section 3 introducing the applied research project DigiPharmaLogNet that aims at evaluating the feasibility of implementing RTI solutions as enablers of logistics 4.0 and circular economies. Section 4 presents the last results of DigiPharmaLogNet. The conclusion in section 5 provides an overview of remaining gaps to be addressed by research.

2 LITERATURE REVIEW

2.1 The Challenges in Integrating Circular Economies with Logistics 4.0 Technologies

The advent of the Industry 4.0 not only triggered new technical innovations and paradigms in the manufacturing industry, but also led to the emergence of new business models and processes [7]. As part of Industry 4.0 in a wider sense, this equally applies to Logistics 4.0, the next generation of logistics, which also aims to solve pressing problems in the area of sustainability [8].

Due to the nature of (pharma) supply chains, the establishment of circular economies relies on the consent of multiple business stakeholders willing to transform their processes with RTI solutions, amongst others wholesalers and other intermediaries [9]. As economic players, their readiness strongly depends on the applicability and long-term profitability of the RTI solution in their specific field. Driven by the European Green Deal policy, the question therefore arises how a practically applicable RTI solution in the context of (pharma) Logistics 4.0 can be designed that creates a technical solution capable of achieving sustainability goals in line with profitable, long-term business models.

In order to build circular economy concepts based on Logistic 4.0, specific use cases and business models are needed to provide quantifiable evidence on the potential of novel technical solutions. In this paper, we aim to contribute to the specific research field of circular economies in pharma logistics by exploring a new technology solution of reusable containers, examine its economic feasibility in two use cases, and explore potential business models designed for implementation of the RTI solution.

2.2 Current Work on Circular Economies in Pharma Logistics Sector

RTIs have been used in the pharmaceutical sector in closed intra-logistics systems or between pharma wholesalers and pharmacies, where there are several deliveries a day and the return logistics of the empty RTIs is done at little or no extra cost [10, 11]. In addition, digitization of supply chains and widespread use of Internet of Things (IoT)-technology, that allow tracking and tracing of shipments suggest increased reliability of RTI-based circular logistics systems [12]. In particular, there are very promising applications of blockchain technology [13] as well as robotics and drone solutions that could streamline reverse logistics and cold chain processes [14, 15].

3 THE RESEARCH PROJECT DIGI-PHARMA-LOG-NET

DigiPharmaLogNet is a consortium project for the development of a returnable transport item solution applicable for the establishment of circular economies within pharma supply chains (Fig. 2). The Institute for Digital Transformation and Strategy of the University of Applied Sciences for Management and Communication in Vienna leads and conducts the project in partnership with non-profit research institutions, start-up companies, and industrial partners involved in Austria's pharma supply chains. The partners provide the process resources needed for the analysis of possible business cases [16].

The project addresses the following research questions:

- What are the organizational requirements for the establishment of circular economies in pharma logistics?
- What are the technical requirements for an RTI solution that enables the integration of logistics 4.0 and reverse logistics?
- What are the potential economic impacts of implementing an RTI solution?
- How to evaluate sustainability aspects of RTI solutions?

In order to achieve the objectives, the project follows an exploratory, incremental design approach, in which a proof-of-concept demonstrator is designed and applied to selected use cases for first feasibility evaluations. According to the chosen approach, the development of the proof-of-concept demonstrator is carried out in four phases:

1) Flare and Focus Phase: in this phase, the initial designs for the proof-of-concept demonstrator were established.
2) Design and Development Phase: in this phase the initial design was refined, and a physical prototype built, including the digital control systems.
3) Demonstration phase: here, feasible use cases from the partner companies were selected and test-settings derived. The proof-of-concept demonstrator was then adopted in these test-use-cases.
4) Business model derivation phase: in the final phase, the findings from the previous tests were combined to develop possible future business models.

This particular project approach not only allowed to utilize the insights from the practical application to be used as a basis for redesigning the initial RTI prototype, but also provided valuable insights for designing possible future business models.

The core technology developed in DigiPharmaLogNet is an RTI designed by BOOXit, which is intended for circular logistic chains. In its current developmental stage, the RTI is made of a durable plastic-based material that allows a high number of reuse cycles (Fig. 1).

Moreover, its design is optimally suited for an application in pharma logistics. It offers opportunities for enhancement with Industry 4.0 sensor and communication technologies and enables temperature monitoring and tracking functionalities which are some of the most discussed problems in pharma logistics [9]. With a multitude of additional hard- and software add-ons it strives to provide an

Internet of Things (IoT) ecosystem in logistics with the RTI as the central component and enabler of value-added services.

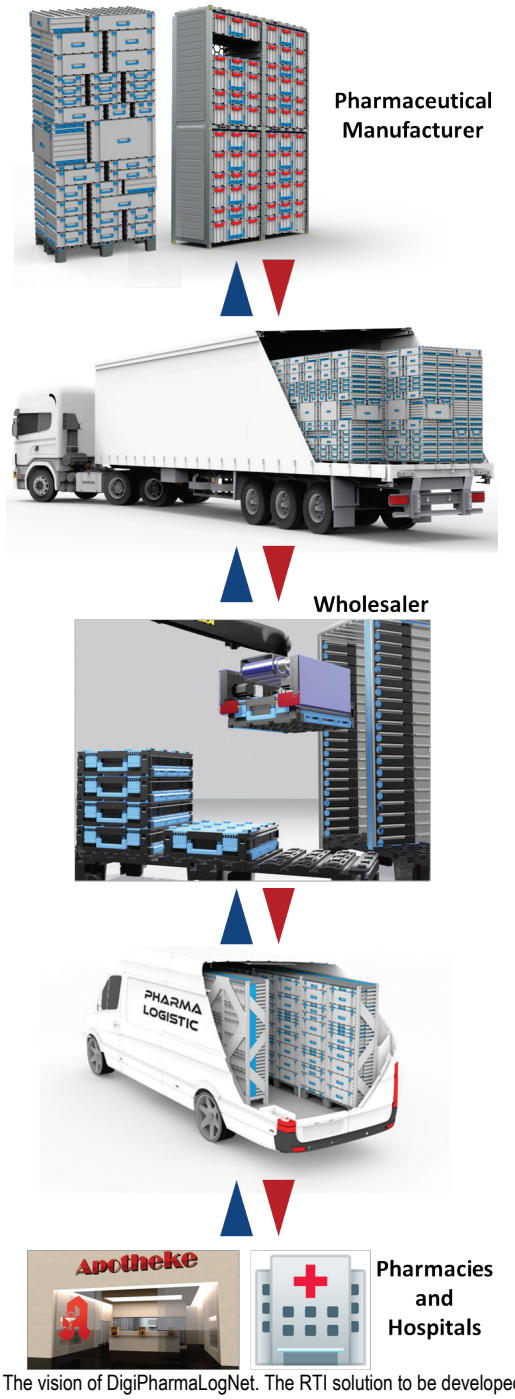


Figure 2 The vision of DigiPharmaLogNet. The RTI solution to be developed by the project team has a modularity mechanism incorporated, which allows stacking boxes with different sizes. Its compatibility with a robotic system enables automation of warehouse processes. Since RTIs are intended for reuse, circular logistic chains could be established in pharma logistics. RTI, returnable transport item.

The design solution includes an integrated grid mechanism that allows stacking boxes of different sizes, comparable to Lego® bricks (Fig. 3A-C). Furthermore, the grid mechanism enables horizontal displacement of boxes placed one upon the other. This allows utilization of individual boxes as drawers in dedicated racks (Fig. 3D), dubbed as "one-shot loading" of multiple boxes on or off delivery systems, such as trucks or trains, becomes possible. In view of the goal to eliminate cardboard-based boxes in supply chains [6], a grid mechanism poses an advantage, since cardboard-based boxes need to be managed single-handedly and demand extra transport security measures like belts or cellophane wrapping.

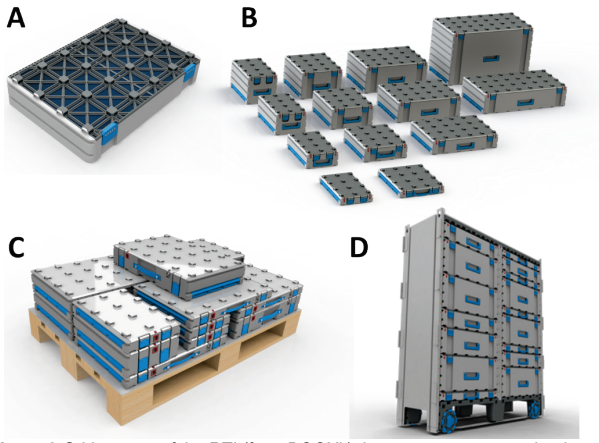


Figure 3 Grid system of the RTI (from BOOXit) that serves as core technology of DigiPharmaLogNet. The integrated grid mechanism (A) allows modular applications of boxes with different sizes (B), i.e., "stackability" (C) or usage as drawers in a proprietary rack (D).

The grid mechanism and overall design of the core technology also offers potential for semi- or full automation of RTI inventory management processes in warehouses. In addition, the RTIs are compatible with a robot arm. Together with the "one-shot load" feature of a specialized rack on wheels the overall RTI system will thus reduce ergonomic problems of the logistics industry.

The ultimate objective of DigiPharmaLogNet is the evaluation of usability and economic feasibility in terms of economic efficiency and reduced CO_2 emissions, that are expected due to a high number of reuse cycles (Fig. 1B). For this purpose, potential use cases were selected based on their improvement potential but also on accessibility to pharma supply processes of the industrial partners.

4 RESULTS

To date, the research and development project delivered several outcomes, which are presented in this section. The primary outcome is a proof-of-concept (PoC) demonstrator that identifies technical benefits and challenges of the new RTI concept. With the limitations that a demonstrator poses, two real-world use cases in pharma logistics could be identified. Those use-cases were analyzed with respect to their economic potential when implementing the new RTI system. Vice versa, the considerations during implementation also led to adjustments and improvements of the PoC demonstrator itself. Based on the data from the use cases, a hypothetical business model was created for the start-

up company BOOXit, which develops and commercializes the new RTI system.

4.1 The Proof-of-Concept Demonstrator

A major objective of DigiPharmaLogNet is the development of an Industry 4.0-based RTI solution that meets the specific demands of the pharma logistics industry, such as temperature monitoring and location tracking. This necessitates the integration of electronic components for automatic communication with an inventory system software program.

The prototypic rack controller, which is conceived as the basic electronics unit of the rack, automatically reads data from a near frequency communication (NFC) chip (as part of a box), such as RTI identification number, content (pharmaceuticals) in the RTI, or current position (slot) in the rack. The prototypic firmware also has writing functionalities for data stored on the NFC.

In addition, a first configuration of an RTI controller was applied to communicate with sensor technologies (e.g., temperature sensor) that might be incorporated in the box. In order to support delivery personnel with drawing boxes from racks or warehouse personnel with putting boxes into racks (of trucks), the firmware has pick-by-light or put-by-light functionalities, respectively.

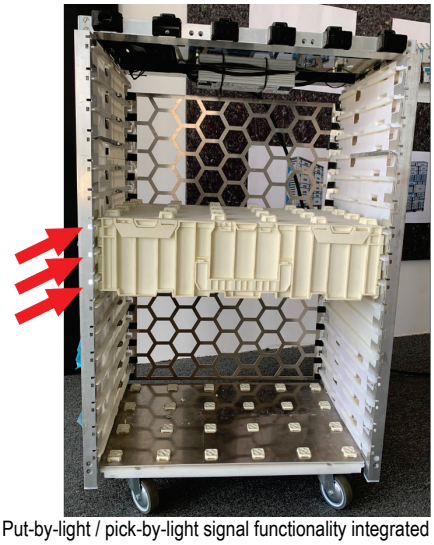


Figure 4 Put-by-light / pick-by-light signal functionality integrated in the PoC demonstrator rack. The flashing LED lights (red arrows) indicate that the respective slot is unlocked, thus ready for removing the box it holds or accepting a box. Once the LED lights turn out, the slot is locked, thus the box is fixed in the slot or the slot is not ready for accepting a box. PoC, proof-of-concept; LED, light-emitting diode.

The rack and RTI controller units were integrated with the first PoC demonstrator rack. In its current developmental stage, the rack controller can be manually triggered to control the lock mechanism of the rack. If the light-emitting diode signals are flashing, the corresponding locks of the slots are ready to accept or release single or stacked boxes (Fig. 4). In combination with the box inventory system, RTI movements can be managed. A prototypic smartphone app for supporting the RTI handling has already been developed (Fig. 5).

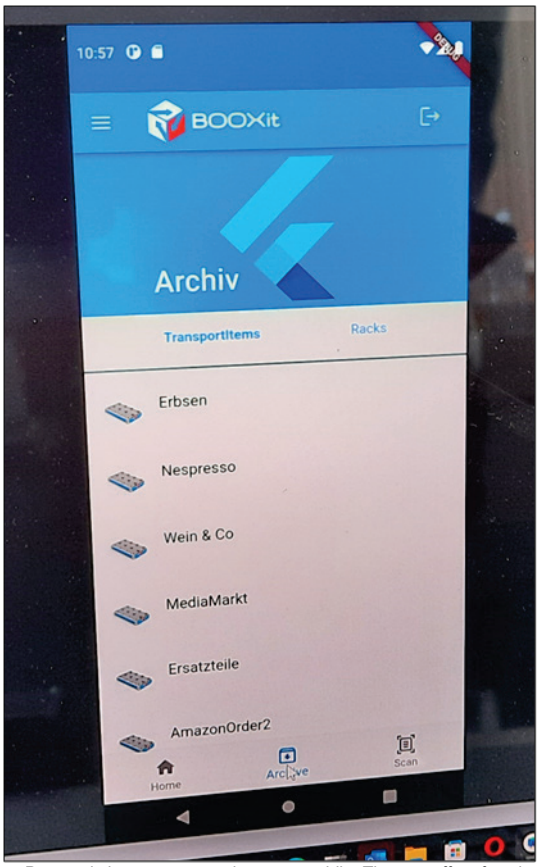


Figure 5 Prototypic inventory control system mobile. The app offers functionalities for RTI management, such as receipts or initiations of shipments, or identifications of RTIs in racks. RTI, returnable transport item.

Later process analysis of the milk run use case (see 4.2.1) revealed that some pharmaceuticals might still need manual loading into boxes. Therefore, the lock mechanism of the core technology was advanced for simplification of the opening of the box to enable subsequent automation. To further facilitate manual loading, the PoC demonstrator box was equipped with an additional mechanism that allows removal of the entire lid. At present, the box supports one-sided lid opening (Fig. 6A) and complete lid removal (Fig. 6B).

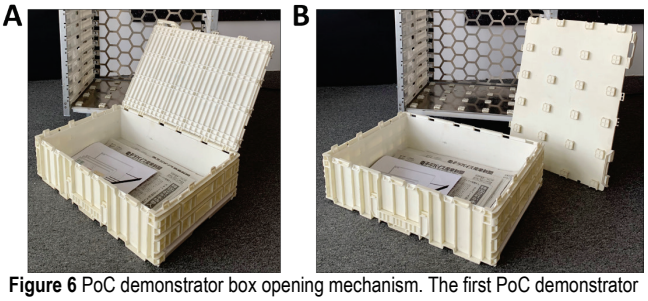


Figure 6 PoC demonstrator box opening mechanism. The first PoC demonstrator of the RTI has an integrated hinge mechanism for one-sided opening (A) or complete detachment of the lid (B). PoC, proof-of-concept.

4.2 Identification and Analysis of Two Potential Use Cases

In order to estimate the logistical improvement potential of the proposed RTI solution, two possible use cases for

deployment were identified together with one industry partner from the pharmaceutical wholesale area and one partner from the shipping sector:

- a wholesaler's pharmacy delivery milk run and
- an international carrier's airfreight shipment of pharmaceuticals that need cooling (pallet shipper).

The use cases were chosen according to the alternative transport containers utilized in each case. Each container represents an ideal target for substitution through either parts or the whole prototypic RTI system.

4.2.1 The Wholesale Milk Run

The wholesale milk run case describes the wholesale supply of pharmaceuticals to local pharmacies, which order various pharmaceuticals for humans and animals, but also goods such as animal feed and other specialty products for animal care. Two types of plastic boxes (standard RTIs and passively cooled polystyrene-based boxes as shown in Fig. 7) are filled in a highly automated picking system and then loaded according to the respective tours.

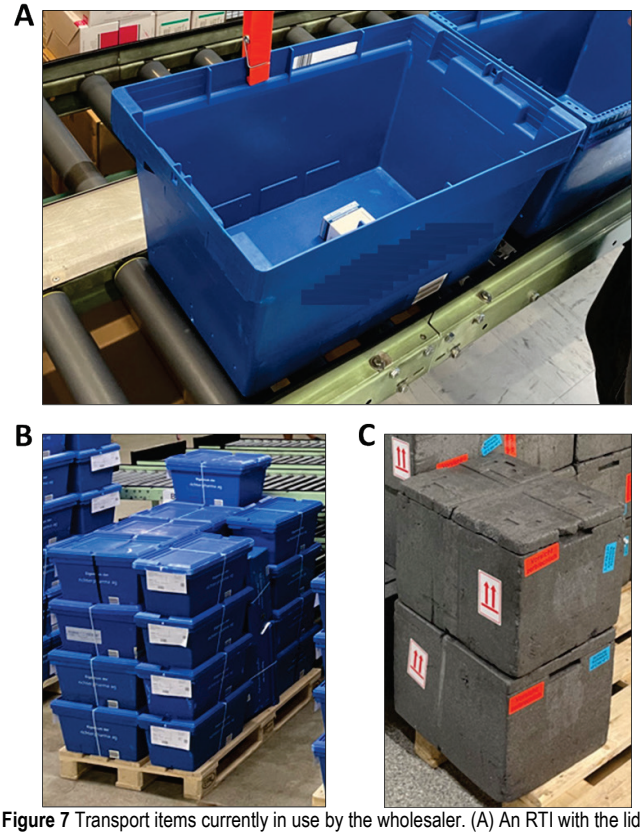


Figure 7 Transport items currently in use by the wholesaler. (A) An RTI with the lid removed. These types of RTIs are used for transportation of drugs that need no cooling. (B) Stacked RTIs. (C) Polystyrene-based cooling boxes to be filled with dry ice for cold chain delivery. RTI, returnable transport item.

Truck drivers sort the boxes corresponding to the stops on their routes and load them into their trucks in reverse order. Each driver then completes his tour and delivers the boxes to the pharmacies on that tour. In the process, the driver also picks up the empty boxes from the pharmacies and returns them to the warehouse at the end of his tour.

While feed bags or bulk packs seem to be less suitable for substitution, the standard box used in the delivery process represents an ideal target for substitution by the proposed prototype. Since the prototype RTI can also be equipped with refrigerated panels, an exchange of the cooled box variant is also conceivable. However, due to simplicity, first calculations were only made with the standard variant. Regarding the economic potential of this use case, the packaging process already seems to be largely automated. However, the sorting and delivery runs are almost entirely carried out manually, thus showing the greatest potential for logistical improvements in terms of cost savings.

4.2.2 The Pallet Shipper

The pallet shipper use case, represents the scenario of a carrier's international pharmaceutical shipping service, specialized to the needs and requirements of specific pharmaceutical business sectors that need temperature control, e.g., manufacture of vaccines or other biotechnology-based products. The process analyzed in this use case thereby consists of arranging the transport from its original destination to the clearing center at the airport and booking the airfreight transportation. Since many goods in the pharmaceutical sector require low temperatures and strict compliance with the cold chain, the means of transportation used in this case is a special isolated pallet shipper, equipped with cooling elements and a temperature logger. Beside the organizational duties mentioned earlier, the carrier is also responsible for providing and assembling the pallet shipper for the customer.

In this use case the pallet shipper represents the means of transport that is to be replaced by the prototype rack. The juxtaposition in Fig. 8 shows the apparent similarities between both systems.

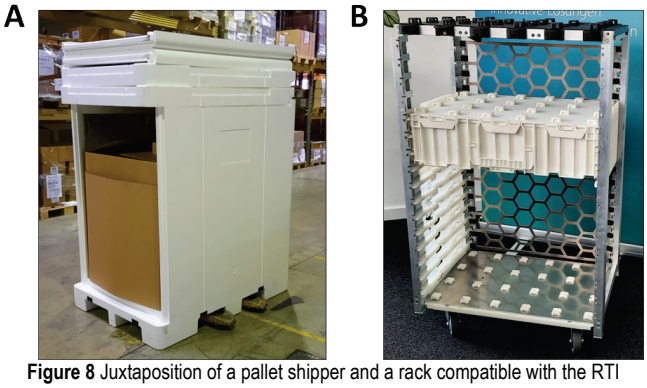


Figure 8 Juxtaposition of a pallet shipper and a rack compatible with the RTI solution. (A) Pallet as currently utilized for international shipments of pharmaceuticals. (B) A first PoC demonstrator designed for the RTI solution to be technically advanced for pharma logistics purposes according to the project objectives. PoC, proof-of-concept; RTI, returnable transport item.

Initial economic considerations of this process suggested that the expense of assembling the pallet shipper, as well as temperature monitoring, could be simplified by the new reusable prototype. However, the current pallet shipper

system is designed as a single-use container and return transport is not currently included in the process. Installing a return transport regimen is an organizational challenge and under current framework conditions will result in considerable additional costs.

In order to evaluate the economic potential of the two use cases (pallet shipper and milk run), actual business process data was exhibited for extrapolating potential cost reduction opportunities from hypothetical implementations of the RTI solution in these cases.

4.3 Logistical Improvements of the Use Cases

Finally, estimating the respective logistical improvement potentials was achieved through the comparison of actual and future cost models, which were derived from underlying business processes and improvement data from expert experiences, pre-tests and videos of the prototypical solution. The potential logistical improvements of those two cases are presented in the following two sections.

4.3.1 Improvements of the Milk Run

For the milk run use case, the team chose to investigate the process section starting after the boxes have left the fully automated picking unit and wait for sorting and delivery till final delivery at the customers as the most promising area for implementation. The primary source for potential savings is time savings during sorting and delivery. The RTI technology facilitates automated or non-automated pre-sorting of filled boxes into the racks directly at the end of the conveyor belt. Thus, staff is not required to manually carry the individual boxes to the assigned truck. Trucks can be equipped with transportation locks that assure the racks are secured in the freight hold. Thus, the racks work as handling aids and load securing device at the same time.

When unloading the boxes at the customers, the RTI system also promises significant time savings. The pick-by light system precludes time-consuming search for the right RTI. Currently, on average a truck driver abides about three minutes at a single customer. With the new technology, this time should be reduced by 38 seconds.

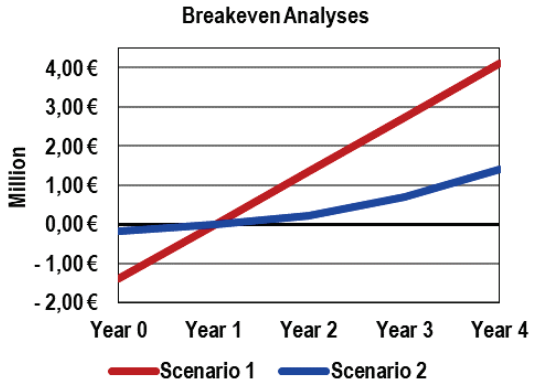


Figure 9 Breakeven analyses for the milk run use case. Scenario 1 assumes full replacement in the first year. Scenario 2 assumes rolling replacement of old boxes with the BOOXit boxes. The x-intercept signifies the year of reaching the breakeven point.

The overall time savings for one tour (currently on average 9.4 hours) sum up to roughly 1.28 hours. Assuming that the company is able to fully utilize the achieved time savings, the calculations show a yearly cost reduction of €1.3 million per year.

Based on the process analysis and optimization a static break-even analysis of two different scenarios was performed. Scenario 1 assumes that all existing boxes are replaced at year 0. Scenario 2 assumes a rolling replacement of old boxes. Under this scenario, complete replacement is achieved after 5.83 years. Irrespective of the scenario, however, the break-even point is reached after almost exactly 1 year (Fig. 9).

4.3.2 Improvements of the Pallet Shipper

Unlike the milk run use case where one RTI technology should be replaced with another RTI technology, the pallet shipper use case investigates the replacement of a single-use system with the RTI and rack system. The pallet shipper use case was selected because of the apparent similarities of the existing pallet shipper and the RTI and rack system (Fig. 8). The RTI system's main advantage in this case is not based on more efficient handling but simply on the cost advantage of reusing the transport boxes. The purchase price of the existing single-use pallet shipper amounts to €1.400. The estimated purchase price of the entire rack with comparable shipping volume is €2,360. The rack should sustain 100 cycles and the boxes 250 cycles. This results in depreciations of €101.12 per cycle.

However, the challenge of this use case is the establishment of a functional return shipment of the boxes. In its current mode of operation, the pallet shipper is primarily used for airfreight, where drugs are sent in irregular intervals and accompanied with individual services. Returning racks with empty boxes over long distances on a low-volume basis proved to be inefficient, both from an economic and an environmental point of view. The calculation yielded an estimated cost increase of 18.4%.

Our obtained insights from investigating the use case are in line with recent literature that highlights transportation distance, delivery frequency, volume, and number of active players as determining factors of the establishment of a return system [17, 18]. At present, these requirements cannot be met by the medium-sized family-owned company in this particular use case. A long-term solution might be the establishment of a circular system with local partners based on frequent orders and stepwise expansion of the system to other businesses.

4.4 A Business Model for the Milk Run Use Case

Based on the preliminary data, a potential business model for time-to-market expedition was derived. The scope of these business models for reaching market entry was set to one or two years after project completion. The business idea builds on a combination of a selling and service model addressing pharma wholesalers and pharmacies as primary

(initial) customer segments, who plan to establish circular economy processes.

The major sources of income are the regular selling of the RTI system (RTIs, racks, IT-infrastructure) and single updates of the global database that supports the box inventory management system. Customers initiate an update each time an RTI enters or exits a rack, which requires updates of its current location and content (e.g., removal of delivered drugs). Additional updates of the database are required by the real-time temperature monitoring feature (if ordered by the customer). Since customers pay per use, single (periodical or manual) inquires of box properties (location, content, temperature log, etc.) will generate income as well.

Next to that core business, the business model also includes service provisions, involving maintenance of the box system and the entire IT-infrastructure. The different building blocks of the initial business model are summarized by the business model canvas according to Osterwalder and Pigneur depicted in Fig. 10 [19].

<table>
<tr><th>Key Partners</th><th>Key Activities</th><th>Value Proposition</th><th>Customer relationships</th><th>Customer segments</th></tr>
<tr><td>•Box manufacturer
•Tool designer
•Developer of a passive cooling system for isolated transport items
•Developer and manufacturer of the electronic equipment
•Developer of the inventory control system
•Box deliverer
•Rack deliverer
•System integrator
•Box cleaning service provider</td><td>•Box selling
•Service provision
•Electronic box management with the inventory control system program software
•Continuous technical and software development
Key Resources
•Box
•Rack
•Box warehouse
•IT-infrastructure</td><td>•Reduction of investment costs
•Reduction of box depletion
•Tracking and tracing of boxes
•Elimination of polystyrene-based cooling boxes
•Elimination of the need to use dry ice for cooling medical products
•Real-time temperature monitoring
•Support of delivery- and warehousemen</td><td>•Business Development
•Customer Service
•Equipment delivery
•Take over of old equipment for recycling purposes
Channels
•Direct sales and distribution
•Direct marketing
•Website
•Service points</td><td>•Pharma wholesaler
•Pharmacies</td></tr>
<tr><th colspan="3">Cost Structure</th><th colspan="2">Revenue Streams</th></tr>
<tr><td colspan="3">•Production per box - scale effect (high initial costs)
•Delivery costs - scale effect (high initial costs)
•Consultancy with and/or execution of the integration of the IT-infrastructure and box technology with existing IT-systems
•Maintenance of boxes and the IT-infrastructure
•Box storage
•Provision of IT resources and access to core data</td><td colspan="2">•Sale of boxes
•Sale of racks
•Sale of IT-resources
•Pay per use - fixed fees for each update of and access to the core data bank
•Temperature monitoring
•Service fees</td></tr>
</table>

Figure 10 Business model canvas of an initial business model

5 CONCLUSION

The preliminary results of the DigiPharmaLogNet project presented in this paper demonstrate self-organizing RTI that are digitally enhanced can substantially foster automation and thus the optimization of the supply chain in pharma logistics. The technological, logistical and economical potential of the novel RTI ecosystem developed by the projects company partners. The two use cases showed that economic feasibility is highly dependent on the specific process and its boundary conditions. While one use case shows a very high return on investment with a break-even after about one year, the second use case is likely harder to implement in a profitable manner, even though it is theoretically feasible. Especially the establishment of the reverse logistics for airfreights seems highly complex and expensive. This coincides with observations from other sectors of the economy (e.g., transportation of milk or other drinks with glass bottles, or RTIs for machine parts), where reverse logistics is only possible, if empty transport capacity is used for the backhaul of the RTIs [11]. In the context of pharma logistics, the presented milk run use case already provides this backhaul opportunity and hence an actual implementation seems promising.

Acknowledgements

The authors would like to thank Andreas Holzleithner and Peter Entenfellner from BOOXit for providing the photo material and the team of DigiPharmaLogNet for sharing their expertise.

6 REFERENCES

[1] Shukar, S., Zahoor, F., Hayat, K., Saeed, A., Gillani, A. H., Omer, S., Hu, S., Babar, Z.-U.-D., Fang, Y., & Yang, C. (2021). Drug Shortage: Causes, Impact, and Mitigation Strategies. *Frontiers in Pharmacology, 12*, 693426. https://doi.org/10.3389/fphar.2021.693426

[2] Francas, D., Mohr, S., & Hoberg, K. (2023). On the drivers of drug shortages: empirical evidence from Germany. *International Journal of Operations & Production Management, ahead-of-print*. https://doi.org/10.1108/IJOPM-09-2022-0581

[3] Stoller, C. W. (Ed.) (2017). *Pharmalogistik*. Springer Fachmedien Wiesbaden, Wiesbaden. https://doi.org/10.1007/978-3-658-15264-2

[4] Yu, H., Sun, X., Solvang, W. D., & Zhao, X. (2020). Reverse Logistics Network Design for Effective Management of Medical Waste in Epidemic Outbreak: Insights from the Coronavirus Disease 2019 (COVID-19) in Wuhan. *SSRN Journal*. https://doi.org/10.2139/ssrn.3538063

[5] Koskela, S., Dahlbo, H., Judl, J., Korhonen, M.-R., & Niininen, M. (2014). Reusable plastic crate or recyclable cardboard box? A comparison of two delivery systems. *Journal of Cleaner Production, 69*, 83-90. https://doi.org/10.1016/j.jclepro.2014.01.045

[6] European Commission. (2023). European Green Deal: Putting an end to wasteful packaging, boosting reuse and recycling. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7155, accessed 20 February 2023.

[7] Siepmann, D. & Graef, N. (2016). Industrie 4.0 – Grundlagen und Gesamtzusammenhang. In Roth, A. (Ed.), *Einführung und Umsetzung von Industrie 4.0*. Springer Berlin Heidelberg, Berlin, Heidelberg, 17-82. (in German) https://doi.org/10.1007/978-3-662-48505-7_2

[8] Strandhagen, J. O., Vallandingham, L. R., Fragapane, G., Strandhagen, J. W., Stangeland, A. B. H., & Sharma, N. (2017). Logistics 4.0 and emerging sustainable business models. *Adv. Manuf., 5*, 359-369. https://doi.org/10.1007/s40436-017-0198-1

[9] Ding, B. (2018). Pharma Industry 4.0: Literature review and research opportunities in sustainable pharmaceutical supply chains. *Process Safety and Environmental Protection, 119*, 115-130. https://doi.org/10.1016/j.psep.2018.06.031

[10] Gouda, M. (2017). Pharmaceutical Distribution in Europe—The Drivers and Challenges of Distributing Pharmaceutical Products in the European Market. In Stoller, C. W. (Ed.), *Pharmalogistik*. Springer Fachmedien Wiesbaden, Wiesbaden, 167-184. https://doi.org/10.1007/978-3-658-15264-2_11

[11] Coelho, P. M., Corona, B., ten Klooster, R., & Worrell, E. (2020). Sustainability of reusable packaging–Current situation

and trends. *Resources, Conservation & Recycling: X, 6*, 100037. https://doi.org/10.1016/j.rcrx.2020.100037

[12] Shashi, M. (2022). Digitalization of Pharmaceutical Cold Chain Systems using IoT Digital Enabler. *IJEAT, 11*, 133-137. https://doi.org/10.35940/ijeat.E3622.0611522

[13] Li, H. & Wang, X. (2021). Drug Traceability System Based on RFID and Alliance Block Chain Technology. Abawajy, J., Xu, Z., Atiquzzaman, M., & Zhang, X. (Eds.). *International Conference on Applications and Techniques in Cyber Intelligence*, Springer International Publishing, Cham, 655-661. https://doi.org/10.1007/978-3-030-79200-8_97

[14] Fanti, M. P., Mangini, A. M., Roccotelli, M., & Silvestri, B. (2020). Hospital Drugs Distribution with Autonomous Robot Vehicles. *The 16th IEEE International Conference on Automation Science and Engineering (CASE)*, 1025-1030. https://doi.org/10.1109/CASE48305.2020.9217043

[15] Parker, C., Evens, E., Stankevitz, K., Eichleay, M., Homan, R., Loogos, M., & Onyango, D. (2021). Adding unmanned aerial vehicles to HIV supply chains in remote settings: modeling feasibility and cost in Turkana, Kenya. *Journal of Global Health Reports, 5*. https://doi.org/10.29392/001c.28349

[16] See https://www.fh-wien.ac.at/forschung/forschung-an-der-fhwien/digipharmalognet/ (Accessed March 30th, 2023).

[17] Limbourg, S., Martin, A., & Paquay, C. (2016). Optimal returnable transport items management. *World Conference on Transport Research - WCTR 2016*, Shanghai. https://orbi.uliege.be/bitstream/2268/200983/1/Optimal%20returnable%20transport%20items%20management.pdf

[18] Taschner, A. (2023). Returnable transport packaging as a company value driver. *BIJ, 30*, 196-214. https://doi.org/10.1108/BIJ-06-2021-0311

[19] Osterwalder, A. & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. Wiley, Hoboken, NJ.

Authors' contacts:

Gerald Schneikart, MBA, PhD
(Corresponding author)
FHWien der WKW, Institute for Digital Transformation and Strategy,
Währinger Gürtel 97, 1180 Vienna, Austria
+43 1 476 77 5761, Gerald.Schneikart@fh-wien.ac.at

Walter Mayrhofer, Prof. Dr.
FHWien der WKW, Institute for Digital Transformation and Strategy,
Währinger Gürtel 97, 1180 Vienna, Austria

Josef Frysak, Dr.
FHWien der WKW, Department of Digital Economy,
Währinger Gürtel 97, 1180 Vienna, Austria

Clemens Löffler, DI Dr.
FHWien der WKW, Institute for Digital Transformation and Strategy,
Währinger Gürtel 97, 1180 Vienna, Austria

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230511165415

Preliminary communication
Received: 2023-05-11, Accepted: 2023-05-15

Towards Logistics 4.0: A Skill-Based OPC UA Communication between WMS and the PLC of an Automated Storage and Retrieval System

Andreas Lober*, Lisa Ollinger, Sven Voelker, Hartwig Baumgaertel

Abstract: In order to bring intralogistics systems to the same level of interoperability as today's modern production systems, logistics must take the essential steps towards Industry 4.0. This requires an increasing abstraction level of control logic as an enabler for horizontal and vertical integration. The abstraction will lead to the interconnection of manufacturing and logistics control with the production planning and warehouse management systems (WMS). A main enabler for these communication paths are service-oriented architectures (SoA). OPC UA has established itself as a widely used and already adopted SoA-based communication standard in industry. The paper describes the realization of an OPC UA-based approach for the communication between a WMS and a PLC of an automated storage and retrieval system (ASRS). The conceptual basis of communication design are skills of the ASRS. The work is supported by an architectural design with a subsequent prototypical implementation.

Keywords: Control Logic; Industry 4.0; Logistic System; OPC UA; Skill-based Engineering

1 INTRODUCTION

The field of intralogistics is experiencing considerable progress in terms of flexibility and adaptability addressed by Industrie 4.0. Replacing individual interfaces with standardized interfaces is a central goal of Industrie 4.0 to enable a holistic network architecture. It has become apparent that production and logistics still have considerable hurdles to overcome for reaching this target [1]. Resources from different system providers are often used for logistics subsystems, which means that a large number of different interfaces must be managed. To overcome this interface problem, an increasing degree of abstraction of the control logic is required as an enabler for horizontal and vertical integration. Increased abstraction enables a linking of technical manufacturing and logistics control with production planning and warehouse management systems. Service-oriented architectures (SoA) are an important enabler for a flexible and adaptive communication pattern.

Concerning the vertical dimension of communication, there is still a noticeable gap between the control layer and higher automation functions for monitoring the production plan [1]. This discrepancy is reflected in different requirements in terms of timing constraints and synchronous/asynchronous communication patterns. OPC Unified Architecture (OPC UA) has emerged as a widely used and already adopted SoA-based communication standard in industry. Traditional communication approaches between Operations Technology (OT) and IT systems cause high development effort, long commissioning times, and inefficiencies, as they require individual arrangements by programmers on both sides for each individual interface. Despite extensive scientific work and trade show demonstrations over the last decade, general migration concepts have not been established, and experts who can perform such a transformation are rare. The migration of new approaches in legacy technical systems into cyber-physical systems poses various challenges, such as non-standard control logic or the use of different programming languages and end devices.

In order to be able to simulate such a scenario in an industry-oriented way and to implement a skill-based control logic via OPC UA, this paper considers the technical implementation of the conversion of an automated rack storage system (ARSS) for small receptacles according to the concept of skill-based engineering (SBE). For this purpose, the original control system, consisting of an industrial PC with Soft PLC, ProfiBus PCI card, and a proprietary Warehouse Management System (WMS) was replaced by a PLC and a commercial WMS. In the next step, a concept for the conversion and transition to a modern, networked architecture was developed. In this concept, OPC UA is used as the communication and information standard. Hence, a Siemens S7-1500 PLC was chosen as the central control unit, as this PLC type is equipped with an OPC UA server. The information model for this server was designed based on a skill model for the ARSS. The last step, for now, is the implementation of a new WMS, which communicates with the ARSS' PLC exclusively via OPC UA. User activities are translated by this WMS into calls of OPC UA methods at the PLC. The paper presents a successful implementation of OPC UA based on the SBE approach, which is not based on OPC UA programs like many existing approaches, but on OPC UA methods. As these methods are applied to automation functions, a so-called skill-based control concept is realized. This allows to use digital twins (DTs) from the beginning of the engineering process [2]. Those DTs can proactively negotiate an optimal production mode that choreographs production capabilities within a market scenario [3]. In addition, the advantages and disadvantages of implementing this approach are discussed.

After discussing related work in Section 2, the authors present the use case of an automated rack storage system in Section 3. After presenting the skill model and communication architecture in Section 4, the implementation of this architecture is described in Section 5. In Section 6, the presented approach is discussed and the advantages and disadvantages of the implementation are presented. Finally, a conclusion is drawn in Section 7 and an outlook on future work is given.

2 RELATED WORK

Research has shown that several concepts built on a service-oriented architecture (SoA) are promising for representing production systems knowledge and capabilities [4]. The SoA concept is characterized by higher flexibility in contrast to the traditional component-based hierarchical and less complex approaches [5]. SoA is a type of possible software design that modularizes previously monolithic IT systems. It is based on the concept of callable services. These services must provide access to the capabilities of a software or hardware system through a predefined interface through a service description [6]. SBE is described as a possible implementation for the implementation of SoA callable services for industrial applications [7]. SBE covers an extensive research area, and therefore approaches dealing with the skill description of production machines are discussed in the following.

2.1 Skill-based Engineering in Production

The research area of automation technology shows that the approach of skill-based engineering is becoming increasingly important. Many approaches focus on mapping capabilities in such a way that possible interfaces are simplified and communication technologies can be easily integrated.

Zimmermann et al. [7] describe three prerequisites for the successful introduction of skill-based engineering in the creation of PLC programs. For this, the machine manufacturer provides SBE automation components and the application engineer orchestrates the provided skills and creates the final process and program flow. In addition, a cross-manufacturer standardization of skills is critically considered. Dorofeev [8] shows a generic interface with specific functionalities in terms of the details of SBE. A metamodel of capabilities represents semantic information. Perzylo et al. [9] describe a knowledge-based digital SBE concept for semantic integration of various data and information along the value chain of manufacturing companies. In Köcher et al. [10], a method for automating and supporting the development of machine capability and capability models is outlined. The presented method enables mechanical engineers and machine developers to create formal models of machines and their abstract capabilities without additional effort. A further work that builds on this by Köcher et al. [11] derives semantic skill models from existing PLC code. However, a holistic approach would derive a PLC code hierarchy from a semantic skill model to derive the skill-based PLC code directly.

According to the literature, a skill-based control logic structure that uses all aspects of OPC UA offers the greatest advantages in the implementation of production systems. Therefore, further consideration will focus on approaches that combine SBE with OPC UA.

2.2 OPC UA & Skill-based Engineering

The combination of the skill-based engineering approach and OPC UA technology has already been used in previous work. To be able to classify the present work, the best-known approaches of SBE and OPC UA are presented below.

Dorofeev and Zoitl [12] demonstrate an SBE approach for OPC UA programs using a state machine to describe the execution logic of skills. As a possible future work, the authors mention the implementation of a skill as a pure OPC UA program. Dorofeev and Wenger [5] compare a SBE approach with a classical hierarchical control logic approach. Despite the capability-based architecture using nonreal-time communication via OPC UA, the performance is high compared to the traditional approach according to IEC 61131-3. Zimmermann et al. [7] do not only focus on the use of SBE for technical purposes but use a SBE control and show that such an architecture can, in principle, be realized with OPC UA. The paper highlights that the current implementation is mainly based on the client/server model of OPC UA and the use of methods over TCP/IP limits the application possibilities due to the lack of real-time synchronization between different components.

Spitzer et al. [13] propose an approach in which the core concept of SBE programming is implemented via capabilities, variables, configurations, and messages. In doing so, the functionality of a device is moved to an abstract level. This abstraction supports a clear separation between orchestration and the subsequent execution of production processes. Profanter et al. [14] present a generic system architecture for a plug & produce system. OPC UA is chosen as the basis for this system. Using a combination of OPC UA's decentralized discovery mechanism and a skill detector, the system automatically detects newly added components and their skills. Jhunjhunwala et al. [15] show that building a SBE architecture using the digital twin concept defined in IEC 61499 with OPC UA is a feasible approach as an enabler for the capability structure of a production system.

A feature in [13] is the use of a specific OPC UA part, the OPC UA programs. Other approaches are using OPC UA methods based on services that are offered by the devices. These approaches call services via OPC UA, whereby the complexity of the control logic remains implemented on the PLC, since the services are programmed as methods on the PLC. In contrast to those approaches, this article will focus on implementation without OPC UA programs and complete services as OPC UA methods, and shows that it is possible to successfully implement a control system for the ARS system based on so-called basic skills that are implemented as OPC UA methods. Basic skills represent basic capabilities of the device that are needed, for example, to offer a certain service. These basic skills are put into a specific order and are executed sequentially or in parallel. The cited work describes the application of the approaches in scenarios where the work of three types of experts is required. The work presented here offers a different approach that has been implemented in a student project. However, as preparation, the capabilities had to have been implemented as callable OPC UA methods prior to the student project. Because otherwise it would not have been possible for the student group to implement their solution, since they had no knowledge about PLC programming and were not supposed to work on the PLC.

3 USE CASE – AUTOMATED RACK STORAGE SYSTEM

An automated rack storage system (ARS) for small receptacles is the considered use case of this paper (Fig. 1). The ARS system consists of one row of racking with 11 rack columns and 9 rack levels, where each compartment can store one small standard bin. The bins are stored and retrieved using an automated stacker crane (ASC), which has a load carrying device (LCD) mounted on its mast. Two belt conveyors connect an operator workplace with the ARS, one for the bins to store and one for the retrieved bins. On the operator workplace side, a transfer carriage connects the two conveyors. Between the conveyor belts, several airlift pumps are installed that can lift boxes from the belt and fix them in a specific position each. This realizes the functionality of accumulating conveyors for the belt conveyors. After the introduction of the new Siemens PLC at the operations technology level, the ARSS was operated at the user level by a commercial WMS. This WMS and the PLC exchanged commands and data by ASCII files called telegrams, which are transmitted via the TCP protocol. Both systems used a Siemens software library to handle telegrams. This library contains routines for buffering, parsing, and interpreting incoming telegrams, and creating and sending telegrams to the other party. Typical action orders are "store a bin to a specified bin location", "retrieve a bin from a specified bin location", and "read the ID barcode of the bin at a specified bin location". A WMS user initiates these orders via the WMS graphical user interface. A problem of the used standard telegram software library is that it offers many parameters for each command whose semantics is not obvious in many cases. This causes high effort at commissioning as experts for PLC programming and WMS programming need to agree on the usage and semantics of parameters for all used telegram types.

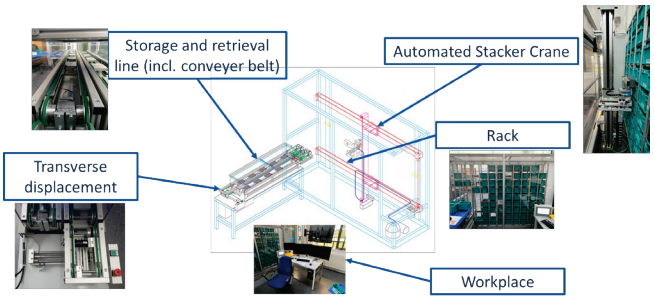


Figure 1 Automated Rack Storage System

In most cases, the PLC sends telegrams to the WMS as responses to received telegrams, e.g. orders or status inquiries. However, the PLC sends a telegram proactively when a bin arrives at the store-in location. Then, the WMS reacts with corresponding action order telegrams. The action orders from the WMS are split by the PLC into several basic activities, like moving the ASC in *X* and *Z* directions, moving the load carrying device towards the rack or backward (*Y* direction), grasping a bin by the LCDs fork, and releasing a bin from the fork when it is placed. Some of these actions can be performed in parallel, like movements in the *X* and *Z* direction. Others require an exactly planned sequence of actions, such as grasping a bin as a combination of movements in the *Y* direction, the *Z* direction upward to lift the bin from its location, and back in the *Y* direction to pull the bin out of the rack. The sequence of actions needed to store a bin is distributed over PLC and WMS. After a new bin is detected at the store-in location by an occupation sensor, the ASC has to move to this location, read the bin ID from a barcode at the front of the bin and pick the bin to its fork.

Then, PLC and WMS exchange telegrams with the bin ID and, if valid, the target bin location. After that, the ASC moves to the target location, checks if it is free, and, if yes, finally puts the bin to this location. Depending on the operation strategy, the ASC can then proceed with retrieving a bin or moving back empty to the store-in location. In most cases, the decision on where to put a bin and what to do as the next step is made by the WMS. However, other parts of the action sequence could be either hard-coded in the PLC or initiated by the WMS. The PLC and WMS programmers agree on how this is organized during the commissioning of the ARSS. As more than one order to the ARSS may be due at a time, the WMS maintains an order queue, as the ASC can process only one order at a time and needs time for that.

To demonstrate the skill-based OPC UA approach to communication between the WMS and the PLC of the ARSS described above, a new WMS was designed and developed by the project team. For this development, the following requirements for the WMS were derived:

- An interface for OPC UA must be established between the WMS and the ARS that enables various services such as "putaway", "retrieval" and "relocation" based on skills.
- The WMS communicates with the ARS exclusively via OPC UA. No other communication channel may be used.
- The WMS provides an inventory function, updates its database accordingly, and sends all non-readable containers to the operator workstation.
- The WMS provides an identification point, i.e., functionality that allows a bin ID to be associated with a material ID of the bin contents.
- The WMS focuses on the operation of the ARS and does not cover other process areas in real warehouses such as goods receipt or picking.
- The data model of the WMS must be as simple as possible, but as realistic as necessary. It must separate the master data of the rack, storage bins, and material from the movement data. It should be possible to change master data via special GUI views restricted to authorized users.

The following chapter describes in detail the underlying skill model of the ARSS and the corresponding communication architecture.

4 SKILL MODEL AND COMMUNICATION ARCHITECTURE

In order to be able to handle a skill-based control logic via an OPC UA communication architecture, the ARS must

first be defined in terms of its services and skills. For the definition of individual skills, a distinction must first be made between processes and tasks (or services).

In this case, the ARS presented has three tasks: Storing, Retrieving, and Relocating. These different tasks can be realized by a specific sequence of skills. In the Storing task example, the following skills are called first:

1) Convey bin
2) Check occupation of storing location
3) Move Stacker Crane
4) Identify bin
5) Pick bin
6) Move Stacker Crane
7) Check occupation of target location
8) Put bin.

In that case, the third and sixth skills are so-called combined skills, which are a combination of two skills performed in parallel. The skill model thus distinguishes combined skills from simple ones. A combined skill corresponds to skills that execute different capabilities in parallel, e.g. to realize a new type of skill. In the example shown, this corresponds to the diagonal movement of the stacker crane, which is realized by the parallel execution of the movement in the *x*- and the *z*-direction. By combining these eight standard skills, the three services mentioned can be realized in the use case presented. The services differ in the orchestration of the skills. For the realisation of a service it is possible that certain skills are not required or are called in a different order compared to another service. The complete skill model is shown in Fig. 2. Based on this skill model, the control logic in the PLC of the ARS system was adapted. For this purpose, the OPC UA server was set up on the PLC and the information model was built based on the skill model.

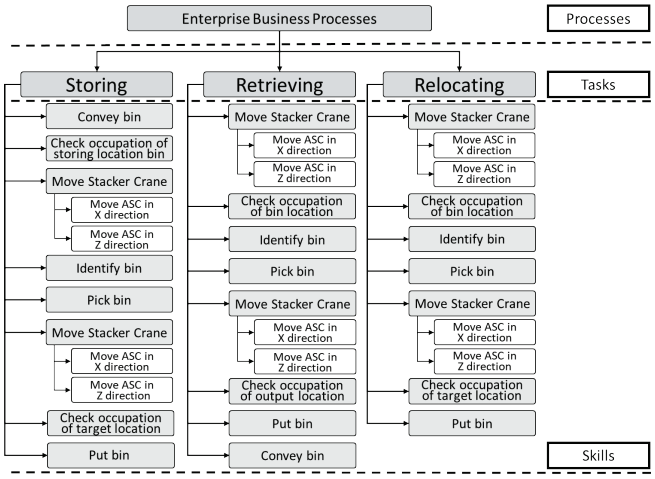


Figure 2 Skill model of the ARS

While the connection of PLC variables with variable nodes in the IM is simple and straightforward, the integration of callable methods requires the implementation of specific wrapper function blocks, whose main task is to transfer the OPC UA method parameters to the PLC code and vice versa. Based on the skill model and system constraints, finally, sixteen callable OPC UA methods are implemented. They can be classified into five groups:

- Methods to be executed by the ASC (including the LCD), e.g. "Put bin to", "Pick bin from", "Read barcode of bin at position"
- Method to be executed by the conveyor at the bin retrieval position, "Move bin to outbound conveyor"
- Methods to be executed by the conveyor and the airlift pumps in bin retrieval direction, e.g. "Move bin from pos L to pos L+1"
- Methods to be executed by the conveyor and the airlift pumps in store-in direction
- Methods to be executed by the transfer carriage at the end of the conveyors, e.g. "Pick bin and move in front of storing conveyor", "Put bin".

The OPC UA methods and the skills from the skill model are not identical. The ASC methods aggregate several basic skills. For example, the method "Pick bin from" consists of the basic skills Move Stacker Crane, Identify bin and Pick bin, and the method "Put bin to" of Move Stacker Crane, Check occupation of target location and Put bin. These aggregated skills occur in all three task processes of the skill model.

That is, they can be seen as reusable building blocks for the operation workflows. On the other hand, the skills from the skill model regarding the conveyors needed to split into several OPC UA methods, as the accumulating behavior of the conveyors was not considered. The skill Check occupation of storing location is not realized as an OPC UA method but can be mapped to a subscription to the corresponding variable by OPC UA clients.

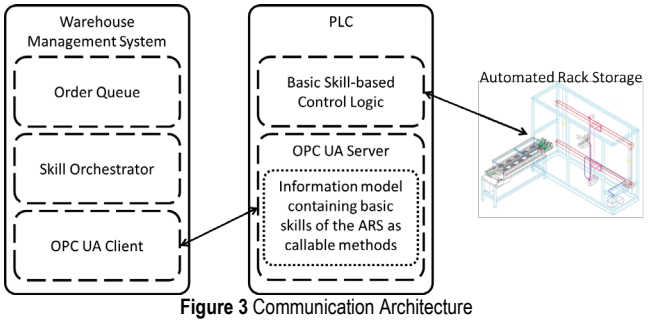


Figure 3 Communication Architecture

Based on these two adjustments to the system, the following communication architecture between the WMS and the PLC can be built (see Fig. 3). The implemented control logic within the PLC can be divided into two areas. First, the basic skill-based control logic is implemented in the PLC in the code, which ensures that the PLC only executes functions based on basic skills. Second, the PLC contains the OPC UA server with an integrated information model that integrates various skills of the ARS system as callable methods. The PLC is connected to the physical ARS via a standard Ethernet port. Its counterpart is the WMS, which has a total of three components: the Order Queue, the Skill

Orchestrator and the OPC UA Client. The order queue manages the orders that are placed with the ARSS and stores them temporarily so that the orders can be processed according to a certain logic.

Based on the information model of the ARSS, the Skill Orchestrator can split the tasks into individual skills and send them as method requests via the OPC UA Client to the PLC.

Fig. 4 shows an example communication flow of the WMS with the PLC via OPC UA between the frontend, backend and PLC. If certain tasks, such as a "store-in bin", are to be managed via the WMS, a command is first transmitted to the back end via the front end through REST. The process request is started in the back end by transmitting a request ID to the front end and triggering the necessary skills via OPC TCP. By issuing commands via the backend of the WMS, the PLC starts the physical processing of the request by controlling the actuators. After a certain time, the backend iteratively checks the status of the request by issuing a command to the PLC via OPC TCP to check the status. The PLC passes the current status of the skills to the backend, which in turn updates the status in the WMS. The frontend can iteratively query the status via the request ID or receives a response about the current status when the process is fulfilled or accepted.

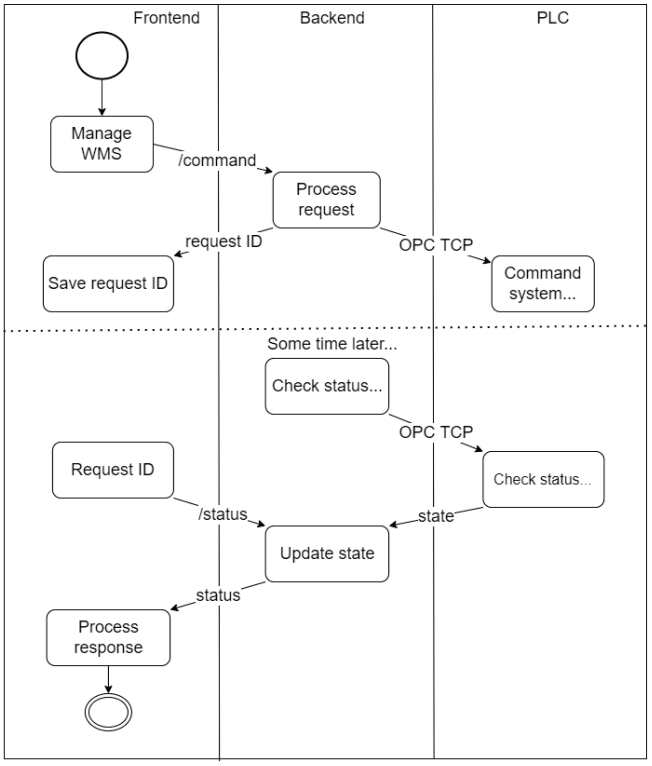


Figure 4 Basic Communication principle

5 IMPLEMENTATION

The WMS system of the ARS is divided into a backend and a frontend part (see Fig. 5). The frontend realizes the Graphical User Interface (GUI) as a thin web client written in VueJS. A Flask web server serves it, which is part of the backend system. The server offers a RESTful API to the web client. The backend system is written in Python and contains, beneath the Flask server part, a database interface to a SQLite database, an OPC UA client. The core logic of the program contains an order queue for user orders from the GUI, procedures that translate user orders into sequences of skills, and the basic methods to call OPC UA methods on the PLC that realize single skills. The Graphical User Interface (GUI) consists of a grid part on the left-hand side and a table and activity part on the right-hand side (see Fig. 6).

The grid part offers several views of the rack content and interacts with the table view, which offers functionalities like

- search for bins or material,
- user actions like retrieving, relocating, or storing bins,
- ARS maintenance like banning or releasing storage locations, assigning material numbers and amounts to bins (the so-called identification point or I-point)
- assigning storage locations to material numbers for fixed position storing strategy, and
- settings of the WMS like the choice of a storing strategy (like fixed position storing, chaotic storing, storing in ABC zones according to the turnover frequency of the material, and manual determination of storing locations).

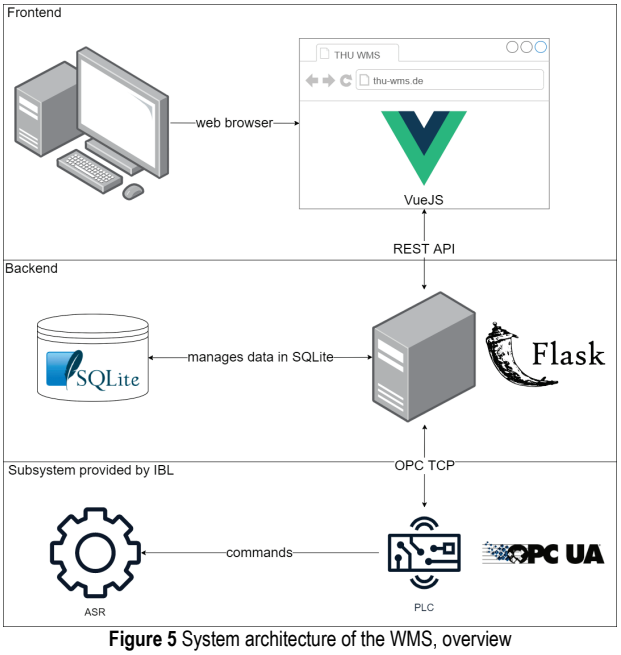


Figure 5 System architecture of the WMS, overview

The grid views contain visualization of the results of the current search, in which matching locations are highlighted, visualisations of the rack zones for A, B, and C classified material, banned locations, and more. An important insight from the tests of the first version of the software in the lab was that the skill orchestration needed to be split for the several technical components of the storage system, like the stacker crane, the conveyor belts, and the transfer carriage, and let these components work in parallel. This was not explicitly required from the beginning, because it was taken for granted by the laboratory staff, as they were used to it from the previous system. In the current system, the PLC takes care of the parallel execution.

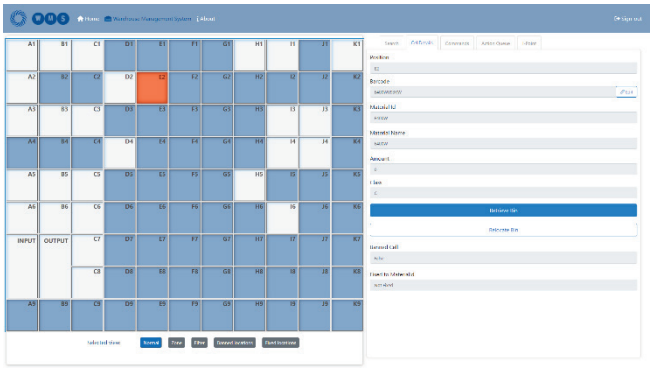


Figure 6 GUI of the WMS, grid view (left), command view (right)

The first version of the WMS software conducted all skill method calls sequentially. This led to the observation that the stacker crane was blocked as long as a conveyor belt conveyed a bin. As this is obviously not a desired behavior, an additional requirement was stated that the system components should work in parallel.

This caused a serious re-design of the skill orchestration part of the backend software. In the first version, only two threads were used: one for the web server to be responsive to the GUI client, and one for the whole ARS operation. The second approach used separate threads for each technical component of the ARSS, like the stacker crane, the two conveyor belts, and the transfer carriage. For each thread, a separate action queue was introduced. Tasks in the queues are either stated by skill orchestration for user commands or in case of the handover of bins from one technical component to another (see Fig. 7). The introduction of these threads leads to the expected behavior of the ARSS components as known from pure PLC control: different components can operate in parallel as long as their activities are independent and synchronized when they have to coordinate their activities, e.g. in the case of bin hand over from one component to another.

Fig. 7 shows an example of the coordination of two threads, one of the ASC and one of a conveyor belt.

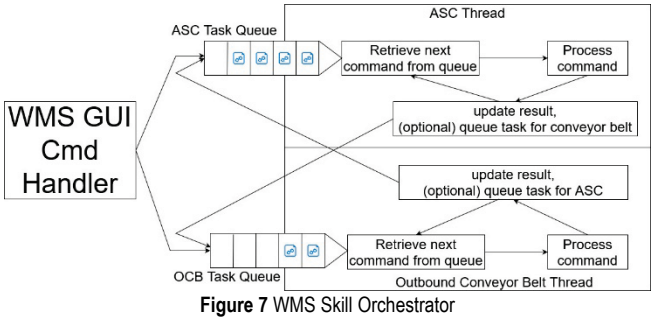


Figure 7 WMS Skill Orchestrator

The design and implementation of the new WMS was conducted as a study project of Computer Science students in their 7th semester. Based on the requirements and additional specifications and recommendations, the students designed and implemented the software up to a certain level of maturity according to an agile software development approach.

A description of this student project and its evaluation and students experiences is reported in [16]. After some further development, the new WMS is now stable and used for student laboratory courses with the ARSS.

6 DISCUSSION & CONCLUSION

The presented approach shows the application of a skill-based control logic that can execute different skills of an ARSS via an OPC UA communication architecture. The WMS makes the skill requests to the PLC based on the OPC UA methods implemented in the PLC. The skill orchestrator implemented in the WMS takes over the task of aggregating the skills into an executable service. Through the use case presented here, it could be shown that some of the skills in the different services only had to be "re-orchestrated" to enable a new service. Based on the information model of the ARSS, the Skill Orchestrator can split the services into individual skills and send them as method requests to the PLC via the OPC UA client.

The advantages over other approaches are the extended reconfigurability and the implementation of new services based on existing implemented basic skills. Due to the skillset, new services can be implemented more easily via the Skill Orchestrator based on an information model of the resource. Conventional approaches that implement their services as OPC UA methods in the PLC also require an adaptation in the PLC program when creating new services. In the PLC program, these approaches require new control sequences to be implemented as OPC UA methods. In the approach presented here, an adaptation in the PLC program is only necessary if the new service cannot be realized by the basic skills already implemented. A new service based on the already known basic skills can be implemented faster by an adaptation in the Skill Orchestrator. The unfavourable situation by applying the standard telegram software library for calling services implemented on the PLC whose semantics are not unique in many cases, cannot occur with this approach. This problem, as with the services implemented as OPC UA methods, causes a high effort during commissioning and adaptations, since experts for PLC and WMS programming have to agree on the use and semantics of the parameters for all telegram types used. This coordination problem would not be necessary with a PLC based on a basic skill control logic, since the skills could be implemented directly as OPC UA methods and the necessary services for the WMS could be compiled from them. However, since such an approach requires the WMS programmer to understand which basic skills make up a swap service. The research on this approach will be continued so that, on the one hand, an automatic generation of OPC UA methods from the basic skill-based control hierarchy in the PLC is implemented, and, on the other hand, an automatic orchestration of these OPC UA methods into a service based on the OPC UA information model.

7 OUTLOOK

The project provided many insights into different topics, from skill definition via the creation and implementation of the OPC UA information model on the Siemens PLC to the

necessary design of the client software to operate the technical system properly.

For the skill definition, it turned out that the initial skill design was not complete. The skill set of the conveyor belts needed to be extended by skills that allow conveying bins over longer distances without lifting them at every single location. These skills should be used if no other bins are on the belts that force a bin to wait. For the ASC, a skill was missing that checks whether a storage location is free or not. The skill for reading a bin ID barcode is not suitable for this as in the case that no valid bin ID could be read it cannot distinguish if the reason for this was that the location was empty or the barcode of an existing bin was not readable. It will be analyzed in a further step of the project, how these mistakes in the skill definition could have been avoided in the design phase. However, the PLC program is currently updated, and the missing skills will be added thereby.

In a further branch of the project, a digital twin of the ARSS will be created that can be controlled by the same WMS as the original one. The basis for that digital twin is a simulation and 3D animation model of the ARSS that already exists. To extend this model to a digital twin, it has to be enhanced by control logic of the simulation elements that corresponds to the behavior of the real system components and accompanied by a new software module as interface to the WMS. This new module has to include an OPC UA server like the ARSS' PLC, so that the WMS can connect to it as an OPC UA client, and needs to be linked to the simulation model elements control units to steer them like the PLC steers the physical components of the ARSS. Finally, the simulation model has to create values of parameters that are mapped to variables of the OPC UA information model so that the WMS can monitor these variables in the same way as it monitors variables on the PLCs OPC UA server.

A further extension of the physical ARSS was started recently: the ASC will be extended by an RFID reader and its communication infrastructure to accompany or as alternative to the existing barcode reader. All receptacles of the system are already equipped with RFID tags according to the ISO 15693 standard, i.e. with a frequency of 13,56 MHz, and a readable and writable memory of 112 Bytes. The RFID reader will be mounted on the load carrying device. It can then read and write the tags of the bin in front of it. This offers several advantages compared to the barcode reading, which is sometimes unreliable, since it needs good light conditions for successful reading. The reader will be connected to a communication module. This module could be connected to the PLC via ProfiBus, but this would require an adaptation of the PLC program as well. Nevertheless, the communication module can also be connected to a computer network via an Industrial Ethernet connection and an OPC UA server with an information model for AutoID devices on board. This offers the opportunity to connect it directly to the WMS without affecting the PLC. As the WMS is an OPC UA client, it can establish connections to different OPC UA servers at the same time by instantiating the OPC UA client object more than once. As soon as this RFID system is commissioned and the TCP/IP connection is established between the communication module and the WMS machine, the WMS code can be extended to access to this second OPC UA server and the skill orchestration can be extended by the skills of the RFID system. An OPC UA client will use this as a proof of concept for the orchestration of skills of different technical systems into common services. This can generalized to a concept for easy integration of new technical components into Service-oriented control architectures for production and logistics systems.

8 REFERENCES

[1] Schuh, G., Anderl, R., Dumitrescu, R., & Krüger, A. (2020). Using the Industrie 4.0 Maturity Index in Industry - Current challenges, case studies and trends. acatech COOPERATION. http://www.acatech.de/wp-content/uploads/2020/04/aca_KOOP_MatInd_en_Web.pdf

[2] Lober, A., Lehmann, J., Hausermann, T., Reichwald, J., & Baumgartel, H. (2022). Improving the engineering process of control systems based on digital twin specifications. In *the 4th IEEE International Conference on Emerging Trends in Electrical, Electronic and Communications Engineering (ELECOM)*. https://doi.org/10.1109/ELECOM54934.2022.9965259

[3] Lehmann, J., Lober, A., Rache, A., Baumgärtel, H., & Reichwald, J. (2022). Collaboration of Semantically Enriched Digital Twins based on a Marketplace Approach. *International Conference on Wireless Information Networks and Systems*, SciTePress. https://doi.org/10.5220/0011141200003286

[4] Leitao, Colombo, K. (2016). Industrial automation based on cyber-physical systems technologies: Prototype implementations and challenges. Emerging ICT concepts for smart, safe and sustainable industrial systems. *Computers in Industry, 81*, 11-25. https://doi.org/10.1016/j.compind.2015.08.004

[5] Dorofeev, K. & Wenger, M. (2019). Evaluating skill-based control architecture for flexible automation systems. In *the 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. https://doi.org/10.1109/ETFA.2019.8869050

[6] OASIS. (2006). Reference Model for Service Oriented Architecture 1.0: Committee Specification 1. http://www.oasis-open.org/committees/download.php/19679/soa-rm-cs.pdf (Accessed 27 March 2023)

[7] Zimmermann, P., Axmann, E., Brandenbourger, B., Dorofeev, K., Mankowski, A., & Zanini, P. (2019). Skill-based engineering and control on field-device-level with OPC UA. *The 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Zaragoza, Spain, 1101-1108. https://doi.org/10.1109/ETFA.2019.8869473

[8] Dorofeev, K. (2020) Skill-Based Engineering in Industrial Automation Domain: Skills Modeling and Orchestration. In *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Companion Proceedings (ICSE '20)*. Association for Computing Machinery (ACM). https://doi.org/10.1145/3377812.3381394

[9] Perzylo, A., Kessler, I., Profanter, S., & Rickert, M. (2020). Toward a knowledge-based data backbone for seamless digital engineering in smart factories. *The 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. https://doi.org/10.1109/ETFA46521.2020. 9211943

[10] Köcher, A., Hildebrandt, C., Caesar, B., Bakakeu, J., Peschke, J., Scholz, A., & Fay, A., (2020). Automating the development of machine skills and their semantic description. *The 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Vienna, Austria, 1013-1018.

https://doi.org/10.1109/ETFA46521.2020. 9211933
[11] Köcher, A., Jeleniewski, T., & Fay, A. (2021). A method to automatically generate semantic skill models from PLC code. *The 47th Annual Conference of the IEEE Industrial Electronics Society (IECON 2021)*. https://doi.org/10.1109/IECON48115.2021. 9589674
[12] Dorofeev, K. & Zoitl, A. (2018). Skill-based Engineering Approach using OPC UA Programs. *The 16th IEEE International Conference on Industrial Informatics (INDIN)*. https://doi.org/10.1109/INDIN.2018.8471978
[13] Spitzer, F., Lindorfer, R., Froschauer, R., Hofmann M., & Ikeda, M. (2020). A generic Approach for the Industrial Application of Skill-based Engineering using OPC UA. *The 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. https://doi.org/10.1109/ETFA46521.2020.9211872
[14] Profanter, S., Perzylo, A., Rickert, M., & Knoll, A. (2021). A generic plug produce system composed of semantic OPC UA skills. *IEEE Open Journal of the Industrial Electronics Society*, 128-141. https://doi.org/10.1109/OJIES.2021.3055461
[15] Jhunjhunwala, P., Atmojo, U. D., & Vyatkin, V. (2021). Applying skill-based engineering using OPC-UA in production system with a digital twin. *The 30th IEEE International Symposium on Industrial Electronics (ISIE)*. https://doi.org/10.1109/ISIE45552.2021.9576342
[16] Baumgärtel, H., Lober, A., Som, J., Völker, S., & Ollinger, L. (2023). Design of a WMS with OPC UA-based Control of an Automated Rack Storage as Study Project in Computer Science and Proof-of-Concept for Skill-Based Engineering. In *ICLF 13th Conference on Learning Factories, Reutlingen University*.

Authors' contacts:

Andreas Lober, M. Eng.
(Corresponding author)
Institute of Industrial Engineering and Supply Chain Management,
University of Applied Science Ulm,
Prittwitzstraße 10, 89075 Ulm, Germany
Andreas.Lober@thu.de

Lisa Ollinger, Prof. Dr.-Ing.
Institute of Production Technology and Material Science,
University of Applied Science Ulm,
Prittwitzstraße 10, 89075 Ulm, Germany
Lisa.Ollinger@thu.de

Sven Völker, Prof. Dr.
Institute of Industrial Engineering and Supply Chain Management,
University of Applied Science Ulm,
Prittwitzstraße 10, 89075 Ulm, Germany
Sven.Voelker@thu.de

Hartwig Baumgärtel, Prof. Dr.-Ing.
Institute of Industrial Engineering and Supply Chain Management,
University of Applied Science Ulm,
Prittwitzstraße 10, 89075 Ulm, Germany
Hartwig.Baumgaertel@thu.de

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230511184836

Preliminary communication
Received: 2023-05-11, Accepted: 2023-05-15

Location of Emergency Treatment Sites after Earthquake using Hybrid Simulation

Simona Cohen Kadosh*, Zilla Sinuany-Stern, Yuval Bitan

Abstract: A mass-casualty natural disaster such as an earthquake is a rare, surprising event that is usually characterized by chaos and a lack of information, resulting in an overload of casualties in hospitals. Thus, it is very important to refer minor and moderately-injured casualties, that are the majority of casualties and whose injuries are usually not life threatening, to ad hoc care facilities such as Emergency Treatment Sites (ETSs). These facilities support the efficient use of health resources and reduce the burden on permanent healthcare facilities. In our study, a hybrid simulation model, based on a combination of discrete events and an agent-based simulation, provides a solution to the uncertainty of positioning temporary treatment sites. The simulation methodology used compares between "rigid" and "flexible" operating concepts of ETSs (main vs. main+minor ETSs) and found the "flexible" concept to be more efficient in terms of the average walking distance and number of casualties treated in the disaster area.

Keywords: earthquake; emergency; temporary emergency facility; hybrid simulation; humanitarian logistics

1 INTRODUCTION

In an earthquake scenario, it is well known that the first hours and days are the most crucial for saving lives [1]; [2]. Following an earthquake, assuming that severe and moderate casualties are evacuated directly to hospitals, one of the main goals of the Israeli government is to prevent overburdening the hospitals during those first hours and days; hence, care of minor injuries elsewhere can leave more resources for severe and moderate casualties. Thus, the Israeli Ministry of Health has decided to establish 120 Emergency Treatment Sites (ETSs) nationwide, covering each city or locality. Their purpose is to handle minor condition casualties and up to 10% of moderate casualties that might arrive. The medical designation of these ad hoc treatment points and their equipment and personnel, which we define conceptually as "rigid" containing only main ETSs, is based on fixed locations throughout the country to treat casualties. The location of an ETS should follow safety guidelines such as being situated in an open space and in a central area. These operational decisions are imposed on Magen David Adom (MADA – the Israeli national emergency medical service) and local authorities, in cooperation with the Israel Police and the Home Front Command as part of preparedness. Unfortunately, the plan to establish and operate ETSs has not yet been implemented. MADA claims that there are disadvantages in preparing these sites ahead of time since the location of future destruction sites cannot be determined in advance.

Our study proposes a "flexible" concept with a new type of ETS that we define as a minor ETS that can be located near a destruction site. The activation of a minor ETS is subject to a specific main ETS, and its uniqueness is its closeness to a destruction site and the ability to set its location after earthquake damage occurs. Assuming that an evacuation procedure will be performed by evacuators (passers-by, volunteers, neighbors, etc.) on foot, under conditions of uncertainty, accompanied by the destruction of infrastructure and poor communications, the constraints considered include the maximum distance between a destruction site and the nearest main or minor ETS for evacuees and evacuators to walk. Early preparation for deployment of the ETS will constitute a basic plan for a tailored response and establishment of minor ETSs in real time. This paper is based on [3].

2 REVIEW OF LITERATURE

The basic plan in Israel for an earthquake event is called "The Preparation Framework" [4]. This framework states a forecast of total casualties that includes approximately 37,000 with minor injuries, 8,600 with moderate or severe injuries, and approximately 7,000 dead. Considering these figures, it is very important to refer minor and moderate-injured casualties, which include most casualties and whose injuries are usually not life threatening, to ad hoc care facilities to support more efficient use of health resources by reducing the overload on the health system.

While extensive research has been devoted to locating permanent healthcare facilities for emergency medical services, the location of temporary facilities in emergency situations has not been sufficiently investigated [5]. Planning the location of permanent facilities is done in a way in which their placement will enable them to respond to a variety of future events for all kinds of casualties and over time. In contrast, the location of temporary facilities that emerge in a case of high and immediate need for medical treatment in a specific location must be done so that their deployment will provide a quick response, with the facility placed near the event site. Patients who receive treatment at these temporary facilities can be discharged home or referred for further treatment to a permanent health facility.

An example of a temporary medical facility is the "Casualty Collection Point" (CCP), which appears in several studies [1, 6, 2]. Such a facility provides a response to various cases and different needs (not only medical) and is usually designed for a longer period (more than three days). Another example of a temporary facility are field hospitals that are established a few days after earthquakes (Armenia, Turkey, India, and Haiti) and provide care for moderate and severe casualties [7]. However, out of 150 articles about medical emergency facilities [5] only 5% of them dealt with

temporary facilities – a rate that indicates a significant lack of research in this field.

An important key feature of the temporary medical facility that affects uncertainties in the case of an earthquake is derived from the fact that the location of a planned facility and the facility itself may be damaged during the disaster event. In a survey on the location of medical facilities [8], researchers noted that there is a need to develop a more flexible model and to expand the options of locations and coverage in the solution. The distinction between a temporary facility and a permanent one will enable a suitable and quick response to the needs of the casualties and help decision and policy makers in such circumstances.

A recent study [9] used a two-stage stochastic mathematical model in which recommended locations of temporary medical facilities after an earthquake (a field hospital) were examined. Cohen-Kadosh [10] presented a deterministic mathematical model for locating main and minor ETSs with two objectives: minimizing the overall distance of evacuations from all the destruction sites and maximizing the number of casualties that could receive treatment in the main and minor ETSs.

The knowledge gap we found in the literature regarding using and locating temporary facilities under an assumption of casualty evacuation by relatives or volunteers, while doing so on foot and caring mainly for minor condition casualties needed to be investigated and satisfied.

Our use of simulation aimed to provide an answer to the most known stochastic scenario such as an earthquake. Using simulation methodology allows for the creation of large emergency scenarios with variant damage combinations of random and other characteristics involved in the event.

Fikar at al. [11] presented a study dealing with disaster situations based on multi-agent simulation (agent- based simulation), combined with optimization to determine the location of logistical assistance centers after a disaster (for supplying food, water, medicine, and emergency equipment). Another study [12] examined a robust optimal simulation model to locate aid and supply centers before a disaster such as an earthquake. The study used simulation and established assumptions including known parameters that indicated the extent of the relationship between the various infrastructure types (for example between an electrical system and a gas system or drinking water infrastructure) and the damage following a variety of disaster scenarios.

Brailsford et al. [13] indicated a new trend in combining several types of simulations: discrete event simulation, agent-based simulation, and dynamic simulation, thus, offering a solution for complex situations. About 22% (out of 139) of the articles in that survey dealt with health services. Therefore, in our research, we decided to employ a hybrid simulation combining a discrete-event simulation, which describes the chain of events in the formation of destruction sites as a result of an earthquake and the appearance of evacuees and evacuators, with an agent-based simulation representing what actually happens in the ETSs and the destruction sites, i.e., dealing with the movement and transfer of casualties from destruction sites to the emergency facilities.

3 RESEARCH OBJECTIVE

Considering the above, we examined the ability of positioning ETSs (main and minor) in the case of an earthquake disaster and compared between the two operating concepts – the rigid concept (only main ETSs) versus the flexible concept (main and minor ETSs) – under uncertainty conditions. The hybrid simulation examining different disaster scenarios can support authorities' decisions as to where to position ETSs to result in minimal evacuee transfer distances and a maximal percentage of casualties that could receive medical treatment.

4 HYBRID SIMULATION MODELS FOR AN ETS

4.1 Operating Mode

For a test case, we used a specific city with several destruction sites (7). In the rigid mode, before an earthquake event, candidate main ETSs (3) are pre-determined. In the flexible mode of operation, there are also minor ETSs (5 candidate locations) considered shortly after the earthquake event. The solution for the flexible mode, in Figure 1, shown by arrows the selected ETSs (2 main and 3 minor), those directed by arrows, from the destruction sites as casualties which arrive to the various ETSs. The arrows represent the movement of evacuees carried/accompanied by evacuators from a destruction site to an ETS (main and minor).

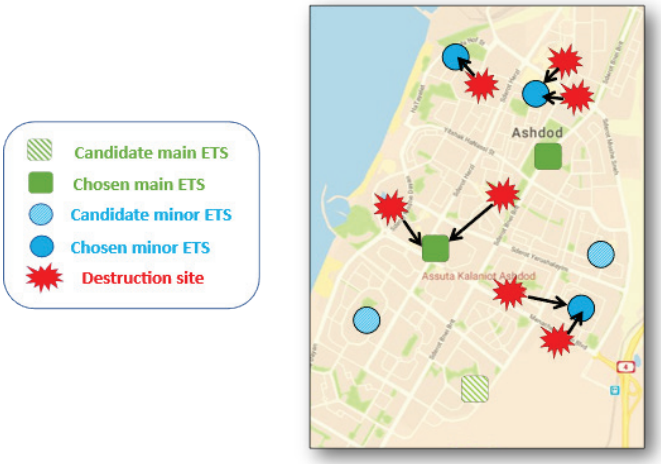


Figure 1 Flexible & Rigid operating concept

4.2 Stochastic Simulations for Casualty Evacuation in an Earthquake

The selected simulation for our study is a hybrid of a discrete simulation, which creates destruction sites, casualties, and evacuators; and an agent-based simulation, which expresses physical evacuation of casualties from destruction sites to ETSs and returning to the evacuation site. In each simulation run, there are entities and resources which have properties that are determined randomly or deterministically.

4.3 Simulation: General Description

We describe here the simulation of evacuation for comparing between two operational modes:

- The rigid concept with one ETS level: only main ETSs – Simulation model A
- The flexible concept with two ETS levels: minor and main ETSs – Simulation model B.

Comparison between the two concepts was done for various levels of each parameter, identifying earthquake scenarios (such as the number of casualties, number of ETSs and their locations) and on the stochastic behavior of the entities or resources (such as the rate of casualty appearances and evacuators, matching between casualties and evacuators, length of walking, road condition and its effect on the length of walking, etc.). For an earthquake event, we determined the locations of the two ETS types: main and minor.

The simulation model was built by using an RStudio interface, combined with a Simmer package [3]. The simulation was defined as a system with no stops, ending either after 48 hours or when no further resources remained. The results were tested by two objective measures:

1) The average distance per casualty from the destruction site to the relevant ETS (as a proxy for evacuation time) was to be minimized.
2) The proportion of casualties who received medical treatment from the total number of casualties was to be maximized.

The chain of events in each simulation run began with an earthquake event, followed by a random number of destruction sites scattered in a pre-defined area. At each destruction site, the simulation randomly created casualties and evacuators. The connection between destruction sites and the various ETSs were limited by the maximal distance allowed between them, and the capacity limitation (number of casualties) in the ETS. The movement of casualties and the ETSs' operation was performed according to the specific concept (rigid or flexible) and ended at the specified time defined for the simulation, either when all casualties from the sites received medical treatment or when the ETS capacity was reached – whichever occurred first.

4.4 System Description

We describe here the system components, model assumptions, and main connections between the entities and resources.

4.4.1 System Components

- A main ETS is a resource providing medical treatment for a given capacity of casualties (most often 250). Its location was taken from several possible given locations.
- A minor ETS is as above, but with a much smaller capacity (most often 50). It is connected to a specific main ETS, as it receives its medical supply only from one main ETS which is connected to it as its source. There was a minimal distance limitation between the minor ETS and the main ETS connected to it, given that the main ETS had enough capacity for the minor ETS.
- An evacuator is a resource representing those persons who establish the minor ETS (in the flexible case) and evacuate the casualties to it. Such a resource is created by the main ETS connected to it. A main ETS served each minor ETS connected to it directly.
- An evacuee is an entity representing a casualty from the disaster site who is waiting to be taken by evacuators to the ETS associated to it (minor or main ETS). These entities were created at each disaster site separately. Each evacuee had a property determining the number of evacuators needed for their evacuation (1–4).

4.4.2 Basic Assumptions of the Simulation Model

- The simulation was modeled without stops, ending at most after 48 hours.
- The time units were in minutes.
- The main ETS was created at the beginning of the simulation.
- The simulation will be repeated under various sets of parameters as shown later.
- A case is a container with medical supplies (for one evacuee/the capacity of a minor ETS usually is 50 casualties) brought from the main ETS connected to the minor ETS.
- The time duration for loading cases in the main ETS and unloading them in the minor ETS was negligible.

4.4.3 Main Stochastic Parameters in Simulation A and B

Some of our assumptions are based on the literature, whereas others are based on interviews with experts during the study or on orders and procedures of the relevant authorities. The parameters related to the walking time are based on experiments performed by the authors.

- The number of destruction sites was selected randomly between 4–10 uniformly. Similarly, the coordinates of their locations were drawn randomly.
- A destruction site was considered randomly from a set of unconnected sites, and the closest available ETS which met the capacity and distance requirements was selected. If no minor ETS was found, then the simulation identified main possible ETSs within the distance limitations, and with sufficient available capacity in relation to the number of evacuees predicted from the site. A minor ETS was connected to the closest main ETS with available capacity.
- The number of casualties at each destruction site was determined randomly according to the number of destruction sites drawn. The sum of all casualties at all destruction sites was determined in advance as the *TC* (total casualties) given by forecast.
- The arrival time between consecutive evacuees was distributed exponentially, where TC_K was the number of evacuees at destruction site.

- The number of evacuators had a Poison distribution with a rate of 10 per minute.
- Evacuator rests-The evacuators rested every 3 hours [3].
- Evacuee properties – Every evacuee had a property defining the number of evacuators needed, so that 10% of the evacuees needed one evacuator (the evacuees could walk unassisted), evacuees under the age of 14 (evacuees who needed to be carried) who needed two evacuators, and the remainder (older evacuees who needed to be carried) who needed four evacuators (based on information from evacuation authorities).
- Waiting for evacuees in line – The evacuees waited for the number of evacuators they needed. Priority was given to those who needed four evacuators, then two evacuators, and last one evacuator.
- Walking time – The duration of evacuators' walking was based on an experiment that we performed [3] for one evacuee carrying weight as the basic walking time, whereas walking back without weight was adjusted by factor of 0.95.
- Accounting for various terrains (easy, medium, or difficult) – The type of terrain affects the evacuation time. The factor used had the same probability for each type of terrain (1/3). The easy terrain factor was 1, the medium terrain factor was 1.2, and difficult terrain factor was 1.4, based on the method used by the Israeli army.
- The number of evacuators affects the walking duration – The need for team synchronization among the evacuators increases the evacuation time; thus, the larger the team, the longer the evacuation time of the evacuee. Therefore, a single evacuator had a factor 1, two evacuators had a factor of 1.2, and four evacuators had a factor of 1.3.
- The walking duration between the destruction site to ETS – was determined by multiplying the above factors:
- The walking speed from our experiment [3]
- The distance between the destruction site and the ETS
- The terrain factor (easy, medium, difficult)
- Evacuator team size (single, two, or four evacuators).

The goal was to minimize the average distance of casualties, and to maximize the proportion of treated casualties.

4.5 Key Parameters Characterizing an Earthquake Event

For the typical example area in Fig. 1, we examined the sensitivity of the simulation models for five defined key parameters that characterize an earthquake event as follows:

- *TC* – The total number of casualties, which received the following values: 500, 600, 750, 900.
- *RS* – The maximum possible distance from the destruction site to the main ETS ranging from 1,000 to 4,000 meters.
- *RM* – Possible distance from the destruction site to the minor ETS, which received the values: 500, 1,000, 1,500 meters.
- *MS* – Maximal patient capacity in the main ETS, which received the values: 250, 300, 450.

MM – Maximal patient capacity in the minor ETS, 50 or 100.

In addition, there was a sample set of possible locations for three main ETSs and five minor ETSs.

The simulation models were tested according to different combinations of the above sensitivity parameters, where 100 different earthquake scenarios were simulated for each combination of parameters, in which each created a variable number and layout of destruction sites; namely, 4,800 row results were calculated for the rigid perception, and 10,800 row results were calculated for the flexible perception.

To determine which layout performed better and which model we recommend, the results were examined by comparing the findings of the two measures we defined in Section 4.3:

1) The average distance per casualty, which represents the distance that a patient/team would walk from the destruction site to the relevant ETS, as an estimate (proxy) of the evacuation duration; and
2) The proportion of treated casualties in the ETS compared to total casualties in the area.

5 RESULTS

Comparing the simulation of the rigid concept to the flexible concept, we found that for the average walking distance per casualty index, the flexible concept (Model B) was better than the rigid concept (Model A). However, for the average percentage of treated casualty measure, we found that Model A outperformed Model B. Since the simulation was run 100 times for each combination of parameters (as listed in Section 4.6), we were able to test the statistical significance of the results via t-tests and found that most of the results had a 95% significance.

Fig. 2 summarizes the results of the two measures for all the combinations of parameters multi-dimensionally. The left column ($RM = 0$) represents the rigid concept (Model A), and the rest of the *RM* columns represent the flexible concept (Model B). The four dots in each rectangle in each *RM* column provide the results of the two measures as a function of *TC* (in the same order as they appear on the right side of the figure).

In our sensitivity analysis of the parameters, we found that for a possible maximum distance *RS* of 2,000 meters from a main ETS to the destruction site and maximum distance RM 1,000 from a minor ETS to the destruction site, an improvement of the two measures was obtained. Increasing to an *RS* of 3,000, we observed a stabilization in the index values. Along with an improvement in the average distance index, we observed a decrease in the proportion of treated patients. The relative improvement in the distance index (in percent) was usually higher than the relative improvement in the index of treated casualties.

In testing the effect of the ETS's capacity (*TC*) on the results, in comparing the two concepts, we found that as the capacity of the main ETS increased alongside that of the minor ETS, there was an improvement in the average distance measure per evacuee (shorter average distance)

compared to a slight decrease in the proportion of treated patients (deterioration in the index) based on the range capacity between the main ETS and the minor ETS. Examining the effect of the total number of casualties on the indicators based on the layout of the ETSs, it can be said that in the rigid concept, as the number of casualties increases, the location of the main ETS has a negative effect. Thus, the main ETS was located, the casualties walked more, while on the other hand, fewer casualties were treated. Checking the flexible concept, we found that as the number of total casualties increased, adding the minor ETSs resulted in smaller average walking distance (improvement of the index), while again decreasing the ratio of treated casualties.

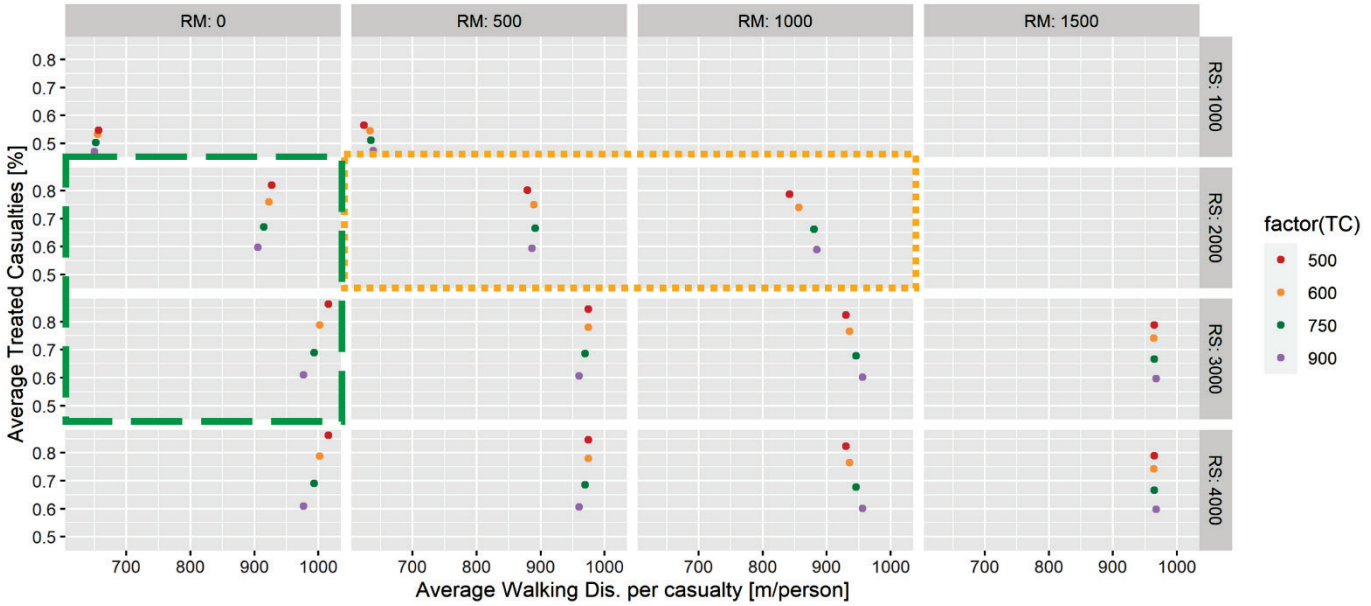


Figure 2 Average walking distance and percentage of treated casualties by various parameters (*RM* from 0-1500, *RS* from 1000-4000 and *TC* from 500-900)

To sum up, it is evident that, for the most part, for the patient proportion index, Simulation model A according to the rigid concept is clearly the preferred one, whereas for the average distance index, the preferred alternative is Simulation model B, according to the flexible concept. It can be said that there is interchangeability between the two indices.

We see, for example, that the capacity of the ETS of both types, and their relative location to the destruction site, affects the ability to determine the preferred alternative. Therefore, a decision must be made on the weight of each index, and this decision remains an open question for policy makers.

6 CONCLUSIONS

This simulation was formulated as a hybrid simulation, incorporating many stochastic parameters that characterize an earthquake event. By averaging multiple cases of earthquake scenarios, we conclude that the flexible concept yields better results regarding the average walking distance index. The analysis in the study shows that there is substitutability between the two indices – the average distance and the proportion of treated casualties – and, therefore, a decision must be made on the weight of each index, and this remains an open question for policy makers.
We emphasize that the proposed flexible model should be adapted, and it can be utilized both as part of the preparation phase before an earthquake, and as an autonomous decision support system that supports local authority real-time response to a disaster. The proposal to develop an applicable decision support tool was recently approved by the Israeli Ministry of Science and Technology, and a grant was awarded.

Future research should analyze larger areas of destruction with more main and minor ETSs. Another direction is to consider better rules to improve the percentage of treated casualties using the flexible concept, such as by sending untreated casualties from a minor ETS (due to lack of capacity) to the closest main ETS. Moreover, when a minor ETS finishes treating all its casualties and has leftover supplies, these should be sent to the closest main ETS. Such policies have the potential to reduce the percentage of untreated casualties.

7 REFERENCES

[1] Arl, C., Chultz, H. S., Risti, K., Oenig, L. K., & Oji, R. K. N. (1996). A medical disaster response to reduce immediate mortality after an earthquake. *New Engl J Med, 334*(7), 438-444. https://doi.org/10.1056/NEJM199602153340706

[2] Fawcett, W. & Oliveira, C. S. (2000). Casualty Treatment after Earthquake Disasters: Development of a Regional Simulation Model. *Disasters, 24*(3), 271-287. https://doi.org/10.1111/1467-7717.00148

[3] Cohen Kadosh, S. (2021). Preparedness for an Earthquake Disaster: Modeling Optimized Deployment of Emergency Treatment Sites. *PhD dissertation*, Ben-Gurion University, Beer Sheva, Israel (in Hebrew).

[4] Shapira, A. (2011). Framework for preparing for devastating earthquakes in Israel. (in Hebrew)
[5] Ahmadi-Javid, A., Seyedi, P., & Syam, S. S. (2017). A survey of healthcare facility location. *Comput Oper Res, 79*, 223-263. https://doi.org/10.1016/j.cor.2016.05.018
[6] Drezner, T. (2004). Location of casualty collection points. *Environ Plan C Gov Policy, 22*(6), 899-912. https://doi.org/10.1068/c13r
[7] Bar-On, E., Abargel, A., Peleg, K., & Kreiss, Y. (2016). Coping with the Challenges of Early Disaster Response: 24 Years of Field Hospital Experience after Earthquakes. *Disaster Medicine and Public Health Preparedness, 7*(5), 491-498 https://doi.org/10.1017/dmp.2013.94
[8] Peng, Q. & Afshari, H. (2014). Challenges and Solutions for Location of Healthcare Facilities. https://doi.org/10.4172/2169-0316.1000127
[9] Oksuz, M. K. & Satoglu, S. I. (2020). A two-stage stochastic model for location planning of temporary medical centers for disaster response. *Int J Disaster Risk Reduct, 44*. https://doi.org/10.1016/j.ijdrr.2019.101426
[10] Cohen Kadosh, S., Sinuany-Stern, Z., & Bitan, Y. (2018). Preparedness for earthquake disaster: optimal deployment of emergency treatment sites. *The 14th International Conference on Industrial Logistics (ICIL 2018)*, 55-62.
[11] Fikar, C., Hirsch, P., & Nolz, P. C. (2018). Agent-based simulation optimization for dynamic disaster relief distribution. *Cent Eur J Oper Res, 26*(2), 423-442. https://doi.org/10.1007/s10100-017-0518-3
[12] Ghasemi, P. & Khalili-Damghani, K. (2021). A robust simulation-optimization approach for pre-disaster multi-period location–allocation–inventory planning. *Math Comput Simul, 179*, 69-95. https://doi.org/10.1016/j.matcom.2020.07.022
[13] Brailsford, S. C., Eldabi, T., Kunc, M., Mustafee, N., & Osorio, A. F. (2019). Hybrid simulation modelling in operational research: A state-of-the-art review. *Eur J Oper Res, 278*(3), 721-737. https://doi.org/10.1016/j.ejor.2018.10.025

Authors' contacts:

Simona Cohen Kadosh, PhD, Lecturer
(Corresponding author)
(1) Sapir Academic College, Logistics Department,
Mobile Post Hof Ashkelon, 7915600, Israel
(2) Ben-Gurion University of the Negev,
Department of Industrial Engineering & Management,
Ben-Gurion Avenue 1, POB 653, Beer-Sheva, 8410501, Israel
cohensim@mail.sapir.ac.il

Zilla Sinuany-Stern, PhD, Full Professor
Ben-Gurion University of the Negev,
Department of Industrial Engineering & Management,
Ben-Gurion Avenue 1, POB 653, Beer-Sheva, 8410501, Israel
zilla@bgu.ac.il

Yuval Bitan, PhD, Associate Professor
Ben-Gurion University of the Negev,
Department of Health Policy & Management,
Ben-Gurion Avenue 1, POB 653, Beer-Sheva, 8410501, Israel
ybitan@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230511175500

Preliminary communication
Received: 2023-05-11, Accepted: 2023-05-15

A Dynamic Systems Model for an Economic Evaluation of Sales Forecasting Methods

Lara Kuhlmann*, Markus Pauly

Abstract: Sales forecasts are essential for a smooth workflow and cost optimization. Usually, they are assessed using statistical error measures, which might be misleading in a business context. This paper proposes a new dynamic systems model for an economic evaluation of sales forecasts. The model describes the development of the inventory level over time and derives the resulting overstock and shortage costs. It is tested on roughly 3,000 real-world time series and compared with the commonly used approach based on statistical measures. The experiments show that different statistical measures have no coherent evaluation, making their usage even less suitable for a practical economic application.

Keywords: forecasting; inventory management; sales forecast; supply chain analytics; time series

1 INTRODUCTION

Working with precise sales forecasts is crucial for the supply chains of production companies or retailers [1]. It enables a smooth workflow and reduces waste along the value chain [2]. The longer the lead times of products are, the more important it is to accurately plan ahead [3]. Due to current crises, the lead times of most products have increased substantially [4], amplifying the relevance of reliable forecasts. The challenge is to find the most suitable forecasting method among plenty of existing ones, each with its own benefits [5]. Usually, a forecasting method is chosen based on statistical accuracy measures, also called error metrics.

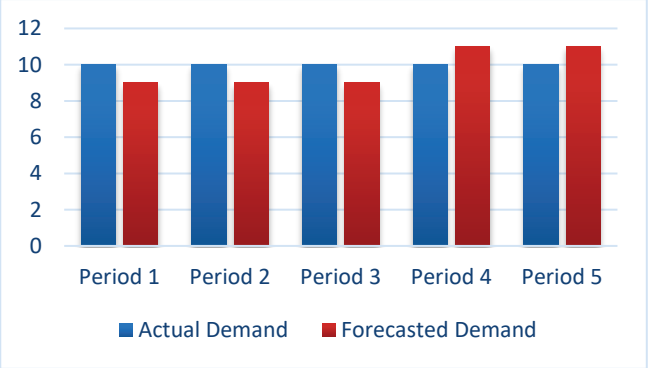


Figure 1 Forecasting with Model 1 in the toy example

However, solely choosing a sales forecast method based on statistical accuracy measures can have some disadvantages, as the following toy example illustrates: We assume a given demand for a product and we want to compare two forecasting models that forecast the demand for five periods. The forecast of Model 1 is one unit below the actual demand for the first three periods and one unit above the actual demand for the last two periods (see Fig. 1). The forecast of Model 2 is alternating one unit above and then one unit below the actual demand (see Fig. 2). Both models had the same absolute deviations in every period. Thus, statistical measures as the Root Mean Square Error (RMSE) or the Mean Average Percentage Error (MAPE) are the same for both models, suggesting an equally good performance.

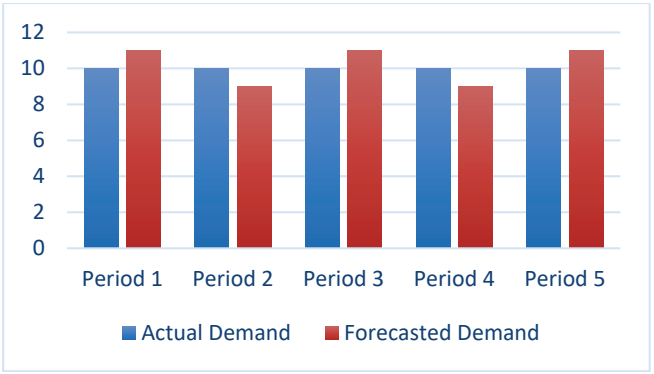


Figure 2 Forecasting with Model 2 in the toy example

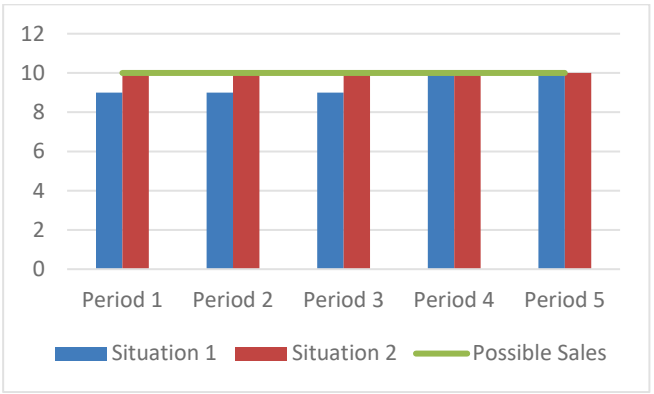


Figure 3 Sales based on the chosen forecasting Models 1 and 2 in the toy example

Nevertheless, from a business perspective, Model 2 outperforms Model 1 if the goods are non-perishable and their value does not decrease over time. If the company had acted upon Model 1, it would not have been able to meet the customers' demand in the first three periods (see Fig. 3). If the company had relied upon Model 2, it would have satisfied the customers' demand in every period. Thus, considering the costs of product shortage (which would have been higher for Model 1) and the costs of overstock (which would have been

equal) over time seems to be a more suitable way for evaluating sales forecasts.

Some papers already considered costs in the context of forecasting, but never the development of stock levels over time. In order to fill this gap, we introduce an intuitive dynamic systems model for an economic evaluation of sales forecasting methods, considering overstock and shortage costs that derive from the development of inventory levels. The new method is tested on a real-world dataset of about 3,000 product sales time series of a German raw materials wholesaler. Sales forecasts are created and evaluated using the dynamic systems model but also statistical measures. The evaluations are compared with regard to the similarity and preference for a certain forecasting method.

2 RELATED WORK

2.1 Sales Forecasting Methods

There are many different methods for sales forecasting [5]. They rely on time series analysis and forecast future demand based on historic sales [6]. External data can be integrated into forecasts to improve them [7]. The forecasting models are usually univariate point forecasts. The models can either be statistical time series models or machine learning models. The statistical models are easy to implement and intuitively explainable. Among what [8] considers classical sales forecasting methods are ARIMA (Autoregressive Moving Average) and Holt-Winters. As the name suggests, the ARIMA model combines an autoregressive component, that links past and present values in a similar way as autocorrelation is computed, and a moving average component [9]. Holt-Winters is a seasonal smoothing method that can capture both trends and seasonal behavior within a time series [10, 11]. Recently, a lot of research has been done on forecasting with machine learning methods [12]. They can provide forecasts with higher accuracy [13], but require more run time and are not intuitively explainable. However, as the no-free lunch theorem suggests, there is no single best method [14]. Depending on the characteristics of a dataset, one method achieves more precise forecasts than another method. Subsequently, it is advisable to test several forecasting algorithms and choose the most suitable one.

2.2 Statistical Sales Forecast Evaluation

The performance of sales forecasts can be evaluated by applying point forecast error metrics to a test dataset or multiple test datasets in case of (rolling) time series cross-validation. Ref. [15] conducted a survey on forecast error measures and found that 23 different measures are in use (see Tab. 1). A list providing the full names of the error measures, whose acronyms are given in the table, can be found in the appendix.

In the context of sales forecast, the most commonly applied error measures are the RMSE [13, 8, 16, 17, 18], the MAPE [17, 18, 19] and MAE (Mean Absolute Error) [16, 18, 20]. The MAPE is easy to interpret but cannot be computed if the time series contains zeros [21]. As a remedy, its symmetric version sMAPE (Symmetric Mean Absolute Percentage Error) can be applied. It has the further advantage of penalizing under-forecasts more severely than over-forecasts. As the time series for our experimental evaluation (see Section 4) contain zeros, we focus on the three error measures RMSE, MAE and sMAPE. The error measures are computed based on the actual sales y_t and the forecasted sales $\hat{y}_t$ at time t [15]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2} \quad (1)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_t - \hat{y}_t)| \quad (2)$$

$$sMAPE = \frac{1}{n}\sum_{i=1}^{n} 200 * \left|\frac{(y_t - \hat{y}_t)}{(y_t + \hat{y}_t)}\right|. \quad (3)$$

Table 1 Forecasting error measures based on [15]

Category	Error measures
Absolute forecasting errors	MAE, MdAE, MSE, RMSE
Measures based on percentage errors	MAPE, MdAPE, RMSPE, RMdSPE
Symmetric errors	sMAPE, sMdAPE, msMAPE
Measures based on relative errors	MRAE, MdRAE,
Scaled errors	MASE, RMSSE
Relative measures	RMAE, RRMSE, LMR, PB(MAE)
Other error measures	mRMSE, inRSE, Std_AE, Std_APE

Despite the challenge of choosing an error measure, the problem arises that none of them can provide us with a unique best forecasting method. What is an excellent RMSE value for one forecasting method in one dataset, might be a mediocre value for another dataset. Depending on the demand stability, some products are easier to forecast than others. Thus, to achieve an objective evaluation of a forecasting method, it is recommendable to create a baseline forecast, to which other more advanced methods can be compared to in the process of benchmarking [5]. A very common baseline method is naïve forecasting, which simply assumes the future sales to be equal to the most recently observed sales [22].

2.3 Inventory-Related Costs

Holding inventory entails different kinds of costs. In order to minimize the overall costs, the different kinds need to be balanced. Literature on reducing inventory-related costs focuses on capital commitment costs, order costs, and shortage costs. Sometimes, costs for the inventory control system are also considered [23]. The capital commitment costs are usually priced with the weighted average cost of capital (WACC) [24]. Storage costs can usually be disregarded because most companies own their warehouses, making the storage costs fixed [23]. Only if companies are charged fees based on a variable level of inventory the companies should consider these fees when deciding on inventory order politics [24]. Shortage costs measure the costs that occur from not being able to sell a product. These include the lost profit margin and also reputational damage and a decrease in customers' loyalty [25].

2.4 Cost-Considering Sales Forecast Evaluation

The evaluation of sales forecasts with costs has only been done rarely. The authors of [26] focused on reducing overstock costs with the help of statistical forecasting methods but disregarded shortage costs. Ref. [27] compared the sales forecasts of five neural networks in terms of supply chain-related costs. They derived the costs by looking at the deviations of the forecast and the actual demand for every period individually, not considering the development of inventory levels.

However, the question of how many goods to order to maximize the profit has been investigated a lot. The so-called newsvendor problem is a popular problem in logistics and many approaches to solving it have been published [28]. It also considers shortage costs and tries to optimize the order quantity [25] but it only regards one period. This is why it is not applicable in this paper.

3 THE DYNAMIC SYSTEMS MODEL

A dynamic systems model is an analytical model that describes the changes in a system over time [29]. The proposed dynamic systems model aims to describe the changes in inventory level and the resulting costs over time. To use the model as an economic evaluation for sales forecasting methods, some assumptions had to be made:

- Goods are non-perishable and their value does not decrease over time
- Storage costs are fixed costs
- Goods are ordered at the beginning of each period, order costs do not change
- The ordered goods are delivered at the beginning of period $t + LT$ (lead time)
- The lead time per product is fixed
- The costs for holding a safety stock are calculated into the product price, thus only costs for an inventory level above the safety stock are considered overstock costs.

Ref. [30] reviewed and compared deterministic and statistical methods for the calculation of a safety stock. They found the statistical methods to be more accurate. Thus, we rely on the statistical method, for which the aforementioned assumptions hold, e.g. that the lead time is fixed. The dynamic systems model determines the safety stock s based on a service factor Z, which depends on the desired service level, the lead time l, and the standard deviation of the product's demand σ_y:

$$s = Z * \sqrt{l} * \sigma_y. \tag{4}$$

In the first step, the dynamic systems model describes the inventory level development, later the resulting costs. For every period t, the model calculates the inventory level at the end of that period (I_{E_t}) by subtracting the actual sales (y_t) from the inventory level at the beginning of that period (I_{B_t}). Naturally, an inventory level can never be negative. Thus, the formula to calculate I_{E_t} is given by

$$I_{E_t} = \max(I_{B_t} - y_t, 0). \tag{5}$$

Here, I_{B_t} depends on the inventory level at the end of the prior period ($I_{E_{t-1}}$) and the goods that had been ordered and are delivered at the beginning of each period (D_t) via

$$I_{B_t} = I_{E_{t-1}} + D_t. \tag{6}$$

The number of goods delivered equals the number of goods that have been ordered in a previous period. The time gap between order and delivery is determined by the lead time l of a product:

$$D_t = O_{t-l}. \tag{7}$$

The number of goods ordered depends on the predicted sales for the period in which the goods will be delivered ($\hat{y}_{t+l}$), the predicted sales for the period in which the goods are ordered ($\hat{y}_t$), the safety stock s and the inventory level at the moment of order (I_{B_t}):

$$O_t = \max(\hat{y}_{t+l} + s + \hat{y}_t - I_{B_t}, 0). \tag{8}$$

The derivation of this formula can be found in the Appendix. As the dynamic systems model aims to describe the inventory level development over a certain period of time, periods prior to $t = 0$ need to be modeled to calculate the number of goods that are delivered in $t = 0$. For the periods prior to $t = 0$ the following assumption holds:

$$D_{0-l} = \hat{y}_{t-l}. \tag{9}$$

Moreover, due to the mutual dependency of I_{B_t} and I_{E_t}, an assumption needs to be made about the initial inventory level:

$$I_{1_{0-l}} = s + \hat{y}_{t-l}. \tag{10}$$

Tab. 2 displays another toy example of how the development of the inventory level is calculated. The initial inventory level $I_{B_{-2}}$ is the sum of the predicted sales for the period $\hat{y}_{-2}$, which are 51, and the safety stock s, which is 634. The number of goods ordered in $t = -2$ are calculated as sum of the predicted sales for that period, the predicted sales of the period in which the goods are meant to arrive ($t = 0$), and the safety stock. From this sum, only the inventory level at the beginning of the regarded period needs to be subtracted. Thus, $O_{-2} = 51 + 50 + 634 - 685 = 50$. $I_{E_{-2}}$ results from subtracting 40 from 685.

The example shows a sharp increase in sales in period $t = 1$. This leads to a stockout at the end of that period. However, due to the application of the naïve forecasting method, the increase in sales in $t = 1$ also leads to very large order and to nearly 7,000 goods being delivered in period $t = 4$.

Table 2 Second toy example: Calculation of inventory level over time with *l* = 2, *s* = 634, forecasting method: Naive

t	D_t	I_{B_t}	y_t	$\hat{y}_t$	O_t	I_{E_t}
–2	51	685	40	51	50	645
–1	263	908	300	263	29	608
0	50	658	50	50	326	608
1	29	637	6,091	40	87	0
2	326	326	50	300	6,699	276
3	87	363	0	50	371	363
4	6,699	7,062	0	6,091	0	7,062
5	371	7,433	565	50	0	6,868

In the second step, the dynamic systems model calculates overstock and shortage costs based on the previously calculated inventory level development. Overstock costs C_{O_t} per period t are calculated by multiplying the difference of the average inventory level of that period and the safety stock with an overstock cost rate w, most likely the WACC. If the average inventory level is below the safety stock, which implies the lack overstock, the product becomes negative. Therefore, the product is set to be zero at minimal:

$$C_{O_t} = \max\left(\left(\frac{I_{E_t}+I_{B_t}}{2} - s\right) * w\ , 0\right). \tag{11}$$

The shortage costs C_{S_t} per period t are calculated by multiplying a shortage cost rate, e.g. the product's profit margin with the difference of the actual sales that could have taken place y_t and the inventory level at the beginning of that period I_{B_t}. Again, the product is limited to zero as there are no shortage costs, if the inventory level is higher than the sales:

$$C_{S_t} = \max\left(\left(y_t - I_{B_t}\right) * m, 0\right). \tag{12}$$

Finally, the overall costs C_t per period t are calculated as sum of overstock and shortage costs:

$$C_t = C_{O_t} + C_{S_t}. \tag{13}$$

Table 3 Continuation of the second toy example: Calculation of overstock and shortage costs, *s* = 634, *w* = 0.5% and *m* = 6%

t	I_{B_t}	y_t	I_{E_t}	C_{O_t}	C_{S_t}	C_t
–2	685	40	645	-	-	-
–1	908	300	608	-	-	-
0	658	50	608	0	0	0
1	637	6,091	0	0	327.24	327.24
2	326	50	276	0	0	0
3	363	0	363	0	0	0
4	7,062	0	7,062	32.14	0	32.14
5	7,433	565	6,868	32.58	0	32.58

Tab. 3 displays the continuation of the second toy example from Tab. 2. Based on the inventory levels at the beginning and end of each period, I_{B_t} and I_{E_t}, the overstock and shortage costs are calculated. The enormous increase in sales in period *t* = 1, which could have been observed in Tab. 2, leads to shortage costs of 327.24. The delivery of nearly 7,000 goods in period *t* = 4 however, leads to overstock costs in period *t* = 4 and the following periods.

4 EXPERIMENTAL SET-UP

The dynamic systems model is empirically tested using data from a German raw materials wholesaler. The company provided documentation of its sales from 1st September 2016 until 31st July 2022. Every product sale can be assigned to one of five material divisions. For each division, estimated profit margins were given, which serve as shortage costs. The WACC was used to measure the overstock costs. Because some products were sold very infrequently, filters were applied to the data. [31, 32] found that for time series forecasts on a monthly basis, at least 24 observations as training data are required in order to create reasonable forecasts. Rolling time series cross-validation was applied and for all time series, the test data were the sales from August 2021 until July 2022 (the last year of observations). The training data were accordingly the sales prior to the test data, the exact time period also depended on the lead time of the products.

Time series were included if they met the following criteria:

- Date of last sale was between May and July 2022 (to exclude products that are not sold anymore)
- Average sale frequency ≥ twice per month, meaning that the number of sales has to be higher than twice the number of months between the first and last sale (to exclude products with highly intermittent demand)
- The time series contained at least 36 observations plus the amount of lead time in observations (thus, at least 24 observations could be used for training the forecasting model).

These requirements limited the dataset to time series from 2,911 different products. Tab. 4 displays how many product time series from which division and with which lead time were included.

Table 4 Number of time series that met the inclusion criteria per product division and lead time

Division	Lead time in months				
	1	2	3	4	5
1	483	352	0	29	4
2	171	246	39	43	22
3	598	457	9	1	67
4	54	243	91	2	0
Sum	1,313	1,302	139	75	93

For these products, one-step rolling forecasts were created using the statistical forecasting methods naïve forecasting, ARIMA, and Holt-Winters' additive approach. They are considered classical sales forecasting methods and are easy to implement. As we propose a new dynamic systems model and focus on the comparison of forecast evaluation measures and not forecasting methods, we only consider statistical time series approaches. If forecasts happened to be negative, they were adjusted to zero. All forecasts were rounded to integers due to the application context. The forecasts were evaluated using the dynamic systems model and the statistical measures RMSE, MAE, and sMAPE. The evaluation with the dynamic systems model

was performed assuming three different service levels and thus three different safety stocks per product. Considered service levels were 90%, 95%, and 99%, corresponding to service factors Z equal to 1.3, 1.6, and 2.3 [30].

5 RESULTS

Each evaluation metric assessed one forecasting method to be most suitable for one product. Tab. 5 displays how often these assessments were coherent and how often different evaluation measures came to different results. The upper part of the table displays the assessments for all 2,911 time series. The middle and lower part show the results only for the divisions with either the lowest or highest profit margins.

Table 5 Percentage of time series for which the measures assessed the same forecasting method to be best

	Statistical evaluation			Economic evaluation with different service levels		
	RMSE	MAE	sMAPE	90%	95%	99%
For all-time series						
RMSE	100	75.37	59.64	55.82	55.14	54.14
MAE	75.37	100	67.06	55.62	55.93	56.51
sMAPE	59.64	67.06	100	43.15	42.39	41.39
90%	55.82	55.62	43.15	100	92.75	83.79
95%	55.14	55.93	42.39	92.75	100	89.42
99%	54.14	56.51	41.39	83.79	89.42	100
For the division with the lowest profit margin						
RMSE	100	73.51	59.31	55.09	54.51	53.93
MAE	73.51	100	67.37	57.20	58.15	58.35
sMAPE	59.31	67.37	100	43.19	43.95	42.99
90%	55.09	57.20	43.19	100	94.43	89.44
95%	54.51	58.15	43.95	94.43	100	92.90
99%	53.93	58.35	42.99	89.44	92.90	100
For the division with the highest profit margin						
RMSE	100	75.13	62.05	51.79	50.51	54.62
MAE	75.13	100	71.03	53.33	52.82	56.67
sMAPE	62.05	71.03	100	43.33	43.08	45.13
90%	51.79	53.33	43.33	100	91.54	77.44
95%	50.51	52.82	43.08	91.54	100	84.36
99%	54.62	56.67	45.13	77.44	84.36	100

The highest coherence among the statistical measures can be found between the RMSE and the MAE. This is plausible because both measures are absolute forecasting errors. However, even these two similar measures found different best methods for a quarter of the time series. The coherence among the dynamic systems model's evaluations with different safety stock is higher, up to 92.75%. This might imply that the safety stock level does not have a huge impact on the choice of the most suitable forecasting method. When comparing the dynamic systems model's evaluation with those of the statistical measures, it strikes that they rarely come to the same conclusions. The coherence between the RMSE and MAE and the dynamic systems model's evaluation ranges between 54% and 57%. However, for the sMAPE, it is even lower with values between 41% and 44%.

The lower part of Tab. 5 displays the coherence for the divisions with the lowest and the highest profit margin/ shortage costs. The comparison shows that the higher the profit margin, the more the results diverge. For the division with the lowest margin, there is a 92.90% coherence between the dynamic systems model with a service level of 95% and 99%. In the division with the highest profit margin, this value is 84.35%, considerably lower. Also the coherence with the statistical measures decreases with increasing profit margins.

Tab. 6 displays the average of the ranks the evaluation measures assigned to the forecasting methods. The lower the rank, the better the performance of the method.

Table 6 Average rank per forecasting method and performance metric

	Performance measures					
	Statistical evaluation			Economic evaluation with different service levels		
For all-time series						
Forecasting method	RMSE	MAE	sMAPE	90%	95%	99%
Naïve method	2.51	2.32	2.23	2.28	2.29	2.30
ARIMA	1.41	1.68	1.72	1.77	1.78	1.79
Holt-Winters	2.07	2.00	2.05	1.95	1.94	1.91
For l = 1						
Naïve method	2.49	2.27	2.18	2.24	2.24	2.25
ARIMA	1.40	1.71	1.73	1.75	1.77	1.81
Holt-Winters	2.11	2.02	2.09	2.00	1.99	1.94
For l = 5						
Naïve method	2.53	2.34	2.04	2.37	2.37	2.37
ARIMA	1.58	1.78	1.86	1.89	1.88	1.83
Holt-Winters	1.89	1.88	2.10	1.74	1.75	1.81

Among all evaluation measures and all-time series, ARIMA performed best. Interestingly, ARIMA is given a better evaluation by the statistical measures and a slightly worse evaluation by the dynamic systems models. The opposite happened for the Holt-Winters method, it was ranked better by the dynamic systems models than by the RMSE, MAE, and sMAPE. The naïve method is clearly the least accurate forecasting method. This is not surprising because it was used as a simple benchmark method.

The comparison of the ranks for different lead times shows that the naïve method performs better for a short lead time measured in all measures except the sMAPE. The change is intuitive because the naïve method assumes the next sales to be equal to the last observed sales. If the lead time is 1 and the demand quite stable, the naïve prediction can be reasonable. Moreover, it can be observed that the higher the lead time, the better performs Holt-Winters compared to ARIMA. For $l = 5$, the dynamic systems models rank Holt-Winters better than ARIMA. However, the statistical measures still rate ARIMA best method.

Fig. 4 displays the development of the average overstock and shortage costs per forecasting method with an increased service level. Naturally, the overstock costs increase with the service level and the shortage costs decrease. However, the decrease in shortage costs seems higher than the increase in overstock costs, especially going from a service level of 95% to 99%. Also striking is that the overstock costs are more than double the shortage costs. And the difference between the overstock costs of different forecasting methods is considerably higher than the difference in the shortage costs.

When comparing the forecasting methods, it is striking that the naïve method has both the highest overstock and shortage costs. ARIMA has the lowest overstock costs, but the second highest/ lowest shortage costs. Holt-Winters has subsequently the lowest shortage and the second highest/ lowest overstock costs.

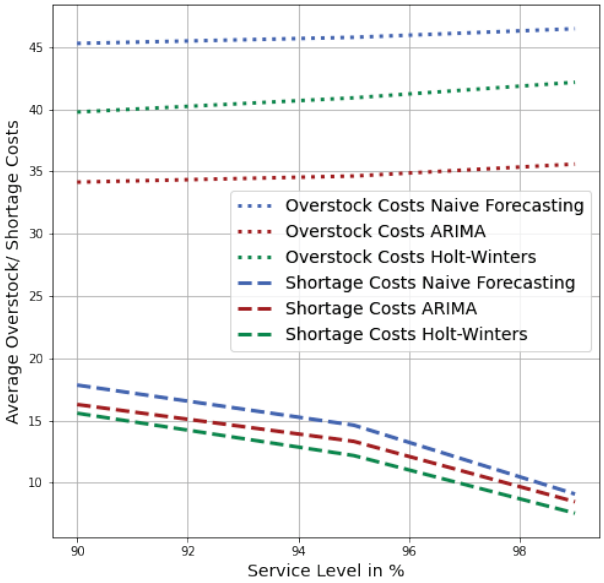


Figure 4 Overstock and Shortage Costs depending on the Service Level and Forecasting Method

6 DISCUSSION

The experiments showed a weakness of the traditional statistical evaluation measures. Even the commonly used RMSE, MAE, and sMAPE did not make coherent statements about the most suitable forecasting method for a time series. Moreover, for roughly half the time series, the cost-considering dynamic systems model came to another conclusion about the best method than the evaluations based upon statistical measures. However, considering costs is crucial for every business context.

The dynamic systems model describes the development of the inventory level over time and derives the resulting shortage and overstock costs. Nevertheless, the shortage costs might have been underestimated because in the real-world dataset, purchases that customers wanted to make but could not due to stockouts of the raw material wholesaler were not documented. Another weakness is that the model does not include order-related costs. It just assumes regular orders and fixed order costs. In practice, order costs vary and for example, quantity discounts could encourage ordering a larger number of goods at once.

Moreover, the assumptions stated in section 3 could be questioned. One assumption is e.g. that lead times are fixed. Recently during the pandemic, we have seen that lead times can significantly change and disturb a steady good supply.

7 CONCLUSION

Literature provides many different methods for evaluating a sales forecast. We have empirically shown on a real-world dataset of 2,911 time series that evaluations by the most commonly used measures RMSE, MAE and sMAPE (as an adaption of MAPE) differ considerably. Moreover, it has introduced a dynamic systems model to evaluate sales forecasts based on shortage and overstock costs. The costs are measured as the percentage of a product's price and can be adapted to the company's needs. In the experiments in this paper, the shortage costs were assumed to be equal to a product's profit margin and the overstock costs to a company's WACC. However, they could also be adapted to include penalty shortage costs for reputation loss in case of a stockout.

The main advantage of the dynamic systems model is that it considers the development of the inventory level over time. It can reproduce the costs that would arise from ordering based on a certain sales forecast. Thus, it helps to choose the most suitable forecasting method, minimizes inventory-related cost, maximizes profit and enables a smooth inventory flow.

In the experiments, three different service levels were regarded. It was shown that the evaluations of the forecasting methods do not differ considerably for the service levels. But, especially for products with a high profit margin, they have an impact on the evaluation. In future work, the service level could be integrated into the dynamic systems model as a tuning parameter to optimize the overall costs. Moreover, the dynamic systems model could be extended into a more complex model that considers variable lead times, storage, and order costs.

8 REFERENCES

[1] Schmid, L., Roidl, M., & Pauly, M. (2023). Comparing statistical and machine learning methods for time series forecasting in data-driven logistics--A simulation study. arXiv preprint arXiv:2303.07139.

[2] Gružauskas, V., Gimžauskienė, E., & Navickas, V. (2019). Forecasting accuracy influence on logistics clusters activities: The case of the food industry. *Journal of Cleaner Production, 240*(1), 118225. https://doi.org/10.1016/j.jclepro.2019.118225

[3] Wheelwright, S., Makridakis, S., & Hyndman, R. J. (1998). *Forecasting: methods and applications*. John Wiley & Sons.

[4] Moosavi, J., Fathollahi-Fard, A. M., & Dulebenets, M. A. (2022). Supply chain disruption during the COVID-19 pandemic: Recognizing potential disruption management strategies. *International Journal of Disaster Risk Reduction*, 102983. https://doi.org/10.1016/j.ijdrr.2022.102983

[5] Petropoulos, F. et al., 2022. Forecasting: theory and practice. *International Journal of Forecasting, 38*(3), 705-871.

[6] Ickstadt, K. et al., (2022). Lernverfahren der Künstlichen Intelligenz zur Inwertsetzung von Daten: Auto-matisierte Erkennung und Prognose. In Silicon Economy: Wie digitale Plattformen industrielle Wertschöpfungsnetz-werke global verändern, 229-250. Berlin, Heidelberg: Springer Berlin Heidelberg. (in German) https://doi.org/10.1007/978-3-662-63956-6_11

[7] Huang, H., Pouls, M., Meyer, A., & Pauly, M. (2020). Travel time prediction using tree-based ensembles. In *Computational Logistics, the 11th International Conference, ICCL 2020*, Enschede, The Netherlands, September 28-30. Proceedings 11, 412-427. Springer International Publishing. https://doi.org/10.1007/978-3-030-59747-4_27

[8] Kiefer, D., & Ulmer, A. (2019). Application of Artificial Intelligence to optimize forecasting capability in procurement. In *Wissenschaftliche Vertiefungskonferenz: Informatik-Konferenz an der Hochschule Reutlingen*, 27. November 2019, 69-80.

[9] Ord, K., Fildes, R. A., & Kourentzes, N. (2017). *Principles of business forecasting*.
[10] Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages. *O. N. R. Memorandum No. 52*, Carnegie Institute of Technology, Pittsburgh USA.
[11] Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science, 6*(3), 324-342. https://doi.org/10.1287/mnsc.6.3.324
[12] Liu, N., Ren, S., Choi, T. M., Hui, C. L., & Ng, S. F. (2013). Sales forecasting for fashion retailing service industry: a review. *Mathematical Problems in Engineering*, 2013. https://doi.org/10.1155/2013/738675
[13] Jiang, H., Ruan, J., & Sun, J. (2021). Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast. In: *The 6th IEEE International Conference on Big Data Analytics (ICBDA)*, March, 5-8, Xiamen, China. https://doi.org/10.1109/ICBDA51983.2021.9403224
[14] Adam, S. P., Alexandropoulos, S. A. N., Pardalos, P. M., & Vrahatis, M. N. (2019). No free lunch theorem: A review. *Approximation and Optimization: Algorithms, Complexity and Applications*, 57-82. https://doi.org/10.1007/978-3-030-12767-1_5
[15] Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A., & Kamaev, V. A. E. (2013). A survey of forecast error measures. *World applied sciences journal, 24*(24), 171-176.
[16] Long, S. & Liu, Q. (2021). Research on New Energy Vehicle Sales Forecast and Product Optimization Based on Data Mining. In *the 2nd IEEE International Conference on Electronics, Communications and Information Technology (CECIT)*, 1019-1024. https://doi.org/10.1109/CECIT53797.2021.00181
[17] Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., & Zhang, R. (2020). ForeXGBoost: passenger car sales prediction based on XGBoost. *Distributed and Parallel Databases, 38*, 713-738. https://doi.org/10.1007/s10619-020-07294-y
[18] Ramos, P., Santos, N., & Rebelo, R. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and computer-integrated manufacturing, 34*, 151-163. https://doi.org/10.1016/j.rcim.2014.12.015
[19] Li, J. (2022). A Feature Engineering Approach for Tree-based Machine Learning Sales Forecast, Optimized by a Genetic Algorithm Based Sales Feature Framework. In *the 5th IEEE International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 133-139. https://doi.org/10.1109/ICAIBD55127.2022.9820532
[20] Schmidt, A., Kabir, M. W. U., & Hoque, M. T. (2022). Machine learning based restaurant sales forecasting. *Machine Learning and Knowledge Extraction, 4*(1), 105-130. https://doi.org/10.3390/make4010006
[21] Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International journal of forecasting, 9*(4), 527-529. https://doi.org/10.1016/0169-2070(93)90079-3
[22] Dhakal, C. P. (2017). A naïve approach for comparing a forecast model. International Journal of Thesis Projects and Dissertations, 5(1), 1-3.
[23] Axsäter, S. (2015). *Inventory control* (Vol. 225). Springer. https://doi.org/10.1007/978-3-319-15729-0
[24] Waller, M. A. & Esper, T. L. (2014). *The definitive guide to inventory management: Principles and strategies for the efficient flow of inventory across the supply chain*. Pearson Education.
[25] Wu, J., Li, J., & Ou, H. Y. (2006, October). Impact of shortage cost in a risk-averse newsboy model. In *IEEE International Conference on Service Systems and Service Management, 1*, 278-282. https://doi.org/10.1109/ICSSSM.2006.320626
[26] Hasbullah, H., & Santoso, Y. (2020). Overstock Improvement by Combining Forecasting, EOQ, and ROP. *J. PASTI, 14*(3), 230-242. https://doi.org/10.22441/pasti.2020.v14i3.002
[27] Borade, A. B. & Bansod, S. V. (2011). Comparison of neural network-based forecasting methods using multi-criteria decision-making tools. *Supply Chain Forum: An International Journal, 12*(4), 4-14. https://doi.org/10.1080/16258312.2011.11517276
[28] Qin, Y., Wang, R., Vakharia, A. J., Chen, Y., & Seref, M. M. (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research, 213*(2), 361-374. https://doi.org/10.1016/j.ejor.2010.11.024
[29] Fishwick, P. A. (Ed.). (2007). *Handbook of dynamic system modeling*. CRC Press. https://doi.org/10.1201/9781420010855
[30] Radasanu, A. C. (2016). Inventory management, service level and safety stock. *Journal of Public Administration, Finance and Law, 9*, 145-153.
[31] Lorek, K. S. & McKeown, J. C. (1978). The effect on predictive ability of reducing the number of observations on a time-series analysis of quarterly earnings data. *Journal of Accounting Research*, 204-214. https://doi.org/10.2307/2490418
[32] De Alba, E. & Mendoza, M. (2007). Bayesian forecasting methods for short time series. *The International Journal of Applied Forecasting, 8*, 41-44.

Authors' contacts:

Lara Kuhlmann, M.Sc.
(Corresponding author)
Graduate School of Logistics, Department of Mechanical Engineering and Department of Statistics, TU Dortmund University,
Leonhard-Euler-Straße 5,
44227 Dortmund, Germany
004915775241591, lara.kuhlmann@tu-dortmund.de

Markus Pauly, Prof. Dr.
Research Center Trustworthy Data Science and Cybersecurity and Department of Statistics, TU Dortmund University,
Otto-Hahn-Straße 14,
44227 Dortmund, Germany
pauly@statistik.tu-dortmund.de

APPENDIX

Acronyms of Error Measures

Acronym	Full name
MAE	Mean Absolute Error
MdAE	Median Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
MAPE	Mean Absolute Percentage Error
MdAPE	Median Absolute Percentage Error
RMSPE	Root Mean Square Percentage Error
RMdSPE	Root Median Square Percentage Error
sMAPE	Symmetric Mean Absolute Percentage Error
sMdAPE	Symmetric Median Absolute Percentage Error
msMAPE	Modified Symmetric Mean Absolute Percentage Error
MRAE	Mean Relative Absolute Error
MdRAE	Median Relative Absolute Error
MASE	Mean Absolute Scaled Error
RMSSE	Root Mean Square Scaled Error
RMAE	Relative Mean Absolute Error
RRMSE	Relative Root Mean Square Error
LMR	Log Mean Squared Error Ratio
PB(MAE)	Percentage Better (MAE)
mRMSE	Normalized Root Mean Square Error
inRSE	Integral Normalized Mean Square Error
Std_AE	Standard Deviation of Absolute Error
Std_APE	Standard Deviation of Absolute Percentage Error

Derivation of Formula O_t

In Period *t*, there is no information about the actual demand (y_t) of future periods. Thus, it needs to be assumed that y_t equals the predicted demand ($\hat{y}_t$). For the periods after placing the order and before the goods arrive, we assume that the number of delivered goods equals the predicted demand.

Subsequently, we only need to consider the predicted demand for the current period *t* and the period *t+l* (lead time), in which the ordered goods will arrive. The sum of the inventory level at the beginning of period *t* (I_{B_t}) and the ordered goods (O_t) should equal the sum of the demand for the current period ($\hat{y}_t$), the demand for *t+l* ($\hat{y}_{t+l}$) and the safety stock (*s)*:

$$I_{B_t} + O_t = \hat{y}_t + \hat{y}_{t+l} + s. \tag{14}$$

Rearranging the formula and considering that the number of ordered goods cannot be negative leads to:

$$O_t = \max\left(\hat{y}_{t+l} + s + \hat{y}_t - I_{B_t}, 0\right). \tag{15}$$

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230518081537

Preliminary communication
Received: 2023-05-18, Accepted: 2023-05-22

Digital Supply Chain Twins in Urban Logistics System – Conception of an Integrative Platform

Lars Tasche*, Maximilian Bähring, Benno Gerlach

Abstract: Current trends in urban areas pose several challenges to city logistics stakeholders while also offering opportunities for optimization. With its analytics, modelling and simulation capabilities, the Digital Supply Chain Twin (DSCT) technology provides a possibility to optimize urban logistics processes. However, a number of barriers have limited the implementation of holistic DSCTs so far. An integrative, collaborative platform could decrease these barriers. By applying design science research methodology and expert interviews, this paper develops an architecture for a high-level cross-institutional platform for the generation of DSCTs. This framework includes a modular design of the platform through eight functional modules. The platform can facilitate the implementation of DSCTs for urban stakeholders and thus optimize urban logistics processes.

Keywords: cross-institutional platform; design science research; digital supply chain twin; digital twin; urban logistics

1 INTRODUCTION

The 21st century is the century of cities. For the first time ever, more people worldwide lived in cities than in rural areas in 2008. And the trend is clear: cities are the living space of the future. The United Nations (UN) predict a global population share of almost 70 % in urban areas by 2050 [1]. Cities are places of transformation and innovation, but at the same time this development condenses challenges. Congestion, air pollution, and growing demand for goods in urban areas are difficult to balance with an increased awareness of environmental and economic sustainability.

Several trends can be identified that influence urban logistics. Consumer behaviour in many relevant markets has shifted, and purchasing transactions are increasingly taking place online. The growing e-commerce has additionally been strongly influenced in the recent past by the COVID-19 pandemic [2]. Deliveries are getting more and more small-scale, the number of parcel deliveries is increasing, and customers have increasing demands for fast and efficient deliveries [2]. Goods can hardly be bundled due to a small-scale delivery quantities and distributed delivery points. These drivers cause that the so-called last mile is regarded as an important part of the supply chain, but is often linked with enormous costs and inefficiency [3].

Innovative measures must be developed to fulfil the increasing requirements for sustainable city logistics and planning in the inner-city areas. The UN Sustainable Development Goal 11 is a commitment to making cities safe, resilient and sustainable, and simulations can play an important role in the implementation [1]. Simulations are frequently used to analyse new logistics and supply chain concepts, providing a reliable method of assessing logistical challenges with a diverse range of applications [4]. However, current simulation models and associated optimizations have so far primarily been used for strategic planning [5]. These models are often constrained to single-use scenarios, failing to leverage the complete process optimization potential within a continually evolving urban landscape. To generate models and decision-making also for operational processes, timely access to up-to-date data is crucial [5]. The long-term, bidirectional exchange between a digital simulation model and a real logistics system with timely data enables a new approach called digital supply chain twin (DSCT) [6]. DSCTs are virtual replicas of logistics systems and can mirror everything from real supply chains – assets, inventories, logistics flows, transactions and relationships, to name a few [7]. The technology makes it possible to create continuous improvement and adaptation of the entire value chain, as well as end to end visibility and thus improving resilience [8].

However, urban players still face the challenge of how to integrate DSCTs into their existing processes to reap the benefits. The technology increases complexity within enterprises, as specific expertise is required for integration and security standards [9]. In addition to implementation costs, economic factors for underlying technologies are major hurdles [10]. To exploit the potential of DSCT and facilitate its implementation, close collaboration among stakeholders is needed. Urban actors need a technical platform to collect and analyse data in one place and to introduce DSCT technology [11]. These requirements pave the way for the development of an integrative, collaborative platform that can reduce the challenges for generating DSCTs. The objective of this study is to conceptualize such an integrative platform for the generation of DSCTs that reduce the growing challenges in urban logistics.

2 DIGITAL SUPPLY CHAIN TWINS

To meet the ecological and economic requirements of sustainable urban logistics, it is necessary to optimize and redesign existing logistics and transport systems [12]. Innovative measures are needed to implement these changes. Simulation models are already widely used in logistics and supply chain management (LSCM) as well as in urban planning. These simulations are a reliable method for evaluating LSCM problems with their dynamic and complex interrelationships, as they can be used to create a database with various indicators, show implementation scenarios and alternative actions as well as support the stakeholders in decision making [4]. Furthermore, simulations in a LSCM context offer particular diversity regarding their use. Individual assets up to entire supply chains and transport networks can be simulated, and the models can also be

applied in various ways to the problems of urban logistics [13]. However, data-driven simulation models and related optimizations have been used mainly for strategic planning so far [5]. They are often limited to a one-time application and do not exploit the full potential of supply chain optimization in a constantly changing urban environment. To enable models and decision making also for operational processes and in long-term, it is crucial that data is available in a timely manner [5]. To address these challenges of logistic simulations, digital supply chain twin technology has gained importance in various business areas and industries in recent years. Based on existing literature, three essential characteristics of DSCTs can be identified: a bidirectional data exchange between the twin and the logistics system is necessary, the model and the real system are in timely synchronization, and it is a long-term approach [6]. In the context of this research, a DSCT is therefore defined as "a digital simulation model of a real logistics system, which features a long-term, bidirectional, and timely datalink to that system. Through observing the digital model, it is possible to acquire information about the real logistics system to conclude, make decisions and carry out actions in the real world [6]. DSCTs have become a critical component of Industry 4.0, that emphasizes the integration of advanced technologies such as artificial intelligence (AI), Internet of Things (IoT), and automation to create smarter, more connected factories and supply chains. Following [6], the digital twin technology in LSCM can be categorized into three levels. The so-called asset level does not represent a DSCT, it is a digital twin of logistics assets. The site level describes a DSCT of a logistics site (e.g., warehouse, production hall, etc.), while the network level represents a DSCT of a multi stakeholder value network. In the context of the research, mainly the network level is considered.

Urban stakeholders are faced with the challenge of finding the most effective way to integrate DSCT technology into their processes to fully exploit the benefits. The technology is associated with high costs and security requirements while also increasing the complexity within an organization. Specific expertise is required to ensure successful integration and data protection [9]. Additionally, the costs for the underlying technologies such as application programming interfaces (APIs), open standards, AI, IoT, or cloud computing are considered obstacles for the integration [10]. Consequently, it has rarely been feasible to introduce network-level DSCTs in urban areas. To make DSCT implementation feasible for companies, close cooperation among stakeholders in the urban environment is necessary. Actors need a centralized platform for data aggregation and analyses to fully exploit the technology's potential [11]. These challenges presented that the use of an integrative and collaborative platform could alleviate the difficulties associated with generating DSCT.

Current research shows that both DSCT technology and corresponding platform approaches have gained importance. For instance, ref. [14] proposed the potential of DSCT for policymaking and planning in urban freight, taking a multi-stakeholder approach and urban focus. However, this study did not provide any conceptual architecture or implementation details. Ref. [15] proposed a framework for designing and assessing urban logistics policies, involving all relevant actors and adopting a multi-stakeholder approach. [16] presented an architecture following a platform approach and conceptualised a freight parking management use case for last-mile delivery. Considering the increasing use of automated vehicles, ref. [17] proposed a reference architecture as an intelligent transportation system, following a network view, thereby analysing aspects of transport mobility, safety and environmental applications. Ref. [18] conducted relevant application scenarios of a whole digital twin city. In regards to urban management, ref. [19] propose requirements and a basic structure of a city digital twin to improve urban planning and optimize asset management. Additionally, ref. [20] focused on an urban digital twin for urban planning and the development of smart cities. The authors analysed a variety of applications and considered a platform approach to be a promising concept for the implementation of an urban digital twin. However, no logistics focus was given in their research. The analysed literature has shown that DSCT with an urban focus are gaining relevance, and single applications areas were already analysed. However, current research lacks an approach, in which a holistic, multi-stakeholder view of urban logistics is combined with a DSCT implementation approach. This paper addresses this research gap by providing an architecture for a cross-institutional platform approach that can support stakeholders in the implementation of a DSCT.

3 RESEARCH DESIGN

For this work, design science research (DSR) methodology was chosen as a suitable research method. DSR is intended to develop new research results, so-called artefact, through a practice-oriented approach [21]. It is less about contribution to knowledge in the sense of classical science and more about a result contribution [22]. Through the integration of research and practice, intense and short feedback cycles are generated, which lead to an early research output and enable a fast integration into the problem areas. All in all, it can be stated that DSR is useful as a method, if a practical result is targeted [22]. In engineering disciplines, DSR is accepted as a valid and valuable research method because the research culture values incrementally applicable, effective solutions to problems [21]. The aim of this work is to obtain the framework of such an applicable solution to address urban logistics and supply chain management problems. Therefore, DSR is an appropriate research methodology for this work.

Authors in ref. [21] have created a generally accepted, common and comprehensible framework for presenting and conducting DSR. It consists of six iterative activities: the first step is problem identification and motivation, whereupon requirements/objectives for a solution are defined within the second step. From this, in the third step a solution approach (artefact) is designed and developed, which is afterwards demonstrated. Finally, step five includes the evaluation of the artefact. The last step communication is realized by this paper. The following figure visualizes the procedure.

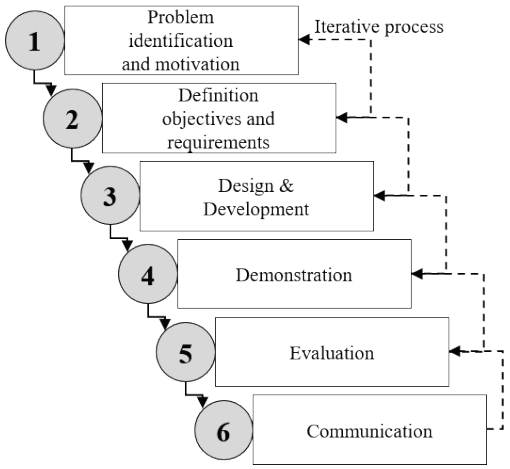


Figure 1 Design Science Research Methodology Process Model [21]

3.1 Problem Identification and Motivation

In this initial phase of DSR, specific research problems are defined, and the value of a solution is justified. In this case, the focus is on identifying problems in a urban LSCM context. The problem definition is the foundation for developing an artefact in the subsequent process, which will represent a solution to the identified problems [22]. According to [23] urban LSCM problems can broadly be categorized into four major areas: urban logistics systems and management, infrastructure, technology and economic factors.

The field of *urban logistics systems and management* faces significant challenges, primarily arising from the high degree of uncertainty and dynamic properties inherent in urban systems [24]. While technology trends such as Industry 4.0 have facilitated increased connectivity among customers, partners, and suppliers [25], companies still struggle with the coordination and optimization of networks and collaborations involving various actors [25]. Additionally, the lack of information and transparency within supply chains hinder an optimal control in complex urban systems, especially with the increasing social, ecological and economic awareness of customers. Not all companies are able to meet these demands, as creating transparency and changing to more sustainable processes may require additional effort and costs [23].

Relating to the second area of *infrastructure*, companies face congestion and high carbon footprints arising from increasing traffic volumes. Addressing LSCM problems and last-mile steps in urban systems requires a consideration of both logistical and urban planning challenges [23]. Logistics performance in cities is influenced by three key dimensions of urban planning: the built environment (with attributes such as population density), planning control (a system that addresses issues of parking, loading, and unloading) and transportation control (a system that addresses speed limits, traffic signals, bus lanes, railroad crossings, etc.) [23]. This interdependence of LSCM processes and urban planning highlights the complexity of urban systems. While emerging technologies and alternative fulfilment concepts may help reduce these problems, they also require new transportation and facility infrastructure.

The ongoing *technological transformation* is leading to the emergence of novel technologies, such as autonomous delivery vehicles or drones, which may replace or significantly transform existing urban logistics systems [26]. Such developments pose significant challenges to companies due to the associated uncertainty factors, especially when considering the integration into existing structures. Furthermore, the trend towards e-commerce is associated with large losses for many businesses operating in urban areas due to inadequate goods delivery planning and fulfilment [27]. This is further compounded by the fragmentation of orders leading to a high number of small orders, as well as narrow delivery windows, which add to the complexity of urban logistics systems [23].

Finally, the *economic problems* that can arise for companies in urban logistics systems are considered. In general, the last mile and urban LSCM are associated with high costs, with freight last mile logistics contributing on average to 28 % of the total delivery costs along the entire value chain [28]. In particular, the growing online and omni-channel grocery business involves high costs and complex fulfilment processes [29]. The fulfilment costs depend on various factors such as service time, service area, distance from distribution centre, driver costs and investment costs [23]. Not only the cost of delivery, but also the cost estimation and planning for goods delivery is challenging in urban areas. It is associated with many constraints related to traffic, driver working hours, refuelling intervals, battery charging processes and capacity (for electric transportation) and more [30].

Several tools are available to face the problems, including online freight exchange platforms, enterprise resource planning (ERP) systems, advanced planning and scheduling (APS) systems, transportation management systems (TMS) and supply chain management (SCM) systems [31]. Online exchange platforms connect transportation, logistics, and freight forwarding companies to market transportation capacities and search for specific freight. However, these platforms do not offer holistic optimization possibilities or sufficient details about freight and order information. Meanwhile, ERP, APS, TMS, and SCM systems are used to plan, control, and manage supply chains. Modern systems are user-oriented and offer tools of process optimization as well as an interface to external partners. Compared to DSCTs, these applications have major disadvantages that hinder problem-solving: timely data updates, insufficient applied analytical capabilities and lack of simulation capabilities for holistic supply chains [31]. As described in the previous chapter, DSCTs have the ability to improve these dimensions.

3.2 Requirements Definition

According to [21], the second step of DSR involves the definition of requirements for a solution. For this purpose, several requirements were derived from the problem identification and expert interviews. The interviewees represented all players of a generic supply chain. In total, five semi-structured interviews were conducted as part of the research process. The requirements were then developed

based on the insights and inputs gathered from these interviews and the mentioned problems.

The problem definition revealed that the supply chain actors suffer from inadequate information exchange due to insufficient data transfer, leading to a lack of comprehensive process insights for all players. As a result, none of the players has an overview of all processes that affect him, certainly not of all supply chain processes. To address this issue, a solution is required to strengthen the transparency and visibility of supply chain processes with available capacities, disruptions, process status information and all further status information [31]. Consequently, to address this issue, the artefact is required to enhance of data exchange capabilities between companies.

Data sharing is a critical aspect of urban logistics, not only due to the need for supply transparency and visibility but also for enhanced communication and coordination within urban systems. The benefits of data sharing extend beyond businesses and can involve other urban actors as well. Through extensive data exchange facilitated by platforms, communication and coordination among various urban stakeholders can be improved. Therefore, it is imperative that proposed solutions encourage data exchange between all relevant urban actors to ensure a holistic approach to urban logistics. Such data sharing can have a significant impact on enhancing the overall efficiency and sustainability of urban systems. Therefore, it is crucial to emphasize the importance of data sharing in the development of proposed solutions for urban logistics.

The existing solutions lack holistic optimization and are characterized by limited analytical and simulation capabilities. To address this limitation, the artifact is a DSCT that integrates modelling, analytics, and simulation capabilities with a high degree of accuracy. Achieving this requires the inclusion of both internal data from the stakeholders of an urban system and external data, such as weather, traffic, and strikes. The combination of data from various sources, including internal and external, is necessary to develop a comprehensive solution that can support holistic optimization in urban systems.

Data obtained from LSCM processes is often not provided in a timely manner. Although real-time data may not always be necessary for all use cases, it is crucial for time-critical tasks. However, it is important to ensure that data is regularly updated to create a DSCT that reflects the current real status of the system without significant uncertainty. Continuous data updates are required to maintain the integrity of the proposed solution.

LSCM processes are not always sustainable, reliable, and resilient. Factors contributing to this include challenges with warehousing and inventory, vehicle utilization, fleet management, vehicle routing, order management and fulfilment planning. To address these weaknesses, accurate models and analytics must be generated that leverage the high-quality data from previous requirements. A thorough modelling and analysis of the current state of the supply chain is required to identify areas of weakness and opportunities for improvement, with the aim of enhancing overall supply chain performance.

The previous chapter has highlighted a significant knowledge and experience gap in managing logistics processes within urban areas. One approach to prevent this is the use of models and simulations to increase stakeholders process understanding. However, there are currently few tools for simulating entire supply chain processes. Without the ability to simulate different scenarios, including what-if analyses, future impacts of possible decisions cannot be determined without affecting entities of the real supply chain. This limitation applies to both strategic and operational levels and results in decision-making being associated with high degrees of uncertainty. Adequate modelling and simulation of LSCM processes can minimize such risks by increasing the process understanding of the decision makers. Therefore, it is required that the artefact can model and simulate (with what-if-scenarios) the supply chain processes.

Digital transformation presents significant opportunities for modifying existing LSCM systems. However, numerous barriers exist that impede the implementation of new technologies and fulfilment concepts, including high costs, lack of knowledge, and established habits. As a result, the integration of new technologies and fulfilment concepts is obstructed. Accurate simulations considering new technologies and concepts minimize this problem as it creates a degree of certainty about how the new approach would operate in the existing system. To effectively reduce the barriers associated with integrating new technologies and concepts, the artefact is required to accurately simulate LSCM processes.

Existing process optimization tools have proven insufficient, causing problems related to cost estimation, goods delivery planning, warehousing and inventory, vehicle utilization, fleet management, vehicle routing and order management. To overcome these issues, comprehensive data-driven decision support can help urban players in their decision and thus optimize LSCM processes in these areas. Therefore, it is required that the solution should incorporate comprehensive data-driven decision support capabilities.

The theoretical foundation has already demonstrated that the DSCT technology offers promising solutions to many of the urban LSCM problems described before. However, the widespread adoption of this technology in businesses has been impeded by various challenges. Especially small and medium-sized enterprises require an affordable and accessible means of implementing DSCT technology into their operations. This can be achieved by integrating them into an appropriate platform and providing software development kits for DSCT generation. It is required that the artefact provides organizations with an IT fundament (platform).

The present section represents the culmination of the requirement gathering phase of the DSR process. While the requirements are able to address a majority of the multiple problems identified in chapter 3.1, it is important to note that certain challenges, such as high traffic volume and urban-specific impediments, may require further attention. Nonetheless, the requirements defined in this section serve as a robust foundation for the subsequent chapter, which will

propose a solution approach for optimizing LSCM processes in urban areas.

3.3 Design and Development

The third step of the DSR methodology involves the design and development of an artefact using a systematic methodology [22]. Authors in ref. [21] describe these artefacts as "potentially constructs, models, methods, or instantiations (each defined broadly) or new properties of technical, social, and/or informational resources." Based on the requirements described in the previous chapter, we developed an artefact using ideation techniques with several iterations. At each iteration, we evaluated the extent to which the defined requirements were met, and continued to refine the artefact until they were fully satisfied. The solution was a high-level architecture of a platform approach for the generation of a DSCT in urban areas. This artefact served as the basis for the next step in the DSR process, in which we conducted five additional expert interviews to theoretically demonstrate the platform approach.

3.4 Demonstration

The subsequent stage of the DSR process is the demonstration phase, which seeks to validate the functionality of the developed artefact. The artefact can then be optimized and adapted by reverting back to the previous stage and repeating the development process. Various activities, such as experiments, simulations, case studies, or proofs, may serve as effective demonstrations. In the context of this work, further expert interviews were selected as an appropriate activity. This qualitative approach is chosen on the fact that DSCTs are a rather new and innovative concept. Generally, resources needed for the demonstration include deep knowledge of the application area as well as knowledge on how to use the artefact to solve the problems. The experts were chosen in order to meet these requirements as they have industry-specific expertise. The architecture and functionality of the platform were presented and demonstrated to the same five experts. During the interviews, the platform architecture and functionality were presented and demonstrated to the interviewees. Following each interview, the artefact was refined based on the feedback received, with the aim of optimizing the architecture. Finally, the last two expert interviews saw no potential for optimisation, leading us to conclude that the development and demonstration stages were completed. The final and adjusted artefact is the platform architecture, which is presented in chapter 4.

3.5 Evaluation

During the evaluation phase of the DSR methodology, the effectiveness of the developed artefact in addressing the problems identified in chapter 3.1 is assessed. The objective is to compare the intended outcomes of the artefact with the actual results obtained through its application. In general, this fifth process step of DSR can also be performed using quantitative methods or, as in this case, the expert interviews. By selecting the suitable interviewees, we were able to generate the necessary knowledge about the relevant industries, enabling us to assess the potential of the platform approach to address the current urban LSCM problems. To achieve this, the final platform architecture was presented to the experts, who were then tasked with evaluating and justifying how specific platform components could potentially solve the urban LSCM problems. The problems were further considered in four categories (see chapter 3.1). The results and justifications are presented in chapter 4.

4 RESULTS

This chapter introduces the platform architecture for generating DSCTs and evaluates its effectiveness in addressing urban LSCM problems.

4.1 Platform Architecture

Based on the requirements and expert interviews, a framework for a cross-institutional platform to generate DSCTs in urban areas is proposed, based on [31]. The platform consists of the integration of various stakeholders, a central instance for the provision of external data, and eight modules. The central instance integrates data from multiple external sources, including city data such as traffic, environmental monitoring, and weather data, which is then made available to all relevant actors. This approach streamlines data sharing and avoids redundant data collection. Specific (aggregated) data from individual urban actors form an information and communication network, whose data is integrated into the DSCTs. The final output is presented in figure 2. A detailed description of the individual modules is presented below.

The proposed solution is designed to transfer a large amount of data to a platform, utilizing technologies such as IoT, 5G, and cloud computing. Within this platform, the data is stored in the *Interface Module*, which is located alongside the SDK Module. The Interface Module plays a crucial role in the further process, fulfilling multiple tasks. On the one hand, all data are available in a wide variety of formats and does not always meet criteria for high data quality (e.g., consistency, completeness, uniqueness, timeliness, validity, accuracy). Interface Module filters the data from various sources and translates it into a suitable format for processing in subsequent steps [31]. Applications of AI can help to process the unstructured data with different quality. On the other hand, the Interface Module also processes the data regarding data privacy and confidentiality. As not all companies are willing to share their entire information, the Interface Module performs filtering and processes the data to comply with privacy protection requirements. All in all, the Interface Module is intended for the storage and processing of data.

The *SDK Module* is an essential component of the platform. It represents a software development kit (SDK) that enables the creation of DSCTs [32]. The platform is accessible to various urban stakeholders who can load the

content onto their machines as executable program files, also known as machine code, that are then loaded into the computer's memory for execution. The primary purpose of this module is to provide urban stakeholders with a suitable system for running a DSCT on their devices [32]. Together with the Interface Module, the SDK Module forms the foundation necessary for companies to generate their DSCT. As illustrated in Fig. 2, the DSCT comprises six additional modules: the Supply Chain Modelling, Simulation, Analytics, Reporting, API and ULC Module.

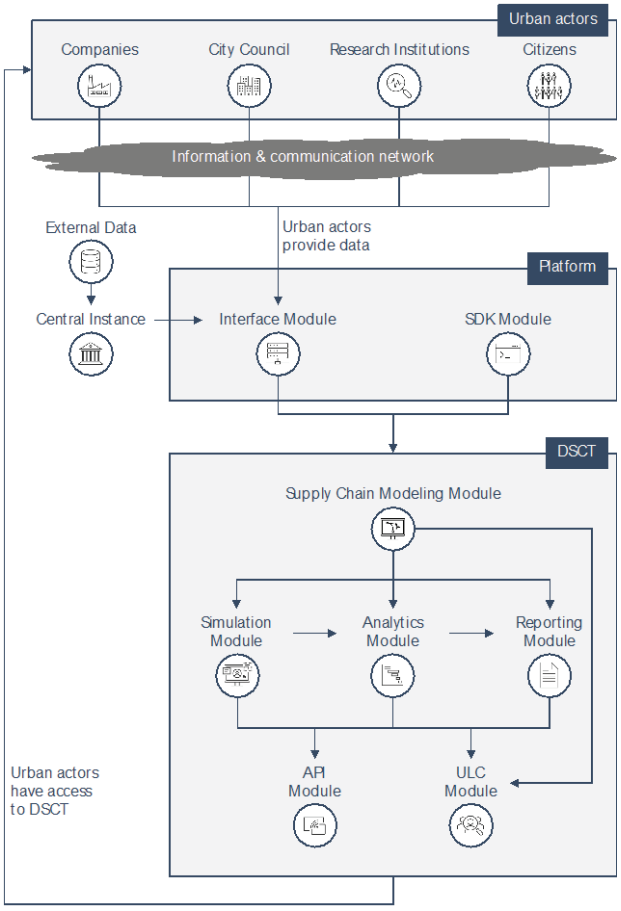


Figure 2 Proposed Architecture and Environment of DSCT Platform [31]

Data and SDK from the platform merge in the DSCT, where they form the *Supply Chain Modelling Module*. It enables to model a real, physical supply chain. The module provides the ability to represent all aspects, interdependencies, and relationships of the intricate urban LSCM systems, thereby offering a better understanding of the underlying processes [31]. By synchronizing real-world data, the current status of the physical supply chain can be monitored (e.g., inventory monitoring, delivery tracking). Classical algorithms or AI techniques can help in achieving this goal.

The *Simulation Module* can determine potential (future) states of the physical supply chain by applying different alternative parameters on the Supply Chain Modelling Module. It provides the ability to determine the impacts and outcomes of decisions and influences without affecting the real supply chain by conducting what-if scenarios. The simulation can be applied to both operational supply chain processes as well as to strategic planning processes. The use of this module optimizes the decision-making process of stakeholders. In addition, the introduction of new technologies and concepts can be simulated, the potential consequences become apparent, allowing stakeholders to anticipate and mitigate any barriers that may arise [31].

The *Analytics Module*, which is based on the Supply Chain Modelling Module, aims both to analyse and provide optimization opportunities of the physical supply chain. This module provides the ability to recognize, interpret, and communicate patterns within the data. The LSCM optimization opportunities made possible by this module can include for instance inventory optimization, maximization of vehicle utilization, efficient fleet management and route planning. These optimization opportunities can be realized with the use of algorithms or AI techniques [31].

The DSCT is further equipped with a *Reporting Module*, which can be regarded as an essential component of the platform [31]. The results obtained from the Analytics Module, the Supply Chain Modelling Module, and the Simulation Module are processed and presented in a structured and clear format, thereby facilitating data-based decision-making by the stakeholders. Effective communication of information, opportunities, and recommendations is critical in enabling the stakeholders to make informed decisions based on the insights gained from the DSCT [31].

As component of the DSCT, the *ULC* (user level customization) *Module* ensures that the DSCT or its individual elements are adapted to the corresponding user. The ULC Module enables the modification of the complexity, information, and decision-making capabilities of the DSCT depending on the user's level of expertise. Authors in ref. [33] have designed a solution approach for user-level customized modelling and simulation in production facilities, which can be used as starting point for the implementation of the ULC Module. However, further research is necessary at the network level to determine the number of modelling levels that should exist and which actor utilizes each level.

Additionally, the *API Module* ensures that the information gained by the DSCT is transferred into appropriate format, enabling the integration of DSCT outputs into existing systems such as SCM software, ERP software, and APS systems. To enable integration with the DSCT platform, existing software must be appropriately interfaced. This, in turn, allows for the integration of the platform with existing LSCM planning operations, thereby supporting an incremental implementation process [31]. The API Module plays an essential role in ensuring interoperability and compatibility between the DSCT and existing systems, thereby enabling the seamless integration of data and information to facilitate process optimization.

4.2 Potential Solutions for Urban LSCM Problems

After presenting the platform architecture, this chapter examines the extent to which urban LSCM problems can be addressed through it. The origin of the described DSCT platform approach are the multifaced problems in urban areas, with a special focus on LSCM. The primary objective

of the platform approach is to enhance the sustainability of LSCM operations in urban settings across the three dimensions of economics, social, and environmental impact. The possible solutions are again considered in the four areas *urban logistics systems* and *management*, *infrastructure*, *technology*, and *economic factors* (see chapter 3.1).

The interviewed experts have assessed that the DSCT platform approach can effectively tackle major problems in the field of urban logistics system and management. According to their consensus, enhanced communication among urban stakeholders can alleviate conflicting interests of different actors in urban systems. The Supply Chain Modelling and Simulation Module can significantly increase process understanding, thereby reducing the lack of know-how regarding last mile processes. In addition, the Modelling, Analytics and Reporting Modules can help to minimise LSCM-specific problems in areas such as warehousing and inventory, vehicle utilization, fleet management, vehicle routing, and order management. The interviewees have also confirmed that the problem of insufficient data acquisition and processing can be addressed with the increased data exchange facilitated by the central instance, Interface and Analytics Modules. The increased data exchange between partners also helps to increase supply chain transparency and visibility, thus enabling the identification of unsustainable and unreliable processes/suppliers, so that the challenges of increasing demands for sustainability and reliability can be addressed. However, there are some problems in the field of urban logistics systems and management that cannot be solved directly by the DSCT platform. These include the increasing complexity of logistics solutions as well as uncertain and dynamic conditions of urban logistics systems.

In the area of infrastructure, the experts also evaluated that some problems can be solved by the DSCT platform. One of the major problems is the adaptation of urban infrastructure to the increasing freight volume. Modelling and simulation can help in the decision-making process for new urban infrastructure, thus facilitating its introduction. The interviewees also stated that the problem of insufficient infrastructure for new fulfilment technologies and concepts can be reduced as the impacts can be simulated beforehand, which can reduce the barriers for the introduction of new technologies and infrastructures. The Modelling and Simulation Module can increase process understanding and replace real experience to a certain extent, which can mitigate the lack of experience in introducing new infrastructure concepts in urban systems. However, the interviewees also acknowledged that the platform approach cannot solve issues related to high traffic volume and challenges resulting from the complex and dynamic condition of urban areas.

In the area of technology, the results are mixed. The experts stated that by simulating the impact of emerging technologies and business models in advance, potential barriers to their integration can be identified.

The Analytics and Reporting Modules also improve the planning of goods delivery, planning and fulfilment. However, the potential of new technologies to replace existing urban systems cannot be addressed.

Finally, the economic issues are considered. The experts confirmed that the Analytics and Reporting Modules can help with data-based analyses, leading to a reduction in cost estimation and better planning for goods delivery. The upfront expenses associated with the implementation and operation of new technologies, as well as the high cost of new infrastructure, cannot be immediately lowered. The expenses associated with urban fulfilment operations are typically high, and the platform may not have a direct impact on reducing them. However, through the simulation of various scenarios, the most financially feasible options can be determined.

5 DISCUSSION

The results illustrate that the DSCT platform has the potential to address numerous urban LSCM problems. Especially in the field of urban logistics system and management, major problems can be addressed. However, certain obstacles cannot be resolved by the DSCT platform, including for instance the complexity of logistics solutions, dynamic conditions of urban systems, high traffic volume and city conditions. Hence, it should be emphasized that the platform is not a universal remedy, but rather a partial solution to urban challenges. Therefore, it is crucial to consider other strategies for the comprehensive optimization of urban LSCM systems. The implementation of the DSCT platform may lead to several issues that have not been considered in this study. These include for instance financial obstacles for both the software and hardware required for data generation and underlying technologies. Furthermore, a high level of technical and economic expertise is necessary to realize the concept, and employees must receive training to operate the platform competently.

From a management perspective, the findings suggest that companies in the LSCM industry should consider the potential benefits and challenges of implementing the DSCT platform. While the platform offers significant advantages such as timely data updates, analytical capabilities, and simulation capabilities for holistic supply chains, it also requires a high level of technical and economic expertise to realize the concept, and employees must receive training to operate the platform competently. Thus, companies need to invest in the necessary technical skills and training to ensure successful implementation and effective use of the platform. Furthermore, companies must also consider the financial obstacles associated with implementing the platform.

From a research perspective, further research is necessary to explore alternative approaches for addressing urban LSCM challenges that are not amenable to the platform. In particular, research should focus on strategies for the comprehensive optimization of urban LSCM systems, considering the complexity of logistics solutions, dynamic conditions of urban systems, high traffic volume, and city conditions. Further research is also needed to investigate the financial obstacles and technical expertise required to implement the DSCT platform effectively. Finally, research should explore the long-term cost-saving potential of the technology and its ability to improve urban systems in terms

of financial and environmental sustainability. By addressing these issues, researchers can contribute to the ongoing development and optimization of the DSCT platform and other innovative solutions for urban logistics challenges.

Although the platform solution cannot solve all problems related to urban logistics and poses several challenges, a clear image emerges from the research: the opportunities associated with the technology integration can surpass the challenges. Compared to traditional solutions like ERP, SCM, or TMS systems, the platform approach offers significant advantages such as timely data updates, analytical capabilities, and simulation capabilities for holistic supply chains. Moreover, the platform has more data generation possibilities and increased connectivity of supply chain actors and urban stakeholders compared to a DSCT utilized by a single supply chain or organization. Additionally, the availability of software development kits simplifies implementation. By combining the advantages of DSCT technology and a platform, the modular architecture provides flexibility for constant adjustment and optimization of the solution. Although the initial financial outlay for implementation may appear very high, the technology offers considerable cost-saving potential in the long term. The technology offers opportunities to improve urban systems in terms of financial and environmental sustainability. All in all, at first glance, introducing the DSCT platform seems challenging. However, in the long term, the benefits can outweigh the challenges. This study has established an initial framework to resolve urban logistics issues.

6 SUMMARY AND CONLUSION

The 21st century is the century of cities. Urban spaces are transforming, and with-it urban logistics. This study explores how a digital platform for generating DSCTs can address urban logistics challenges. It uses DSR methodology to identify and categorize urban LSCM problems, elaborate requirements, design the platform, and evaluate its effectiveness. The final result is a high-level platform architecture and an assessment of its potential to solve urban logistics problems. It can be concluded that the DSCT platform offers a viable approach for addressing numerous urban LSCM problems. These problems include insufficient knowledge about last-mile procedures, inadequate data generation and processing, lack of supply transparency and visibility, inadequate communication among urban stakeholders, and challenges in adopting new technologies Nonetheless, it is important to acknowledge that the proposed platform and our research have limitations and cannot provide a comprehensive solution to all urban LSCM issues. To overcome these limitations, additional research and approaches, such as reducing urban traffic or minimizing the high costs of fulfilment operations, are needed. The limitations of this study include its theoretical and conceptual nature. The DSCT platform was not practically implemented, and the research relied solely on a qualitative approach with a relatively small number of five interviews. As a next step it is essential to practically implement and verify the platform in future research. All in all, based on current trends, it is expected that DSCT technology and platform approaches will become increasingly important for companies in the future, despite some challenges. The technologies provide numerous opportunities to address urban problems and promote sustainability in both cities and LSCM. The elaboration of the DSCT platform approach in this research aims to facilitate and encourage its implementation and address urban LSCM problems.

7 REFERENCES

[1] United Nations, Department of Economic, & Social Affairs, P. D. (2019). World Urbanization Prospects: The 2018 Revision. United Nations.

[2] Rotar, A. (2020). eCommerce Report 2020 - Statista Digital Market Outlook – Market Report. Statista.

[3] Gevaers, R., van Voorde, E. de, & Vanelslander, T. (2014). Cost Modelling and Simulation of Last-mile Characteristics in an Innovative B2C Supply Chain Environment with Implications on Urban Areas and Cities. *Procedia - Social and Behavioral Sciences, 125*, 398-411. https://doi.org/10.1016/j.sbspro.2014.01.1483

[4] Eley, M. (2012). Simulation in der Logistik - Einführung in die Erstellung ereignisdiskreter Modelle unter Verwendung des Werkzeuges "Plant Simulation". Springer-Lehrbuch. Springer Gabler.

[5] Ivanov, D. (2018). Managing Risks in Supply Chains with Digital Twins and Simulation.

[6] Gerlach, B., Zarnitz, S., Nitsche, B., & Straube, F. Digital Supply Chain Twins—Conceptual Clarification, Use Cases and Benefits.

[7] Marmolejo-Saucedo, J. A. (2020). Design and Development of Digital Twins: A Case Study in Supply Chains. *Mobile Networks and Applications, 25*(6), 2141-2160. https://doi.org/10.1007/s11036-020-01557-9

[8] Hosseini, S., Ivanov, D., & Dolgui, A. (2019). Review of quantitative methods for supply chain resilience analysis. *Transportation Research Part E: Logistics and Transportation Review, 125*, 285-307. https://doi.org/10.1016/j.tre.2019.03.001

[9] King, C. (2020). Digital Twin - Mit dem digitalen Zwilling die Zukunft simulieren. Mindlogistik. https://mind-logistik.de/knowhow/digital-twin/

[10] Dohrmann, K., Giesig, B., & Ward, J. (2019). Digital Twins in Logistics. A DHL perspective on the impact of digital twins on the logistics industry. https://www.dhl.com/content/dam/dhl/global/core/documents/pdf/glo-core-digital-twins-in-logistics.pdf

[11] Riedelsheimer, T., Lünnemann, P., Wehking, S., Dorfhuber, L., Samarajiwa, M., Lindow, K., Salamon, D., & Stark, R. (2020). Digital Twin Readiness Assessment - Eine Studie zum digitalen Zwilling in der fertigenden Industrie.

[12] Straube, F. & Junge, A. L. (2019). Smarte Konzepte für zukunftsfähige urbane Logistik- und Verkehrssysteme. *Internationales Verkehrswesen, 71*(2), 35.

[13] Chow, J. Y. J., Yang, C. H., & Regan, A. C. (2010). State-of-the art of freight forecast modeling: Lessons learned and the road ahead. *Transportation, 37*(6), 1011-1030. https://doi.org/10.1007/s11116-010-9281-1

[14] Edoardo Marcucci, Valerio Gatta, Michela Le Pira, & Lisa Hansson and Svein Bråthen. Digital Twins: A Critical Discussion on Their Potential for Supporting Policy-Making and Planning in Urban Logistics.

[15] Belfadel, A., Horl, S., Tapia, R. J., & Puchinger, J. (2021). Towards a Digital Twin Framework for Adaptive Last Mile City Logistics. In *The 6th IEEE International Conference on Smart and Sustainable Technologies (SpliTech)*, 1-6. https://doi.org/10.23919/SpliTech52315.2021.9566324
[16] Li, Q., & Liu, A. (2019). Big Data Driven Supply Chain Management. *Procedia CIRP, 81*, 1089-1094. https://doi.org/10.1016/j.procir.2019.03.258
[17] McKerahan, T. Irfan et al. (2022) Towards Transportation Digital Twin Systems for Traffic Safety and Mobility Applications.
[18] Guo, J., Zhao, N., Sun, L., & Zhang, S. (2019). Modular based flexible digital twin for factory design. *Journal of Ambient Intelligence and Humanized Computing, 10*(3), 1189-1200. https://doi.org/10.1007/s12652-018-0953-6
[19] Ferré-Bigorra, J., Casals, M., & Gangolells, M. (2022). The adoption of urban digital twins. *Cities, 131*, 103905. https://doi.org/10.1016/j.cities.2022.103905
[20] Yang, S., & Kim, H. (2021). Urban digital twin applications as a virtual platform of smart city. *International Journal of Sustainable Building Technology and Urban Development, 12*(4), 363-379. https://doi.org/10.22712/susb.20210030
[21] Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems, 24*(3), 45-77. https://doi.org/10.2753/MIS0742-1222240302
[22] Benner-Wickner, M., Kneuper, R., & Schlömer, I. (2020). Leitfaden für die Nutzung von Design Science Research in Abschlussarbeiten. *IUBH Discussion Papers, 2*(2), 11.
[23] Bosona, T. (2020). Urban Freight Last Mile Logistics - Challenges and Opportunities to Improve Sustainability: A Literature Review. *Sustainability, 12*(21). https://doi.org/10.3390/su12218769
[24] Gómez-Marín, C. G., Arango-Serna, M. D., & Serna-Urán, C. A. (2018). Agent-based microsimulation conceptual model for urban freight distribution. Transportation Research Procedia, 33, 155-162. https://doi.org/10.1016/j.trpro.2018.10.088
[25] Juhász, J., & Bányai, T. (2018). Last mile logistics: An integrated view. *IOP Conference Series: Materials Science and Engineering, 448*. https://doi.org/10.1088/1757-899X/448/1/012026
[26] Edwards, J., McKinnon, A., Cullinane, S., & Potter, A. (2011). Comparative carbon auditing of conventional and online retail supply chains: A review of methodological issues. *Supply Chain Management: An International Journal, 16*(1), 57-63. https://doi.org/10.1108/13598541111103502
[27] Melacini, M., Perotti, S., Rasini, M., & Tappia, E. (2018). E-fulfilment and distribution in omni-channel retailing: A systematic literature review. *International Journal of Physical Distribution & Logistics Management, 48*(4), 391-414. https://doi.org/10.1108/IJPDLM-02-2017-0101
[28] Ranieri, L., Digiesi, S., Silvestri, B., & Roccotelli, M. (2018). A Review of Last Mile Logistics Innovations in an Externalities Cost Reduction Vision. *Sustainability, 10*(3), 782. https://doi.org/10.3390/su10030782
[29] Hübner, A. H., Kuhn, H., Wollenburg, J., Towers, N., & Kotzab, H. (2016). Last mile fulfilment and distribution in omni-channel grocery retailing: A strategic planning framework. *International Journal of Retail & Distribution Management, 44*(3). https://doi.org/10.1108/IJRDM-11-2014-0154
[30] Sheth, M., Butrina, P., Goodchild, A., & McCormack, E. (2019). Measuring delivery route cost trade-offs between electric-assist cargo bicycles and delivery trucks in dense urban areas. *European Transport Research Review, 11*(1), 11. https://doi.org/10.1186/s12544-019-0349-5
[31] Busse, A., Gerlach, B., Lengeling, J. C., Poschmann, P., Werner, J., & Zarnitz, S. (2021). Towards Digital Twins of Multimodal Supply Chains. *Logistics, 5*(2), 25. https://doi.org/10.3390/logistics5020025
[32] Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets, 30*(1), 87-98. https://doi.org/10.1007/s12525-019-00377-4
[33] Kim, B. S., Jin, Y., & Nam, S. (2019). An Integrative User-Level Customized Modeling and Simulation Environment for Smart Manufacturing. IEEE Access, 7, 186637-186645. https://doi.org/10.1109/ACCESS.2019.2961450

Authors' contacts:

Lars Tasche, MSc
(Corresponding author)
Technical University of Berlin, Chair of Logistics,
Straße des 17. Juni 135, 10623 Berlin, Germany
tasche@logistik.tu-berlin.de

Maximilian Bähring, MSc
Technical University of Berlin, Chair of Logistics,
Straße des 17. Juni 135, 10623 Berlin, Germany
baehring@logistik.tu-berlin.de

Benno Gerlach, MSc
Technical University of Berlin, Chair of Logistics,
Straße des 17. Juni 135, 10623 Berlin, Germany
gerlach@logistik.tu-berlin.de

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220920083951

Subject review
Received: 2022-09-20, Accepted: 2022-10-20

Technical Characteristics of Incunabulum in Europe

Stanislav Bolanča, Ivana Bolanča Mirković*, Ivan Pučić

Abstract: Incunabula are printed materials created in Europe from the time of Johann Gutenberg's invention until 1500. Incunabula originate from the Latin language (lat. Incunabulum) and mean cradle or the beginning of something. In this paper, the representation of individual states and cities in the creation of incunabula is investigated and presented. The persons responsible for such development are also listed. Special attention is given to the presentation of Croatian incunabula. The mentioned works describe the characteristic features. Incunabula testify to a high level of culture, standards, and technological development of a particular area. The studied works reveal and confirm, as confirmed in this paper, the attitude of society towards literacy, education, and the national culture of each nation. This paper aims to comprehensively present the importance of incunabula for the development of European and Croatian culture, technological and comprehensive progress.

Keywords: Glagolitic alphabet; incunabulum; Latin alphabet; printing press

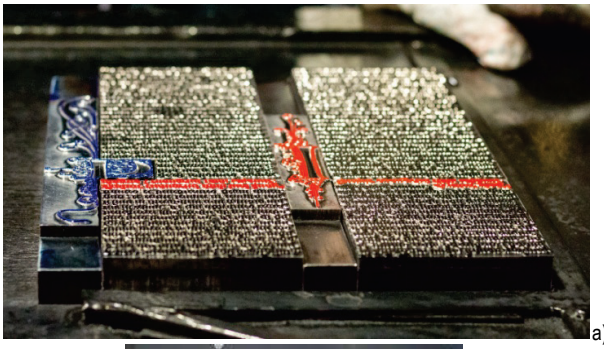
1 INTRODUCTION

The turning point in the rapid development of European civilization, science, and culture, occurred in the middle or second half of the 15th century. The introduction of metal-based movable-type enabled the replacement of handwritten copying with machine multiplication of the original. The inventor of the technology of moving letters, inks, and presses for machine printing of graphic templates was Johann Gutenberg, an inventor from Mainz, Germany (Fig. 1). Gutenberg was trended as a goldsmith, a worker in metals which enabled him to make metal letters [1]. The Gutenberg press was probably derived presses from agriculture or industry presses for grapes, olives, herbs, or papyrus. The wooden press had a large screw attached to a flat platen [2].

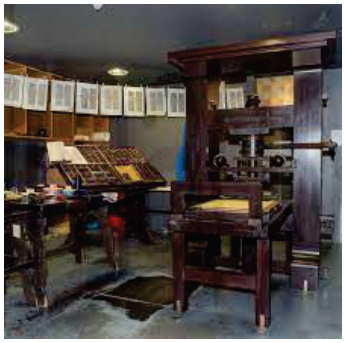
Figure 1 The inventor of movable type and the printing press Johann Gutenberg [5]

To reproduce the original in a lot of copies, the material of the letter had to have certain durability. Gutenberg found an alloy that was easy to pour, cooled quickly, and gave letters of such hardness that they could be used for printing several different editions (Fig. 2a). In doing so, the material of the letter had to accept the ink well and then release it onto the printing substrate [3]. Therefore, the printing substrate, the paper had to be adapted to the printing process. Gutenberg also made an oil-based ink that easily transferred on the printing form and then easily transferred to a printing substrate on which it dried relatively quickly [4].

In 1450, Johann Gutenberg tested his machine by printing a Latin book on speech (Fig. 2b) [6]. This is followed by the printing of small dictionaries, short grammars, farewell letters, and calendars [5].



a)



b)

Figure 2 a) Bible printing form, Johann Gutenberg museum in Mainz [7], b) Reconstructed Guttenberg print shop with the printing press and movable types [8]

Gutenberg started printing the Bible around 1452. and completed it in late 1455. or perhaps early 1456 [7]. Some authors cite 1454. as the year of completion of the printing of the first Latin Gutenberg Bible. [9] Those years are considered the birth of printing [10].

Johann Gutenberg's most famous publication is the 42-line Bible (Fig. 3). The work was printed in two volumes with a total of 1282 pages [11, 12]. About twenty employees of the print shop worked on the production of this work. 290 different decorative figures were printed in the work. It is estimated that 180 copies were printed. About 150 of them on paper and about 30 on parchment. About 49 specimens have been preserved [13, 14].

a)

b)

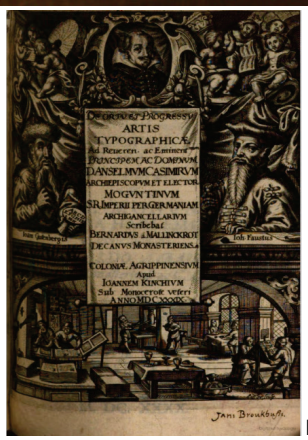
Figure 3 a) The first page of the Gutenberg Bible [8], b) Gutenberg Bible [13, 15]

The Psalter was finished in 1457. and is considered a second masterpiece by some authors[16]. The names of Faust and Schifer are mentioned for the first time as the followers of Johann Guttenberg in the Psalter. It is considered the most beautiful printed work by some authors. The decorations were so beautifully printed that some authors think they were printed in the technique of intaglio printing [17].

Books printed by Gutenberg did not have a special name until the celebration of the 200th anniversary of Gutenberg's invention, in 1639 when the dean of Münster Cathedral and the famous book collector Bernhard von Mallinckrodt published the pamphlet "On the Rise and Advancement of Graphic Art" (Fig. 4). He called the works printed between the time of Gutenberg's invention and the year 1500 "prima typographical incunabula - the first cradle of the press". With that, he defended the primacy of Johann Gutenberg as the inventor of book printing. Since then, the name incunabula has been used for works printed at that [7, 18].

Johann Gutenberg's invention caused a rise in print shop numbers and printed works in Europe. The greatest development was recorded in the area of today's Germany and Italy. It is believed that in each of those areas existed over 40 print shops around the 1480s [20]. The estimates are of 30000 to 40000 printed incunabula printed by the year 1500 [21]. The number of preserved editions is around 22000 copies. The main theme of published works was religion.

a)



b)

Figure 4 a) Bernhard von Mallinchrodt, b) Pamphlet mentioning the term "Incunabula" [7, 19]

2 METHODOLOGY

Research methods that include deductive and inductive methods were used in the research work [22]. A list of research frameworks was compiled by deductive methods, which was processed through a review of previous research with similar topics. The published data are carefully examined, which is the area of this research by reviewing and studying all the literature found based on scientific papers, impartially taking into account all the obtained search results. The applied inductive approach reduces the possibility of overlooking the given search frameworks, which contributes to optimization in expanding or creating new frameworks, which contributes to deepening the research topic and including all relevant data in the research [22]. The mentioned approach can contribute by grouping previously published data and recognizing the omitted data. This can simplify the approach to important facts in this scientific field.

All the above-mentioned literature quotations collected in the described manner have been read several times and provided. The results are presented depending on the geographical area (Europe or Croatia) within a specific period of fifty years (1450 to 1500). All technical characteristics are presented in turn, with additional emphasis on the comparison of different sources cited in the manuscript, to elaborate specific functional features of the framework and contribute to the inconsistency of the source

of information. Articles containing information on similar data were grouped and provided together to better identify the technical characteristics of printed books in the mentioned period.

3 RESULTS

3.1 Print Shop Development in Europe

Although some people opposed the machine replication of printed materials, the majority were supporters of this technological novelty allowing the easier distribution of knowledge [23]. Having a print shop also gave the ruler and the inhabitants of each country prestige (Fig. 5).

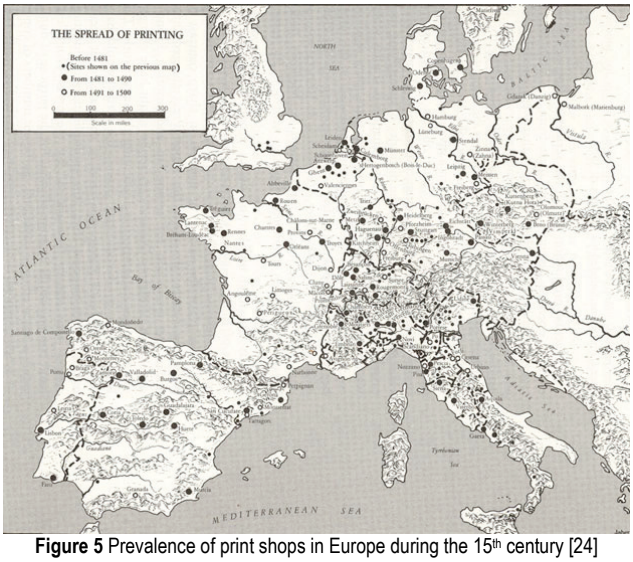


Figure 5 Prevalence of print shops in Europe during the 15th century [24]

The placement of print shops and their equipment in letterpress begins in the German city of Mainz around the year 1440. In Mainz Johann Fust and Peter Schöffer printed Psalterium Moguntium – A double sheet on 14 August 1457. Book was printed in black, red, and blue on vellum. Initials were printed in two- to three-line alternatively in red and blue color (Fig. 6).

Letterpress print shops appear in Köln 1465., Eltvillen 1467., Augsburg 68. – 72., Nürnberg 1470. and Ulm 1472. It is estimated that around 60 print shops existed around 1472. on the territory of today's Germany. In Augsburg, Günther Zainer printed Biblia Germanica between 1474 -1476 and Jodocus Pflanzmann in 1475. Zainer's third edition of the Bible in German has 533 leaves (Fig. 6.). Chapter headings and three-line initials are printed in red while rubricated, versals touched are printed in yellow. Bible has several woodcut Maiblumen initials, and 73 large woodcut initials in contemporary coloring. Pflanzmann printed the fourth German Bible in 456 leaves. Pflanzmann used several colors, rubricated and simple initials he printed in red, and Maiblumen initials in red, blue, green, and liquid gold. The Bible contained an abundance of 57 woodcut illustrations printed from 21 blocks colored by a contemporary hand (Fig. 6) [25].

Hanza built print shops in Lübeck, Rostok, Denmark, Sweeden, and at the beginning of the 16th century even in Russia. In meantime, a large number of letterpress specialized print shops appeared in Germany [26-28].

The first print shop appeared in Italy around the year 1462. or 1463. in the Subiaco Benedictine Abbey near Rome. The print shop was established by clergymen Conrad Sweynheym and Arnold Pannartz [29, 30]. The print shop relocated to Rome around the year 1467. The number of print shops grew continuously reaching around 40 print shops around the year 1500, with around 4500 printed titles. One of the reasons for such a development is due to the largest paper producer in Europe was in northern Italy.

Figure 6 a) Page in Psalterium Moguntium, b) Page in the third edition of Biblia Germanica, c) Page in the fourth edition of Biblia Germanica [25]

It is believed that the press was brought to Venice by Johannes de Spiro in 1469 [31]. Before that, he worked in a print shop in Mainz (1460/1461)[32]. Venice, a large commercial and cultural center, accepted a novelty, letterpress. As early as 1470, Nicholas Jensen perfected the letter by using fewer varieties and sizes of lettering. Then have been introduced major changes in the appearance of printed pages such as smaller, plainer, simpler, and more regular pages, fewer columns and blocks, less ornament, and less color: black and white, regularly spaced, and justified lines, a rectangular block [33, 34]. In 1495, the top graphic artist Aldus Manutius printed a book of supreme artistic value, Hypnerotomachia Poliphili, in 500 copies which was at that time a large edition [35]. In addition to the aforementioned book, the printer introduced innovations

such as the octave format for his books, while the manuscripts and incunabula had the (larger) quarto or folio format, replacing the wooden inlays in the book covers with cardboard, italics typefaces [36, 37] and other. The explosion of printing in Venice can also be seen in the fact that in 1500 it had about 150 printing presses for letterpress printing [26]. In this period are produced over four thousand editions, twice more than Paris and about one-seventh of the entire European production [38].

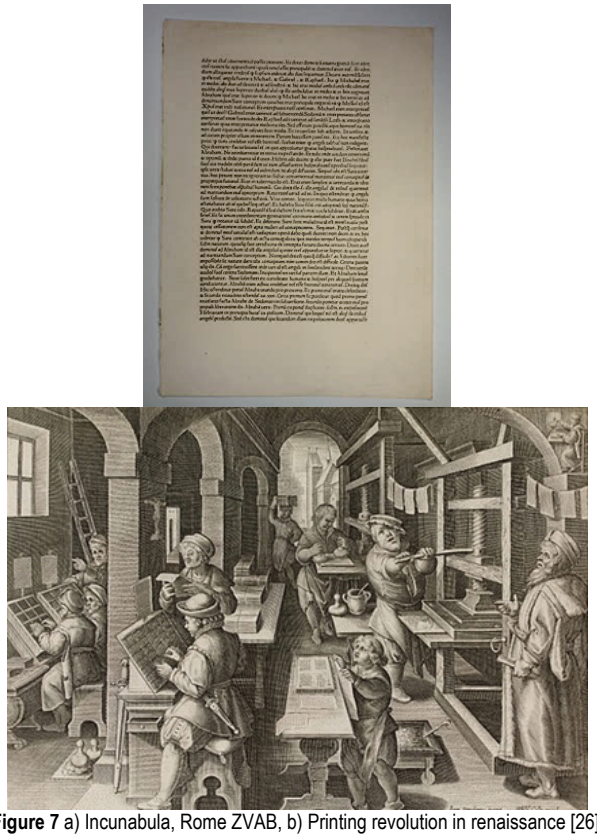


Figure 7 a) Incunabula, Rome ZVAB, b) Printing revolution in renaissance [26]

The first letterpress print shop appeared in England in 1476. The first printer in England was English William Caxton [39]. By occupation, William Caxton was a trader, diplomat, translator, and writer. This brought him to Köln where he learned the printing trade. Afterward, he started his first print shop in Bruges in Belgium, and gains the support of Burgandy dutches [40].

William Caxton returned to England in 1474 and started a print shop in Westminster (Fig. 8a). There he printed the first letterpress work in Great Britain (Fig. 8b) [27, 28]. He immediately expanded the number of titles and personally translated, typesetted, printed, and sold the editions. Of the 90 titles, he published and printed 74 were in the English language [26-28]. Unlike most print shops in Europe, he printed knightly secular novels, historical works, fables, and poems. William Caxton printed his translation of The Recuyell of the Histories of Troye a work by Raoul Le Fevre, and the first book printed in English in 1475 [41].

The first copies of Gutenberg's Bible were sold as manuscripts with aesthetically perfect black letters. Many people at the time viewed the press as black magic, and as a cheap, disgusting medium that could devalue the aesthetics of the beauty of handmade objects. All that led the printers to write elaborate prologues, advertising and describing the book. Caxton's prologues convey his personal credentials and efforts in the production [41, 42]. It should also be noted that Caxton was the first in England to use woodcarving to print ornaments and paintings in some of his works [39].

a)

b)

Figure 8 a) William Caxton in his print shop, b) the First book published in England [27, 28]

The laws in England did not support the rapid development of print shops. Thus, according to some sources at the end of the 15th century in England were only 5 print shops [26].

France as well as Italy had a well-developed intaglio printing at the time. However, letterpress and book printing came to France from Germany, via the city of Strasbourg, where it appeared in 1458. Guillaume Fichet [43] came to the Sorbonne and in the 1470s he founded a print shop at the Sorbonne University with his colleague Jean Heyulin. It was the first printing press in France, and it was primarily focused on the printing of humanitarian texts (Fig. 9). The first book was a collection of letters [44, 45]. In three years, the Paris print shop produced twenty-two books in Latin. It is believed that Fichet directed the press activities and Heynlin supervised production [46].

a)

b)

Figure 9 a) First print in Sorbonne in 1470, b) Drawing of the print shop from 1480 [44, 45].

Expensive books were printed in the beginning and after that, the production of cheap books began. This placed France at the top of book publishers of that time [26, 28]. Both founders left the Sorbonne in 1473, which was a blow to the increase in the number of books of this content [43]. It was not until ten years later that the printing of books continued in Paris, but then a nationally oriented humanism was established. In the meantime, printing expanded not only in France but throughout Europe and even a little wider [44].

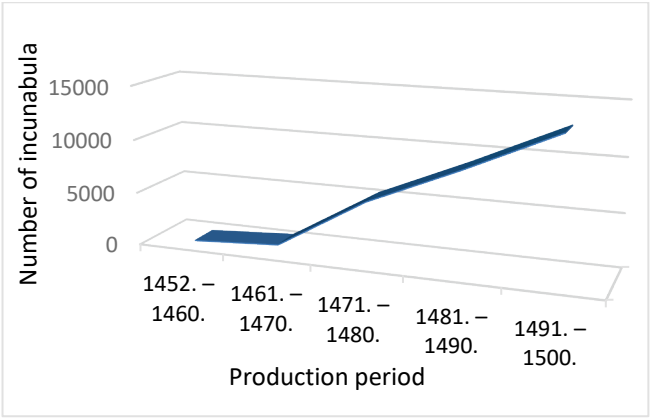


Figure 10 The rise of incunabula in comparison to the production period

The rest of Europe also held on to its reputation and wanted progress in its region [47]. Thus, print shops were opened in many countries, such as Utrecht (1470), Budapest (1473), Valencia (1473), and Krakow (1474). The print shops also appeared in other countries and cities as well (Fig. 10) [13]. In the cities of Germany and Europe, there has been an increase in book printing. It is interesting to see that the growth is not related to the distance from the place of the first print shop (Fig. 11) [23]. When print shops opened in individual cities, the increase in the number of printed books is related to the number of available works in Latin, the ability to write and translate into the language of a particular country, and the degree of civilization.

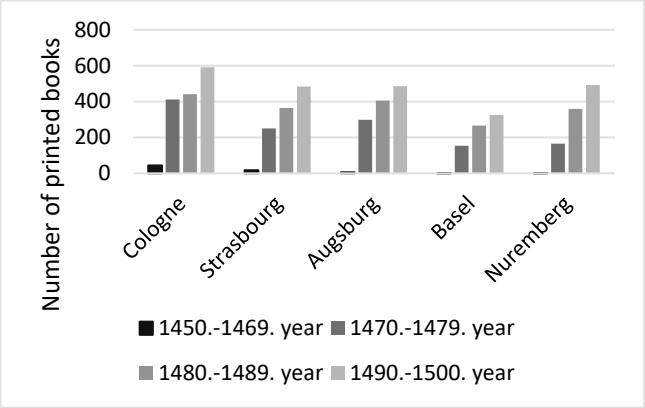


Figure 11 The number of printed books in German cities in the period from to [23]

Incunabula are preserved as cultural treasures all over the world. According to some sources, for example, there are about 19,700 copies in Munich, 12,500 in London, 12,000 in Paris, 8,600 in the Vatican, 8,000 in Vienna, 7,300 in St. Petersburg, and about 1,000 in Zagreb.

3.2 Incunabula in Croatia

The area inhabited by Slavic nations also prints its incunabula. The Check republic has the most published incunabula titles, around 60, then Croatia with 9, Ukraine with 5, and Montenegro with 5 titles. The peculiarity of Croatian incunabula is in the fact that they were mostly printed in Glagolitic, and a smaller number in Latin. Of all the mentioned nations, the Croats, with their diversity in writing, developed a special form of the Glagolitic alphabet, the so-called Croatian angular form. The mentioned alphabet is used both in writing and in the press [48].

The oldest Croatian incunabula is named Missal according to the law of the Roman court, printed in 1483 in Glagolitic script (Fig. 12). This work is the first incunabula printed by all South Slavic peoples and the first missal in Europe not printed in Latin letters and language [40]. No record has been found of where the work was printed, but some authors assume it was printed in Venice [49]. There are scientific discussions about the place of printing that have not yet been harmonized. The Missal is printed on 438 pages, arranged in two columns with 36 works. It is printed in two colors, red and black. 201 typographic characters (uppercase or lowercase letters, initials, ie initials, ligatures, abbreviations, etc.) were used in printing [50]. Most of it is printed on white cotton fiber paper, and a smaller is printed on parchment. The book is printed in two colors, red and black. In terms of typographic, esthetic, and artistic features,

this work is not surpassed by any later Glagolitic edition. 11 specimens have been preserved and none are complete [51]. Only 6 of them are kept in Croatia.

Figure 12 Missal according to the law of the Roman court a) first page, b) inner pages, c) tome

Thanks to the good connection with the Venetians, the first Latin breviary in Zagreb was printed in the print shop of Erhard Radtolt in Venice in 1484 (Fig. 13) [39]. This printed breviary of the Zagreb church was ordered by the Zagreb Bishop Osvald [52]. The typography breviary is unpretentious, containing only two small woodcuts and one decorated woodcut initial letter at the colophon. Two known preserved copies are kept in the Vatican library and the National Library in Budapest [53]. Missals have a few small and several big calligraphic initials in red, blue, or black ink with gilt layer and a printed Crucifixion miniature in front of the mass canon text [52].

The prayer book from 1490 was printed in Latin. The book contains an office in honor of the Virgin Mary and seven obedient psalms. In the same year was printed Office in Latin. The Office consists of two parts, the Office of the Holy Cross, and the Office of the Holy Spirit. The printing place for both books is unknown, and one copy has been preserved. Preserved specimens are bound together and kept in the Vatican libraries.

According to Z. Kulundžić and other scientists, the Glagolitic Breviary According to the law of the Roman court was printed in Kosinj in 1491. The year of printing the Breviary was determined based on a calendar because the only surviving copy is missing the last quaternion with the colophon where the data on the print shop, the date of printing, and other information would be found [55-57]. The copy is kept outside Croatia, in the Vatican. The preserved copy has VIII + 380 sheets, while the text is printed in two columns of 38 lines in two colors: black and red. A copy is printed on paper and a fragment of the calendar is on parchment [54].

Figure 13 a) Zagreb Missal a Crucifixion miniature [52] b) The Glagolitic Breviary, printed in 1491 [54].

The Italian Peregrinus de Pasqualibus printed the Glagolitic incunabula, the Confession (*Ispovid*) or Treatise on Confession (Ispovid ku e vsaki krstjanin drţan imiti i umiti i naučiti). It was compiled by Matej Bošnjak from Zadar. The Incunabula was printed on June 16, 1492, in Venice. Not a single copy has been preserved, evidenced of the existence is evidenced by a manuscript in the Tkon's proceeding (Tkonskom zborniku) [58, 59].

Figure 14 a) a bust of Blaž Baromić, b) Reconstruction of the printing press in Senj c) Senj Glagolitic Missal from 1494 [64]

In 1493, Blaž Baromić printed the Breviary in Venice in the print shop of Andreas Torresani. The breviary is a small format and consists of 544 leaves [60, 61]. The typographic peculiarity of this breviary was ligatures, especially the ones "broken". They were made in the so-called Baromić technique of broken ligatures or Baromić's technique of stacking ligatures [62, 63]. Five copies were saved of which only two are in Croatia.

Blaž Baromić, Silvestar Bedričić, and Gašpar Turčić printed the Senj Missal in Glagolitic script in Senj in 1494 (Fig. 14) [64]. One complete and two incomplete copies have been preserved, of which only one incomplete copy is kept in Croatia. Preserved fragment of the Missal consists of three leaves and has a woodcut with a crucifix (canonical image), which makes the Missal the first Glagolitic book with a woodcut [65, 66]. The missal consists of 216 leaves with two columns of 37 rows each. It is printed in two colors. Frane Paro emphasizes the original technique of casting half-letters, which enables the arrangement of ligature letter groups, further analyzes the details of the typographic process of printing the book, and finds that Baromić's typographic set created for the print shop in Senj is the best of Croatian Glagolitic printing [67].

Bernad Spličanin's Lekcioner was printed in Latin script in Venice in 1495. The book was printed in the print shop of master Damian of Milan (Damianus de Gorgonzola). The Lekcionar has eight volumes and 208 pages, it is printed in two colors. Two complete copies and one fragment have been preserved. One copy is in Odessa, in the Maksim Gorki Library, and the other in Zagreb in the Juraj Habdelić Library in the library of the Zagreb Jesuit College. The fragment is in Zagreb with the Franciscans at the Zagreb Kaptol [68, 69].

Figure 15 Colophon from Spovid općina printed in 1496 [60].

Spovid općina was printed in Glagolitic in Senj in 1496. It was a translation of manual Confessionale generale (Fig. 15) [70]. Confessionale generale was a popular work of the period and was written by Michael Carcano [71]. Blaž Baromić was mentioned as a printer and referred to as priest Blaž. Spovid općina was printed on 40 sheets, in one column with 25 lines. It is printed in one color, black. One specimen has been preserved in Zagreb in Croatia [24].

The development of print shops in northern parts of Croatia is related to the period after 1500, i.e., to the time after the printing of the incunabula.

4 TECHNICAL CHARACTERISTICS OF EUROPEAN INCUNABULUM

The design of the matrix enabled the technology for making type or individual pieces of metal tiles of alloy lead and tin in large series. The matrix was formed by engraving letters in harder metal and fastening them in softer metal. Large batches of type have contributed to a sufficient inventory to stack one page, which contributes even out the print. In this way, it was possible to make the desired number of copies in a short time, after which the types could be used to compose a new page.

According to all the above write out, it can be concluded that technical characteristics of the printing materials in 15 centuries, such as paper or ink, are similar in all incunabula. Inks were prepared for each edition, which could cause small variations in composition, but were not significant. Inks are made from raw materials available in nature or the environment, such as mineral, plant, or animal origin. The mineral used in red inks at the time is mercury vermilion. The white (hydro cerussite), yellow (massicot and lead-tin oxide), and orange (lead tetroxide) inks were based on lead. Blue could have been produced from raw materials from the plant origin as indigo, but from the mineral azure. The earliest recipes are designed specifically for printing and are similar in composition to carbon-based printing ink. Another source is dyes of animal origin such as carbon black obtained from raw materials of animal origin. Soot or lampblack pigments were commonly added to iron gall, which contributed to the good coating properties of the dye, but also its durability [72]. The good durability of natural pigments contributed to the preservation of books of that time, which testify to the level of technological development of that time and the inventiveness of the first printers.

Paper was known in the world and Europe before the invention of printing, but paper production grew rapidly with the invention of the printing press. The main raw material for making paper was used fabrics, especially tulip cloth. The good preservation of paper used in the research period is influenced by the optimal ratio of calcium and potassium. The high calcium content of the papermaking recipe is related to the use of gelatin, while the low amount of potassium binds to the amounts used by Alum [73].

The paper used by Guttenberg for the printing of the Bible was made at Caselle in Piedmont. That is visible from the watermark embedded in the used paper. Watermarks had the symbol of an ox's head (70%), a bunch of grapes (20%), and a walking ox (10%). The sheet format common to all printed matter is called the royal foil, which is about 430 × 620 mm.

5 CONCLUSION

One of the greatest inventions of mankind is the invention of the movable type and printing by Johann Gutenberg. This technology has enabled the rapid development of science, culture, and the economy.

Collecting and compiling religious texts, and translations into national languages contributed to greater accessibility and increased the use of printed books. In this way, all the most important literary works of human civilization become available to the public. Books greatly contribute to getting to know the world and people, which causes the start of a great wave of civilization. In addition to all the above, books are a source of information about the way of life at that time.

In the beginning, there was resistance to the new way of publishing, but soon the number of editions was taken as an unwritten parameter of the development and civilization of a country. In particular, politics in the late 15th century recognized the marketability of language in the press. The measure, the right measure, and the currency fit into it. That is why kings Louis XI (1461-1483) and Charles VIII (1490) strengthened printing by special decrees.

Croatian printers were the first to participate in the printing of books abroad. They brought the acquired knowledge and skills to Croatia, where they expanded their knowledge and introduced innovations. Blaž Baromić published the first book in Croatia and encouraged the reading of technological achievements in that area. From all the above, it is evident that Croatia actively participated in this early civilization process and thus actively contributed to the spread of new knowledge both in Croatia and in the world.

Despite the great distances from the place where the first print was created in Guttenberg's printing house and the poor traffic connection in the 15th century, the new technology spread rapidly throughout Europe. This is reflected in the similarity of technological processes related to the production of paints, paper and printing. Regardless of the above, creative individuals bring their innovative ideas and improve the development of technology, so Caxton, for example, printed ornaments and pictures with woodcarving, and Baromić was the only printer who made "broken" ligatures.

6 REFERENCES

[1] Minguzzi, S. (2012). *The Printing Revolution: Johannes Gutenberg*. https://www.academia.edu/6842453/Gutenberg

[2] Forrester, R. (2020). History of printing-From Gutenberg to the Laser Printer, Available SSRN 3512249. https://doi.org/10.2139/ssrn.3512249

[3] Schwab, R. N., Cahill, T. A., Kusko, B. H., & Wick, D. L. (1983). Cyclotron analysis of the ink in the 42-line Bible. *Pap. Bibliogr. Soc. Am., 77*(3), 285-315. https://doi.org/10.1086/pbsa.77.3.24302918

[4] Kümmel, A., Scholz, L., Schumacher, E., & Kümmel-Schnur, A., Eds. (2004). Die Industria des Buchdrucks. Eine Diskursgeschichte der Vervielfältigung. *Einführung in die Geschichte der Medien*, Paderborn: Fink, 11-33. (in German)

[5] … (2018). Johannes Gutenberg. *History and Biography*. https://history-biography.com/johannes-gutenberg/ (accessed Feb. 01, 2022).

[6] … LEMELSON-MIT. *Johann Gutenberg - Movable Type Printing Press*. https://lemelson.mit.edu/resources/johann-gutenberg (accessed Feb. 01, 2022).

[7] … Die Welt der Habsburger. *Der Anfang der Gutenberg-Bibel, Teil 1: Altes Testament, Vorrede des Hieronymus, 1452-1454*. https://www.habsburger.net/de/medien/der-anfang-der-gutenberg-bibel-teil-1-altes-testament-vorrede-des-hieronymus-1452-1454 (accessed Feb. 01, 2022). (in German)

[8] … die schwarzdrucker. *Johannes Gutenberg - der Erfinder des Buchdrucks*. https://www.dieschwarzdrucker.com/johannes-gutenberg-der-erfinder-des-buchdrucks/ (accessed Feb. 01, 2022). (in German)

[9] Schwab, R. N., Cahill, T. A., Eldred, R. A., Kusko, B. H., & Wick, D. L. (1985). New evidence on the printing of the Gutenberg bible: The inks in the Doheny copy. *Pap. Bibliogr. Soc. Am., 79*(3), 375-410. https://doi.org/10.1086/pbsa.79.3.24303664

[10] Rebić, A. (2003). Gutenbergova Biblija. *Bogosl. Smotra, 73*(1), 225-232. (in Croatian)

[11] Eisenstein, E. L. (1980). *The Printing Press as an Agent of Change*, 1st ed. Cambridge University Press. https://doi.org/10.1017/CBO9781107049963

[12] Füssel, S. (2001). Gutenberg and today's media change. Publ. Res. Q., 16(4), 3-10. https://doi.org/10.1007/s12109-001-0010-7

[13] … *Die 42zeilige Bibel*. http://www.gutenberg.de/bibel/index.php (accessed Feb. 01, 2022).

[14] Lehmann-Haupt, H. E. (2022). Johannes Gutenberg. *Encyclopedia* Britannica. https://www.britannica.com/biography/Johannes-Gutenberg (accessed Feb. 01, 2022).

[15] … *Early printed books*. The British Library. https://www.bl.uk/collection-guides/early-printed-books (accessed Feb. 01, 2022).

[16] Andrés-Sanz, M.-A. (2018). Psalms and Psalters in the Manuscript Fragments Preserved in the Abbey Library of Sankt Gallen. *Fragmentology*, 1, 39-63, https://doi.org/10.24446/ugx4

[17] Stewart, A. (2013). The Birth of Mass Media: Printmaking in Early Modern Europe. Published in *A Companion to Renaissance and Baroque Art*, First Edition, edited by Babette Bohn and James M. Saslow (Chichester: John Wiley & Sons, 2013), 253-273. https://doi.org/10.1002/9781118391488.ch12

[18] … *Bernhard von Mallinckrodt Coins the Term Incunabula: History of Information*. https://historyofinformation.com/detail.php?id=1581 (accessed Feb. 01, 2022).

[19] … *Bernhard von Mallinckrodt*. HowOld.co. https://www.howoldo.co/person/bernhard-von-mallinckrodt (accessed Feb. 01, 2022).

[20] Febvre, L. & Martin, H.-J. (1997). *The coming of the book: the impact of printing 1450-1800*, vol. 10. Verso, 1997.

[21] Farias, P. L., Hanns, D. K., & Dixon, C. (2016). Spatializing design history: Considerations on the use of maps for studies on print culture. *Proceedings of the 10th International Conference on Design History and Design Studies*, Taipei, Taiwan, 348-353. https://doi.org/10.5151/despro-icdhs2016-04_009

[22] Matthes, J. & Kohring, M. (2008). The Content Analysis of Media Frames: Toward Improving Reliability and Validity. *J. Commun., 58*(2), 258-279. https://doi.org/10.1111/j.1460-2466.2008.00384.x

[23] Rubin, J. (2014). Printing and Protestants: An Empirical Test of the Role of Printing in the Reformation. *Rev. Econ. Stat., 96*(2), 270-286. https://doi.org/10.1162/REST_a_00368

[24] … *Der Buchdruck und die Normierung*. https://home.uni-leipzig.de/burr/Intro/html/Gutenberg_Aera_Folien.htm (accessed Jan. 31, 2022). (in German)

[25] Hanke, M. & Nettekoven, I. (2016). Early printed Bibles: 1454-1580. Stalden: J. Günther Antiquariat.

[26] Unwin, G., Unwin, P. S., & Tucker, D. H. 'history of publishing'. https://www.britannica.com/topic/publishing (accessed Feb. 01, 2022).

[27] 'Erfindung des Buchdrucks', Geschichte kompakt. https://www.geschichte-abitur.de/fruehe-neuzeit/erfindung-des-buchdrucks (accessed Feb. 01, 2022). (in German)
[28] Bolanča, S. (2021). *Visoki tisak*. Varaždin: Sveučilište Sjever. (in Croatian) http://library.foi.hr/lib/knjiga.php?B=20&sqlx=100670&H=
[29] Boldan, K. (2022). Lamentatio Nigropontis: An Unknown Print of Sixtus Riessinger, a Prototypographer of Naples. *Acta Musei Natl. Pragae – Hist. Litt., 67*(1-2), 5-12. https://doi.org/10.37520/amnpsc.2022.002
[30] Bolick, L. (2014). Culture, humanism and intellect: Cardinal Bessarion as patron of the arts. https://doi.org/10.21954/OU.RO.0000EEEF
[31] Esih, I. (1970). 5. *The Development of Book-Printing 1483-1940 in Croatia*, F. Eterovich and C. Spalatin, Eds. University of Toronto Press, 251-298. https://doi.org/10.3138/9781487595067-009
[32] Der Buchdruck und die Normierung. https://home.uni-leipzig.de/burr/Intro/html/Gutenberg_Aera_Folien.htm (accessed: Jan. 31, 2022) (in German)
[33] Carlson, D. R. (2004). 4. Nicholas Jenson and the Form of the Renaissance Printed Page. *The Future of the Page*, P. Stoicheff and A. Taylor, Eds. University of Toronto Press, 91-110. https://doi.org/10.3138/9781442657250-006
[34] Brown, H. R. F. (1891). *The Venetian printing press*. John C. Nimmo.
[35] Gibbs, M. L. (1976). Aldus Manutius as Printer of Illustrated Books. Princet. *Univ. Libr. Chron., 37*(2), p. 109. https://doi.org/10.2307/26404001
[36] Meggs, P. B. & Purvis, A. W. (2016). *Meggs' history of graphic design*. John Wiley & Sons.
[37] Davies, M. & Manuzio, A. (1995). *Aldus Manutius: printer and publisher of Renaissance Venice*. British Library London.
[38] Crisci, F. (2017). *The World of Aldus Manutius (1494-1515), a Renaissance Publishing Venture: Materiality, Cultural Entrepreneurship, and Institutional Dynamics in Markets*. https://doi.org/10.7433/SRECP.EA.2017.58
[39] 'Wiliam Caxton - The British Library'. https://www.bl.uk/people/william-caxton (accessed Feb. 01, 2022).
[40] 'William Caxton and the introduction of printing to England - The British Library'. https://www.bl.uk/medieval-literature/articles/william-caxton-and-the-introduction-of-printing-to-england (accessed Feb. 01, 2022).
[41] Lemke, A. B. (1978). William Caxton—The Beginning of Printing in England. *The Courier, 15*(1), 3-13
[42] Blades, W. (1882). *The biography and typography of William Caxton: England's first printer*. Ludgate Hill.
[43] Mureşan, D. I. (2017). Bessarion's Orations against the Turks and Crusade Propaganda at the Große Christentag of Regensburg (1471). *Reconfiguring the Fifteenth-Century Crusade*, Springer, 207-243. https://doi.org/10.1057/978-1-137-46281-7_7
[44] Rüegg, W. & Briggs, A. (1993). Geschichte der Universität in Europa: Mittelalter. Beck, C. H. https://books.google.hr/books?id=QXXRAQAACAAJ (in German)
[45] 'La diffusion de l'imprimerie hors d'Europe.' https://www.cosmovisions.com/civImprimerieMonde.htm (accessed Feb. 01, 2022). (in France)
[46] Booton, D. E. (2009). Guillaume Fichet's Literary Gift to Duke François II of Brittany. *Nottm. Mediev. Stud., 53*, 121-132. https://doi.org/10.1484/J.NMS.3.463
[47] Clair, C. (1976). *A history of European printing*. London; New York: Academic Press.
[48] Uspenskij, B. (2013). Glagolitic script as a manifestation of sacred knowledge. *Studi Slavistici, 10*, 7-27.
[49] Badalić, J. (1952). *Inkunabule u Narodnoj Republici Hrvatskoj: Djela Jugoslavenske Akademije Znanosti i Umjetnosti*. (in Croatian)
[50] Škrbec, S. (1995). *Riječka zvijezda Gutenbergove galaksije: tiskarstvo Rijeke i Hrvatske glagoljske tiskare*. Tiskara Rijeka. https://books.google.hr/books?id=1FPsMQAACAAJ (in Croatian)
[51] Seiter-Šverko, D. (2012). Nacionalni program digitalizacije arhivske, knjižnične i muzejske građe i projekt "Hrvatska kulturna baština", *Vjesn. Bibl. Hrvat., 55*(2), 5-15.
[52] Pelc, M., Leustek, M., Bacic, Z., Cizmek, D., Tomljenovic, G. & Nacionalna i sveucilisna biblioteka (Zagreb). (2018). *Book art in Croatia*. In Zagreb: National and University Library.
[53] Pelc, M. (2017). Ius supremi patronatus regis and the Hungarian Holy Rulers in the Liturgical Books of Zagreb's Bishops around 1500. *Rad. Instituta za Povij. Umjet.*, 41, 7-22. https://doi.org/10.31664/ripu.2017.41.01
[54] Nazor, A. (1993). Croatian Glagolitic Incunabula on the Five Hundredth Anniversary of the Breviary of Blaž Baromić (1493-1993). *Croat. Časopis za Hrvat. Jez. Književ. Kult., 23*(37-38-39), 229-257.
[55] Derossi, J. (1992). Problematika hrvatsko glagoljskoga brevijara iz godine 1491. *Senj. Zb.-Pril. Za Geogr. Etnologiju Gospod. Povij. Kult., 19*(1), 117-124. (in Croatian)
[56] 'Kosinjska tiskara | Hrvatska enciklopedija'. https://www.enciklopedija.hr/Natuknica.aspx?ID=33324 (accessed Jun. 20, 2022). (in Croatian)
[57] Nazor, A. (1984). Tiskana glagoljska knjiga od prvotiska Misala 1483. do Brozićeva brevijara 1561. *Slovo, 34*, 7-16. (in Croatian)
[58] *Senj - inkunabulistika*. http://www.croatianhistory.net/etf/senj3.html (accessed Mar. 17, 2022). (in Croatian)
[59] Bošnjak, M. (1970). Slavenska inkunabulistika. Vlastita naklada. (in Croatian)
[60] Nazor, A. (2014). O glagoljskoj tiskari u Senju i njezinim izdanjima (1494.-1508.). *Senj. Zb.-Pril. za Geogr. Etnologiju Gospod. Povij. Kult., 41*(1), 211-244. (in Croatian)
[61] Bakmaz, I. (1998). *Blaž Baromić as scibe and as printer*. 9-13.
[62] Žagar, M. (2009). Anica Nazor, Knjiga o hrvatskoj glagoljici: "Ja slovo znajući govorim…". Zagreb 2008. *Slovo*, (59), 375-384. Downloaded from https://hrcak.srce.hr/48055 (in Croatian)
[63] Grabar, B. (1984). Tiskani glagoljski Baromićev brevijar. *Slovo, 34*, 159-180. (in Croatian)
[64] *Senjski glagoljski misal 1494*. https://glagoljica.hr/?pr=i&id=22254 (accessed Feb. 01, 2022). (in Croatian)
[65] Čunčić, M. (2015). Marija Agnezija Pantelić, Hrvatsko glagoljsko srednjovjekovlje. Priredio Petar Bašić. Kršćanska sadašnjost doo, Družba sestara milosrdnica svetog Vinka Paulskog u Zagrebu, Zagreb 2013., 725 str.. *Slovo*, 65, 192-205. (in Croatian)
[66] Paro, F. ed. (1994). Neke osobitosti tipografije Senjskoga misala iz 1494. Godine. *Senjski glagoljski misal 1494*, Faks. pretisak., Zagreb: Hrvatska Akademija Znanosti i Umjetnosti, 19-22. (in Croatian)
[67] Dürrigl, M.-A. (1996). Senjski glagoljski misal 1494., faksimilni pretisak. *Slovo*, 44-45-46, 246-248. (in Croatian)
[68] Škunca, S. J. (1995). Pitanje identiteta fra Bernardina Splićanina, izdavača hrvatskog lekcionara (1495). *Croat. Christ. Period., 19*(369, 69-76.
[69] Bratulić, J. (1991). Lekcionar Bernardina Splićanina, 1495. vol. 6. Knjizevni Krug.

[70] Zubčić, S. (2021). Spovid Općena and Kožičić's Zrcalo svestnoe. *Filologija, 77*, 225-242. https://doi.org/10.21857/mwo1vcj4xy
[71] Nazor, A. (1981). Senjska Spovid općena kao izvor Akademijina Rječnika hrvatskoga ili srpskoga jezika. *Senj. Zb. Pril. za Geogr. Etnologiju Gospod. Povij. Kult., 8*(1), 369-376.
[72] Mommsen, H. et al. (1996). X-ray fluorescence analysis with synchrotron radiation on the inks and papers of incunabula. *Archaeometry, 38*(2) 347-357. https://doi.org/10.1111/j.1475-4754.1996.tb00782.x
[73] Albertin, F., Balliana, E., Pizzol, G., Colavizza, G., Zendri, E., & Raines, D. (2018). Printing materials and technologies in the 15th–17th century book production: An undervalued research field. *Microchem. J., 138*, 147-153. https://doi.org/10.1016/j.microc.2017.12.010

Authors' contacts:

Stanislav Bolanča, Prof. PhD
Croatian Academy of Engineering
Kačićeva ulica 28, 10 000 Zagreb, Croatia
+385 4922 559, stanislav.bolanca@hatz.hr

Ivana Bolanča Mirković, Prof. PhD
University of Zagreb Faculty of Graphic Arts
Getaldićeva 2, 10 000 Zagreb, Croatia
+385 2371 080, ivana.bolanca.mirkovic@grf.unizg.hr

Ivan Pučić, Postdoctoral researcher PhD
University of Zagreb Faculty of Graphic Arts
Getaldićeva 2, 10 000 Zagreb, Croatia
+385 2371 080, ivan.pucic@grf.unizg.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221208095644

Subject review
Received: 2022-12-08, Accepted: 2023-01-12

Benefits of Increased Railway Safety and Reliability and their Evaluation

Vít Hromádka*, Jana Korytárová, Eva Vítková, Herbert Seelmann, Tomáš Funk

Abstract: The paper focuses on the presentation of key results of the project of applied research oriented on evaluation of the increased safety and reliability of the railway network due to the implementation of relevant railway network projects and their projection into the investment project socio-economic evaluation using Cost-Benefit Analysis. The subject research described in the article is the design and verification of the functionality of the methodological procedure for the evaluation of these benefits for their further use in the assessment of public projects on railways. The methodological part summarizes the basic outputs of the previous research. These outputs were subsequently applied at the formulation of basic rules for the assessment of benefits connected with increased railway reliability and safety, which were later processed into the form of simplified methodological procedure. The simplified approach uses average data on the occurrence impact for the entire railway network, broken down into line and station sections. The functionality of individual approaches was checked on case studies of actually implemented projects.

Keywords: benefits; economic efficiency; occurrences; railways; safety and reliability; social impacts

1 INTRODUCTION

The article is oriented on the introducing of the final outputs of the research activities oriented on the evaluating the benefits connected with increased user level of the railways as an output of the carrying out of appropriate measures. The key output is a methodological procedure for evaluating the benefits connected with increased safety and reliability level of the railways designed to complement the socio-economic analysis of construction projects in the field of railways. This paper presents the key principles, approaches, methods and values which form the basis of the methodological approach. The contribution is based on the outputs of previous already published articles [1-4]. These results are reached using an extensive database of extraordinary events [5], which involves nearly 6,000 occurrences from 2011-2018 arising in the Czechia and represents a key data set for the subsequent analysis and final synthesis of information into the final version of the methodological procedure for the socio-economic impacts of classified occurrences evaluation. In addition to the above-mentioned Database of Occurrences, statistical data on the Czech railway network [6-9] were used as inputs. The presented paper presents the final unit impact values of sub-categories of occurrences related to the year and transport performance unit (train kilometre) usable within the simplified methodological approach. The functionality of the proposed methodological procedure was consequently checked on two railway infrastructure projects oriented on the modernization divided into a line and station sections.

The benefits described above will be used for the socio-economic evaluation of rail-way infrastructure projects using Cost-Benefit Analysis which is explained in many scientific texts. The Guide to CBA [10] can be considered a basis for the evaluation of transport infrastructure project at the European level. The Departmental Guideline of the Ministry of Transport [11], which developed an approach for the economic and financial assessment of projects of road transport, rail transport and transport-significant water structures is used for the evaluation of transport infrastructure projects at the national level of the Czech Republic. The Guidelines include a comprehensive socio-economic evaluation of these types of projects, including an evaluation of external impacts. by the way, the issue of the impact connected with increased railway user level resulting from the implementation of the projects increasing the level of safety equipment was not ad-dressed by the Guidelines. From the aspect of the methodology, attention should be paid to an extensive study for the evaluation of the Rail Baltica railway corridor [12], which carried out a comprehensive economic evaluation including an assessment of safety change due to the transfer of the part of road transport to the railway. However, the evaluation of the increased transport safety and reliability due to the implementation of modern safety equipment was also not considered in it.

The evaluation of safety and its development in the field of railways transport in re-cent years represents an important issue. The document "Report on Railway Safety and Interoperability in the EU 2020" [13] presents an important source of information regarding this area. The following conclusions were drawn within the field of railway safety in the EU:

- The safety level in terms of the fatal accident rate has been continuously improving since 1990, with an average annual reduction of more than 5%.
- The safety level monitored for 2018 was historically the highest (overall decrease in serious accidents since 2010).
- The "internal" accident (collisions, derailments and fires in rolling stock) level is more or less stagnant.

The material [13] presents a development in the number of serious injuries and deaths due to occurrences across all European Union countries. This trend can be associated with the presented continuous increase in the railway safety level in the European Union.

As already mentioned, a key approach to considering the positive impacts connected with increased reliability ad safety in railway network is the Cost-Benefit Analysis (CBA).

The CBA is generally one of the key tools for the economic evaluation of railway infrastructure projects. The issue of the ex-ante and ex-post evaluation of projects increasing the railway network safety level is addressed, for

example, by a paper [14]. Olsson et al. [15] deal with the utilization of the Cost-Benefit Analysis in tranport infrastructure projects at the international level and focus in their article on the comparison of the approaches to economic evaluation between selected countries. Siciliano et al. [16] apply the CBA approach to innovative transport services and evaluation of the effectiveness of their use. Carteni & Henke [17] provide a practical example of using CBA analysis to evaluate the economic benefits of a tourist railway implementation in northern Italy. A very important area to be presented in connection with the presented research is the issue of railway safety assessment. Evans [18] deals with the economic context associated with the railway network and its safety. Oertli [19] deals with the evaluation of the impacts caused by railway noise and presents the possibilities of inclusion the impacts connected with the reduction of level of noise into the CBA. The issue of time, especially train delays, is the subject of the paper by Tseng et al. [20] who explain the impacts of train unreliability and subsequent changes in timetables on the economic evaluation carried out in the form of CBA. Johnsen [21] and Petrova [22] are interested with the risk management in rail transport and railway infra-structure in their articles.

Evans [23] in his study deals with a detailed analysis of serious railway accidents throughout Europe in the 1980-2009 period. The procedures of systematic analysis of railway occurrences are introduced in the paper by [24]. In their contribution, Klockner and Toft [25] address a comprehensive socio-economic assessment of railway emergencies, Read et al. [26] focus on the causes of occurrence emergence on the railway network, the main attention is paid to the social environment and its impact on the human factor. Prevention of the arising of extraordinary events on the railway and its impact on their emergence is the subject of the paper [27].

Even extraordinary events as a result of human factor or technical mistake are considered in the research project, extraordinary events in the form of suicides or accidents at level crossings in relation to the previous research ([1], [2]) were excluded from the research. Understanding the impact of people on the arising of extraordinary events on the railway in Australia were addressed by Baysari et al. [28], a similar issue, however, from the Indonesian environment, was addressed by Iridiastadi and Ikatrinasari [29]. The use of Human Factor Analysis for the classification and assessment of railway safety in the United Kingdom was presented in the article [30]. The model for assessing the probability of a railway occurrence emergence due to the human factor was presented by Lombardi et al. [31]. The impact of engine drivers' cognitive abilities on the potential occurrence emergence was the subject of the text by Hani Tabai et al. [32].

2 MATERIALS AND METHODS

The main objective of the presented research is the formulation of the methodological approach oriented on considering the positive impacts connected with increased user standard of the railway caused by the safety equipment consideration in the economic evaluation of transport infrastructure projects. The key input for the methodological approach is the unit economic impact of the extraordinary event per appropriate measure unit and the year.

2.1 Relevant Occurrences

The expected impact of the implementation of measures to increase the safety and reliability of the railway network lies in the reduction of the number of occurrences emerging on the railway network, i.e. on the wide line and in the station.

The methodological procedure considers the specific categories of occurrences, which can be avoided or decreased with the realization of specific projects. From this reason, this approach doesn´t consider occurrences emerging at level crossings and the occurrences like suicides. While suicides cannot be avoided by any safety measure, in the case of railway crossing the possibility to avoid exists, e.g. by improvement of safety measures. But these kinds of projects have been already solved by specific methodological approach. The occurrence structure is based on the occurrence classification used in the Data-base of Occurrences [5] and divides occurrences into

- A Serious accidents,
- B Minor accidents,
- C Incidents.

These categories can be divided into subcategories A1-A3 (accidents with impact on health of 5 persons or with big loss), B1-B3 (lower impacts then in the case of serious accidents) and C1-C19 (small impact).

Detailed information on occurrence subcategories is available in the papers [1, 3].

2.2 Determination of Benefits of Extraordinary Events

Within the relevant categories of occurrences there were determined the overall socio-economic occurrence impacts, like impacts on health, losses on property or costs caused by delays of passenger or cargo trains. For the evaluation of these benefits and costs the principles of the CBA analysis were used (e.g. the financial valuation of impact of accidents on health, or the financial valuation of time delays due the accidents – people and freight). The theoretical background of the calculation is explained in the Departmental Guidelines [11].

Table 1 Values of impacts per one extraordinary event according to category (CA 2018)

EE Cat.	Values of impact (€)
A1	913,964
A2	847,877
A3	2,415,941
B1	130,822
B2	90,538
B3	78,651
C1	7,473
C2	5,446
C3	7,043
C6	3,370
C12	4,479
C19	9,388

The detailed principle of the socioeconomic impact calculation is presented in the paper [1], the average total economic impacts per one occurrence caused by a human factor according to the above-listed categories are shown in the papers [2, 3]. Final result of the determination of unit impact of extraordinary events of specific categories caused by human or any other factor is presented in the Tab. 1.

2.3 Costs and Benefits of Extraordinary Events on Wide Lines and in Railway Stations

Costs related of sub-categories of Extraordinary Events must be related to the railway, the wide lines and stations. From the Database of Extraordinary Events [5] it was possible to derive that 94 % of them appear in stations and 5,15 % of them appear on broad lines. Detailed calculation has been already presented in the paper [3].

The size of the railway network is defined in train kilometres for both, train kilometres on wide line and train kilometres in railway stations.

Information on average annual transport performance on the railway network is used for further calculations. The size of the network is given in Tab. 2.

Table 2 The size of the railway transport in the Czech Republic

Input quantity	Value
Number of train kilometres in railway stations (2018)	27.025 mil. tkm/year
Number of train kilometres on wide track (2018)	144.932 mil. tkm/year
The proportion of extraordinary events at the railway station	94.85%
The proportion of extraordinary events on the wide line	5.15%

To calculate the annual costs of the partial categories of extraordinary events, the key relationship for the calculation was defined in the following form:

$$TI_{ik} = \frac{O_i \times R_k}{Q_k \times t} \times UI_i \quad (1)$$

Where:

- TI_{ik} Total annual cost (impact) of extraordinary event of sub-category i per train-kilometre of the wide line (k = 1) or railway station (k = 2) in CZK/tkm
- O_i Number of occurrences of the category i in period t
- i Sub-category of extraordinary event
- UI_i Unit cost of extraordinary event according to table 2
- Q_k Number of train-kilometres on the wide line (k = 1) in in the railway station (k = 2)
- R_k The rate of extraordinary events on the broad line (k = 1) and in the railway station (k = 2)
- k Sub-category of parts of the railway network: 1 – track, 2 –station
- t Reference period (years)

The evaluated period corresponds to the period from which the data from the data-base of occurrences were analysed. This is the period 2011-2018, i.e. 8 years.

The calculated annual costs for each relevant sub-category of extraordinary event per one thousand train-kilometres of wide line and one thousand train-kilometres of the railway station are presented in Tab. 3.

Table 3 Calculated annual cost of sub-categories of extraordinary events (CA 2018)

OC category	Number of OC per year and mil. tkm in station	Number of OC per year and mil. tkm of wide line	Unit impact of OC in €	Average annual impact per 1000 tkm in station in €	Average annual impact per 1000 tkm of wide line in €
A1	0.06581	0.0006657442	913,964	60.15	0.61
A2	0.07458	0.0007545101	847,877	63.24	0.64
A3	0.01316	0.0001331488	2,415,941	31.80	0.32
B1	0.21059	0.0021303813	130,822	27.55	0.28
B2	0.31589	0.0031955720	90,538	28.60	0.29
B3	0.24569	0.0024854449	78,651	19.32	0.20
C1	0.40363	0.0040832309	7,473	3.02	0.03
C2	3.36506	0.0340417183	5,446	18.33	0.19
C3	4.59350	0.0464689427	7,043	32.35	0.33
C6	3.19395	0.0323107835	3,370	10.76	0.11
C12	0.62300	0.0063023781	4,479	2.79	0.03
C19	2.60606	0.0263634689	9,388	24.47	0.25

The total calculated annual costs of extraordinary events were defined as the sum of partial sub-categories. The final values are shown in Tab. 4.

Table 4 Values of total costs of extraordinary events (CA 2018)

Part of the railway network	Value
Railway station	321,95 €/1000 tkm/year
Wide line	3,26 €/1000 tkm/year

The original approach to the calculation of costs developed and published by the author's team in the paper [3] assumed that the impacts of extraordinary events (railway network includes wide line and stations) will be related to the kilometer of broad line and one station, both broken down into the national and regional level. However, in the practical verification, this method proved to be very general and insufficient, as it did not consider the transport performance in the individual sections. However, the traffic performance in the addressed sections of the railway network has a significant impact on the total number of occurrences of all categories. For this reason, the original procedure was abandoned and replaced by the method based on expressing

the impact of extraordinary event on a unit of transport performance (train-kilometre) presented above.

3 RESULTS

The subject of the paper is the suggestion of a methodological procedure for the determination of socio-economic impacts associated with increased user standards in the form of bigger reliability and safety of the railways due to the implementation of projects introducing appropriate safety equipment. When formulating the procedure, it was necessary to consider the availability of information on the addressed railway infrastructure and its history in relation to the number of occurrences and the expected impact of the planned measures on the frequency of occurrences after their implementation.
The author team developed two approaches to evaluate the solved benefits – detailed and simplified. Both approaches are shortly introduced in [33], detailed description of the simplified approach is in following chapters. This paper is oriented on the utilization of the average values described above within the simplified approach.

3.1 Utilization of the Unit Annual Costs of Partial Sub-categories of Extraordinary Events

If the categories of occurrences that shall be eliminated as a result of a project involving measures increasing railway safety and reliability are known, the Tab. 5 can be used. This table provides the average unit annual costs of subcategories of extraordinary events per one train kilometre (tkm) in the railway station or per one tkm on a wide line. It is assumed within the simplified approach that due to the implementation of the corresponding measures; the share of k occurrences of the corresponding category shall be eliminated. The savings associated with the elimination of a specific category of occurrence are determined separately according to the following relations.

Solutions within railway stations:

$$S_{aks} = N_s \times EAI_{ks} \times k \tag{2}$$

Where:
S_{aks} Annual savings when eliminating k category of extraordinary events at the railway station
N_s Number of tkm on railway stations per year in the assessed area
EAI_{ks} Expected annual impact of the k category of occurrences emerging at the station (see Tab. 3)
k Coefficient of decrease in the number of extraordinary events caused by implemented measures

Solutions within railway line sections (wide line):

$$S_{akl} = N_l \times EAI_{kl} \times k \tag{3}$$

Where:
S_{akl} Annual savings when eliminating k category of extraordinary events on the wide line
N_l Number of tkm on wide line per year in the assessed area
EAI_{kl} Expected annual impact of the k category of occurrences emerging on the wide line (see Tab. 3)
k Coefficient of decrease in the number of extraordinary events caused by project realization

The annual savings were calculated as the sum of annual savings for the specific sub-categories of extraordinary events and were consequently used as an economic cash-flow in the CBA processing.

3.2 Utilization of the Overall Costs of Extraordinary Events

In the absence of information on subcategories of occurrences that shall be eliminated as a result of the implementation of a project involving measures increasing railway line safety and reliability, the Table 6 can be used Table 6 presents the average annual costs of extraordinary events for all sub-categories together. Costs are related to one train kilometre in a railway station or to one train kilometre of a wide line as well. It is assumed within the simplified approach that due to the implementation of the corresponding measures; the share of k occurrences shall be eliminated. The savings associated with the elimination of occurrences were determined according to the following relation.

Solutions within railway stations:

$$S_{as} = N_s \times ETAI_s \times k \tag{4}$$

Where:
S_{as} Annual savings when eliminating extraordinary events arising at the railway station
N_s Number of tkm on railway stations per year in the assessed area
$ETAI_s$ Expected annual cost of extraordinary event at the station (see Tab. 4)
k Coefficient of decrease in the number of extraordinary events caused by project realization

Solutions within line sections (wide line):

$$S_{al} = N_l \times ETAI_l \times k \tag{5}$$

Where:
S_{al} Annual savings when eliminating extraordinary events arising on the wide line
N_l Number of tkm of wide line per a year in the assessed area
$ETAI_l$ Expected annual cost of extraordinary event on the wide line (see Tab. 4)
k Coefficient of decrease in the number of occurrences due to implemented measures

Calculated savings were consequently used as an economic cash-flow in the CBA processing. The coefficient of decrease in the number of occurrences due to the implemented measures *k* is assigned value 1. It is therefore assumed that the implementation of the addressed measures will prevent 100% of occurrences. The methodological procedure, which is based on the presented research and which is intended to support the economic analysis of projects in railway infrastructure, mentions the possibility of reduction of eliminated occurrences in the case that the given situation can be expected due to the project characteristics and the addressed transport network part. The percentage of possible reduction is left to the individual consideration of analysts familiar with the situation of the evaluated project.

The authors of this article, however, within their research activities, further focused on the average values of reducing the occurrence frequency due to the project implementation on the railway network. This reduction in frequency was determined based on a statistical comparison of a sample of projects on railways and including the implementation of measures aimed at increasing railway safety and reliability. The change in the number of occurrences in the corresponding section of the railway line caused by the realization of measures intended to improve reliability and safety on railways was monitored within the statistical comparison. The comparison in the article was carried out on a sample of 33 railway infrastructure projects for the situation prior to the measure implementation and after its implementation [4]. The coefficient *k* is equal to 10.77 % resulting from the detailed calculation stated in the cited article. The stated value was determined as an average value for all categories of occurrences. Looking at partial categories of extraordinary events, the following values of the coefficient *k* were calculated:

- Category A 27.23 %,
- Category B 37.77 %,
- Category C 9.82 %.

However, in the case studies presented below, the *k* coefficient was not used. The authors of the article are aware of the fact that on the one hand, this is more or less an indicative calculation showing the decrease of number of extraordinary events, but on the other hand, is not statistically conclusive. If necessary, in practice, a more detailed calculation which considers both the fact that the transport infrastructure project was implemented in the addressed section, which specific measures to increase safety and reliability were implemented and against which type of occurrences it was aimed at, can be designed and performed.

3.3 Case Study

The proposed methodological procedure was checked on a case study and the possibilities of the interpretation of the results resulting from the defined calculations were assessed. The proposed methodological procedure was verified on two projects of implementation of the measures increasing safety and reliability of the railway network. Verification was performed for a detailed approach as well as for both variants of the simplified approach.

The basic characteristics of the projects and the results of the calculations are given in the following part of the paper.

Project 1 Revitalization of the section of the national railway line No. 310A Opava East – Krnov – Olomouc

- Length of the section with an increase in the level of track interlocking equipment – station: 6.039 km,
- Length of the section with an increase in the level of the track interlocking equipment – wide line: 27.090 km,
- Number of freight trains per day: 2.79,
- Number of passenger trains per day: 38.6.

In the case of the station sections revitalization, it is assumed that due to new station interlocking equipment, occurrences of specific categories will be prevented. In the case of the revitalization of line sections, it is assumed that due to the new track interlocking equipment, occurrences of categories next selected categories will be prevented.

The expected annual savings associated with the improvement of user standards in terms of reliability and safety of the railway network were determined for the evaluated project using both versions of the simplified approach based on the input data and mathematical relations given in the outputs of this article.

The outputs are shown in Tab. 5.

Table 5. Annual positive impacts connected with improvement user standards in terms of reliability and safety - Project 1

Approach	Simplified	
Part of the network	According to categories	Overall impact
Stations (CZK/year)	523,090	705,863
Tracks (CZK/year)	21,086	32,032
Totally (CZK/year)	544,176	737,895

Project 2 Revitalization of the section of the national railway line No. 311 Bludov – Hanušovice

- Length of the section with an increase in the level of track interlocking equipment – station: 2.672 km,
- Length of the section with an increase in the level of track interlocking equipment – wide track: 18.748 km,
- Number of freight trains per day: 3.95,
- Number of passenger trains per day: 24.11.

In the case of the revitalization of station sections, it was assumed that due to new station interlocking equipment, occurrences of specific categories will be prevented. In the case of the revitalization of line sections, it is assumed that due to new track interlocking equipment, occurrences of next selected categories will be prevented.

The expected annual savings associated with the improvement of user standards in terms of reliability and safety of the railway network were determined for the evaluated project using both versions of the simplified approach based on the input data and mathematical relations given in the outputs of this article.

The outputs are shown in Tab. 6.

Table 6. Annual benefits associated with increasing railway safety and reliability - Project 2

Approach	Simplified	
Part of the network	According to categories	Overall impact
Stations (CZK/year)	156,906	211,731
Tracks (CZK/year)	9,893	15,029
Totally (CZK/year)	166,799	226,760

4 DISCUSSION

It is necessary to distinguish two points of view of the results of case studies within the discussion. The first view focuses on the differences in the results obtained by the individual defined approaches. A significant difference between the results obtained by de-tailed and simplified approaches was determined in both evaluated projects. However, the differences do not indicate the methodological shortcomings of the individual approaches, they clearly indicate the different principles on which the individual approach-es are based. The detailed approach uses the specific data about the evaluated section. It is therefore based on historical data on extraordinary events arising in the past. The approach builds on the assumption that similar occurrences with the same frequency as in previous years, may emerge in the future. On the contrary, both variants of the simplified approach are based on average annual values for the entire railway network broken down into a wide line and a railway station, both related to a unit of transport performance, i.e., train-kilometres. The detailed approach is therefore connected with the specific situation of a specific area of the railway network, while simplified approaches work with average values for the entire railway network. This information results in the recommendations regarding the use of individual approaches. The detailed approach is particularly suitable for those sections of the railway network which are specific to certain factors and whose characteristics indicate an increased incidence of occurrences of certain categories. On the contrary, the simplified approach can be effectively applied in the case of standard railway sections, which in the historical perspective do not deviate significantly from the average emergence of occurrences on the entire railway network. There are obvious reasons for the differences between the sub-variants of the simplified approach. When calculating according to categories, the methodological approach allows excluding those categories of occurrences which cannot be affected by the planned measures from the evaluation. The calculated benefits are thus reduced in accordance with the actual situation in the section. An approach using the overall impacts does not allow for this reduction, so it can be considered an easier way of evaluating bene-fits, however, at the same time, less accurate.

The second point of view is focused on the results of the calculation of benefits for individual evaluated projects. The benefits for Project 1 appear to be significantly higher than for Project 2. However, the reasons clearly lie in the scope of the projects, i.e., sections (line and station) that are subject to modernization, and transport performance annually carried out within the evaluated sections. With the growing scope of modernized sections and the growing number of trains passing through the modernized section every year, the number of potential occurrences that can be prevented by the implemented measures as well as realized benefits grow. It is also worth mentioning that in the case of Project 1, the benefits determined by the detailed approach are lower than the benefits determined by the simplified approach, and in Project 2 it is exactly vice versa. This difference highlights the recommendations provided in the previous section of the discussion.

The authors of the article believe that from the point of view of the international use of the proposed procedures, the general principles of determining unit benefits as well as the principles of the detailed and simplified approach can be used without problems even outside the Czech railway network. It can be recommended to adapt the input quantities to the values of the corresponding environment when using this calculation internationally.

5 CONCLUSIONS

The article is oriented on the infroduction of key resutls of a research activity oriented on evaluating of positive impacts related with increased user standards in terms of reliability and safety of railways due to the carrying out of modernization projects on the railway network. The paper builds on the previous research activities of the author team and presents the final version of partial methods for the evaluation of the impacts connected with reducing of extraordinary events. This is a detailed approach based on the use of historical data on occurrences emerging within the evaluated section and two variants of a simplified approach, which is based on average data for the entire railway network. In terms of input data, both approaches use a Database of Extraordinary events managed by the Administration of Railways and the Departmental Guidelines [11]. The presented approaches were consequently checked using two case studies focused on the modernization of the railway network, Finally, calculations and results were presented and discussed in this paper.

Both the detailed and simplified approaches are planned to be developed into individual methodological procedures to supplement the Departmental Guidelines [11] which shall subsequently serve to take into account the impacts connected with the increase of user standards in terms of reliability and safety of railways in the economic evaluation of railway infrastructure projects. For these purposes, the methodological procedure was examined by the representatives of the Railway Administration and, after incorporating the comments, it shall be handed over to the representatives of the State Fund for Transport Infrastructure, which is the implementation guarantor of the research project.

The addressed issue will be the subject of further research. The research team expects to continuously update the data in connection with new information that continuously appears in the Database of Occurrences. In addition, the research team expects to examine further possibilities of making the methodological procedure more

detailed, i.e., considering the division of lines into national or regional, possible reassessment of average passenger train occupancy or average freight train occupancy or refinement of the *k* coefficient calculation.

6 REFERENCES

[1] Hromádka, V., Korytárová, J., Vítková, E., Seelmann, H., & Funk, T. (2020). Evaluation of Socio-economic Impacts of Incidents on the Railway Infrastructure. *Current Topics and Trends on Durability of Building Materials and Components*, Serrat, C., Casas, J. R., & Gibert, V. (Eds). https://doi.org/10.23967/dbmc.2020.063

[2] Hromádka, V., Korytárová, J., Vítková, E., Seelmann, H., & Funk, T. (2021). Economic Impact of Occurrences on Railways. *Procedia Computer Science, 181*, 76-83. https://doi.org/10.1016/j.procs.2021.01.107

[3] Hromádka, V., Korytárová, J., Vítková, E., Seelmann, H., & Funk, T. (2020). New Aspects of Socio-economic Assessment of the Railway Infrastructure Project Life Cycle. *Appl. Sci., 10*(20), 7355. https://doi.org/10.3390/app10207355

[4] Funk, T. & Hromádka, V. (2020). Impact of Railway Modernization on the Frequency of Incidents. International Multidisciplinary Scientific GeoConference - SGEM 2020, 303-310. https://doi.org/10.5593/sgem2020V/6.2/s10.38

[5] Czech Railway Infrastructure Administration (CRIA). (2018). Statistika mimořádných událostí, databáze 2009 – 2018 (Database of Occurrences 2009 – 2018).

[6] Czech Railway Group Statistical Yearbook of the Czech Railways Group, 2018.

[7] Czech Railway Infrastructure Administration (CRIA). (2015). Prohlášení o dráze celostátní a regionální (National and regional railway declaration), 2015.

[8] Czech Railway Infrastructure Administration (CRIA). (2021). Jízdní řád (Timetable), https://www.spravazeleznic.cz/cestujici/jizdni-rad.

[9] Czech Railway Infrastructure Administration (CRIA). (2021). Abecední seznam stanic a zastávek (Al-phabetical list of stations and stops), https://www.spravazeleznic.cz/documents/50004227/50157161/ seznam-stanic.pdf.

[10] Sartori, D. (2016). Guide to Cost-benefit Analysis of Investment Projects, Economic appraisal tool for C-hesion Policy 2014-2020. European Commission, Directorate-General for Regional and Ur-ban policy.

[11] Ministry of Transport of the Czech Republic (MoT CZ). (2017). Departmental Guideline for the Evaluation of Economic Effectiveness of Transport Construction Projects. 2017. Retrieved from http://www.sfdi.cz/pravidla-metodiky-a-ceniky/metodiky/.

[12] EY Rail Baltica Global Project Cost-Benefit Analysis. (2017). Final Report 30, April 2017, https://www.railbaltica.org/wp-content/uploads/2017/04/RB_CBA_FINAL_REPORT_0405.pdf.

[13] European Union Agency for Railways Report on Railway Safety and Interoperability in the EU. (2020). Available on https://www.era.europa.eu/sites/default/files/library/docs/safety_interoperability_progress_reports/report_on_railway_safety_and_interoperability_in_the_eu_2020_en.pdf.

[14] Mátrai, T. Cost benefit analysis and ex-post evaluation for railway upgrade projects. (2013). *Periodica Polytechnica. Transportation Engineering, 41*(1), 33-38. https://doi.org/10.3311/PPtr.7102

[15] Olsson, N., Oekland, A., Halvorsen, S., & Olsson, N. (2012). Consequences of differences in cost-benefit methodology in railway infrastructure appraisal: A comparison between selected countries. *Transport Policy, 22*, 29-35. https://doi.org/10.1016/j.tranpol.2012.03.005

[16] Siciliano, G., Barontini, F., Islam, D., Zunder, T., Mahler, S. & Grossoni, I. (2016). Adapted cost-benefit analysis methodology for innovative railway services. *European Transport Research Review, 8*(23), 1-14. https://doi.org/10.1007/s12544-016-0209-5

[17] Carteni, A. & Henke, I. (2017). External costs estimation in a cost-benefit analysis: The new Formia-Gaeta tourist railway line in Italy. *2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)*, 1-6. https://doi.org/10.1109/EEEIC.2017.7977614

[18] Evans, A. W. (2013). The economics of railway safety. Research in transportation economics [online]. Elsevier, 2013, 43(1), 137-147 [cit. 2021-03-03]. ISSN 0739-8859. Available from: doi:10.1016/j.retrec.2012.12.003.

[19] Oertli, J. (2000). Cost–Benefit Analysis in Railway Noise Control. *Journal of Sound and Vibration, 231*(3), 505-509. https://doi.org/10.1006/jsvi.1999.2539

[20] Tseng, Y., Rietveld, P., & Verhoef, E. (2012) Unreliable trains and induced rescheduling: implications for cost-benefit analysis. Transportation, 39(2), 387-407. https://doi.org/10.1007/s11116-011-9345-x

[21] Johnsen, S. & Veen, M. (2013). Risk assessment and resilience of critical communication infrastructure in railways. *Cognition, Technology & Work, 15*(1), 95-107. https://doi.org/10.1007/s10111-011-0187-2

[22] Petrova, E. (2015). Road and railway transport in Russia: safety and risks. *Advances in Environ-mental Sciences, 7*(2), 259-271. https://search.proquest.com/docview/2018602917/

[23] Evans, A. W. (2011). Fatal train accidents on Europe's railways: 1980–2009. *Accident Analysis and Prevention, 43*(1), 391-401. https://doi.org/10.1016/j.aap.2010.09.009

[24] Santos-Reyes, J., Beard, A. N., & Smith, R. A. (2005). A systemic analysis of railway accidents. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 219*(2), 47-65. https://doi.org/10.1243/095440905X8745

[25] Klockner, K. & Toft, Y. (2018). Railway accidents and incidents: Complex socio-technical system accident modelling comes of age. *Safety Science, 110*, 59-66. https://doi.org/10.1016/j.ssci.2017.11.022

[26] Read, G. J. M., Lenné, M. G., & Moss, S. A. (2012). Associations between task, training and social environmental factors and error types involved in rail incidents and accidents. *Accident Analysis & Prevention, 48*, 416-422. https://doi.org/10.1016/j.aap.2012.02.014

[27] Kim, H. J., Jeong, J. Y., Kim, J. W., & Oh, J. K. (2016). A Factor Analysis of Urban Railway Casualty Accidents and Establishment of Preventive Response Systems. *Procedia - Social and Behavioral Sciences, 218*, 131-140. https://doi.org/10.1016/j.sbspro.2016.04.016

[28] Baysari, M. T., McIntosh, A. S., & Wilson, J. R. (2008). Understanding the human factors contribution to railway accidents and incidents in Australia. *Accident Analysis and Prevention, 40*(5), 1750-1757. https://doi.org/10.1016/j.aap.2008.06.013

[29] Iridiastadi, H. & Ikatrinasari, Z. F. (2012). Indonesian railway accidents - Utilizing Human Factors Analysis and Classification System in determining potential contributing factors. *Work, 41*(1), 4246-4249. https://doi.org/10.3233/WOR-2012-0126-4246

[30] Madigan, R., Golightly, D., & Madders, R. (2016). Application of Human Factors Analysis and Classification System (HFACS) to UK rail safety of the line incidents. *Accident Analysis & Prevention, 97*, 122-131. https://doi.org/10.1016/j.aap.2016.08.023

[31] Lombardi, M., Guarascio, M., & Rossi, G. (2014). The Management of Uncertainty: Model for Evaluation of Human Error Probability in Railway System. American Journal of Applied Sciences, 11(3), 381-381. https://doi.org/10.3844/ajassp.2014.381.390

[32] Hani Tabai, B., Bagheri, M., Sadeghi-Firoozabadi, V., & Sze, N. N. (2018). Evaluating the impact of train drivers' cognitive and demographic characteristics on railway accidents. *Safety Science, 110*, 162-167. https://doi.org/10.1016/j.ssci.2018.03.027

[33] Hromádka, V., Korytárová, J., Vítková, E., Seelmann, H. & Funk, T. (2021). Methodological Approach Valuating the Impacts Associated with Occurrences on Railways. *The 74th International Scientific Conference on Economic and Social Development, Book of Proceedings. Book Series: International Scientific Conference on Economic and Social Development*, Varazdin, Croatia, 264-272.

Authors' contacts:

Vít Hromádka, Assoc. Prof., MSc., Ph.D.
Brno University of Technology, Faculty of Civil Engineering
Veveří 331/95, 602 00 Brno, Czech Republic
hromadka.v@fce.vutbr.cz

Jana Korytárová, Prof., MSc., Ph.D.
Brno University of Technology, Faculty of Civil Engineering
Veveří 331/95, 602 00 Brno, Czech Republic
korytarova.j@fce.vutbr.cz

Eva Vítková, MSc., Ph.D.
Brno University of Technology, Faculty of Civil Engineering
Veveří 331/95, 602 00 Brno, Czech Republic
vitkova.e@fce.vutbr.cz

Herbert Seelmann, MSc.
Brno University of Technology, Faculty of Civil Engineering
Veveří 331/95, 602 00 Brno, Czech Republic
seelmann.h@fce.vutbr.cz

Tomáš Funk, MSc.
Brno University of Technology, Faculty of Civil Engineering
Veveří 331/95, 602 00 Brno, Czech Republic
funk.t@fce.vutbr.cz

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230504094426

Subject review
Received: 2023-05-04, Accepted: 2023-05-08

Standardization of Project Management Practices of Automotive Industry Suppliers - Systematic Literature Review

Filip Knapp*, Michal Šimon

Abstract: This paper deals with the issue of standardization of project management (PM) practices of development activities of automotive industry suppliers. Development activities are critically important for ensuring the final quality of the developed products and related production systems while keeping the target costs. The key to their successful management is considered the use of modern PM practices. In order to represent the current best practices and to be apparent and transparent, the PM practices must be properly standardized. A comprehensive branch standard of PM in automotive industry is however missing. In order to fill the gap in the current state of knowledge, consecutive steps of standardization were defined. This article describes one of the initial steps, a systematic literature review of existing resources that could serve as a valuable basis for standardization. The results show that general PM standards are suitable for treating the quality of project processes and branch PM standards of automotive industry for treating the quality of project product – manufacturing processes of automotive suppliers. The paper discusses the individual methods within both groups in more detail and recommends what purpose they are suitable for.

Keywords: automotive industry suppliers; product creation process; project management; standardization

1 INTRODUCTION

In recent decades, the automotive industry has undergone intensive development. The growing market pressure on car prices and on shortening the duration of development on the one hand, and the ever-increasing complexity and technical demandingness of products in the face of increasing competition and increasing demands of end-users on quality and guarantees on the other hand, are forcing car manufacturers (OEMs) to move ever greater depth of development and production towards suppliers. Suppliers have taken over much of the research and development (R&D) and production from car manufacturers, achieving an overall increase of 70-80% in the process. Richly branched dependent supplier-customer production networks have become an integral part of this industry and car production today is heavily dependent on suppliers. This trend defined a completely new qualification profile for suppliers. Suppliers take full responsibility for the results of development of product and related production systems, product launch and delivery, and absolute fulfilment of the list of requirements and target costs defined by the car makers [1-3]. They thus find themselves in a completely new role, which they are forced to adapt to.

In the past, the Serial production stage was considered to be the decisive stage in terms of ensuring the required quality of the developed products and production systems and achieving the target costs. Currently, it is already widely recognized that up to 80% of the resulting quality and costs are decided already in the pre-production stages of the car's life cycle, collectively called the Product Creation Process (PCP). In the pre-production stages, suppliers plan, create and fine-tune the concepts of future products and production systems and make critical decisions that determine whether the final product will meet the customer's requirements, be competitive and ensure a reasonable profit for the manufacturer. The high influence of pre-production stages on the final quality is directly related to the fact that in these stages there are many more non-conformities than in the implementation phase. In addition, practical experience shows that the expenses associated with the elimination of non-conformities in the pre-production stages require only a fraction of the costs necessary for the elimination of non-conformities during the realization and use of the product (The Rule of Ten) [4]. The earlier thus the suppliers discover and treat the non-conformities, the less effort and resources they have to spend on their elimination and the better results they achieve. In such an environment, project management (PM) takes on a decisive importance.

PM is designed to effectively manage complex, unrepeatable, time- and resource-constrained activities with long durations and high levels of uncertainty. It is exactly such attributes that the development activities of the PCP phase show. PM practices have proven to be an effective tool for their management and mastering the practices turned out as a strategic competence of enterprises in this highly competitive environment [8, 21]. However, in order for PM practices to be truly effective, to represent the best current solutions and to be apparent and transparent, PM practices must be properly standardized. Although there are thousands of suppliers per one OEM today (e.g. the global supplier network of the VW concern includes more than 40,000 suppliers [5]), a comprehensive branch standard of PM in automotive industry is missing.

The main motivation of this research was the will to fill this gap in the current state of knowledge and to contribute to the standardization of the PM industry standard. The subject of the research was defined as a systematic literature review (SLR) of information sources dealing with the issue being addressed. The expected output is recommendations which sources, why and for what purpose suppliers can use in the standardization process.

2 EXPLANATION OF THE CONTEXT OF STANDARDIZATION

To understand the role of SLR in the standardization process, it is useful to first explain its context.

Step 1: Mapping the processes on OEM side

The initial key step in the standardization process was mapping of the processes on side of car manufacturers. Car manufacturers stand at the top of supplier-customer networks and their activities influence the entire dependent chain below them. Understanding processes and events running on the side of the OEMs is thus an integral part of the standardization of activities on the side of their suppliers.

As it emerged from the SLR, the basic framework of cooperation between car manufacturers and their suppliers is defined by the model of Product Creation Process (PCP). PCP is the initial part of the car's life cycle, in which its research and development takes place. The activities of this stage are very complex, diverse and are implemented by development teams requiring mutual coordination. For these reasons, OEMs elaborate the PCP stage in detail into a clear graphic model containing key activities and events running on the side of OEMs for the purposes of determining the continuity of individual processes. The analysis and synthesis (reconstruction) of the PCP model of OEMs took place in the previous step of the research and its closer understanding became the inspiration for the creation of a process model on the suppliers' side.

Step 2: Defining the architecture of the PCP model of Tier 1 automotive suppliers

The construction and hierarchy of the proposed PCP model of Tier 1 automotive suppliers is shown in Fig. 1. From the point of view of complexity and scope of detail, the model consists of 4 vertical levels.

- **Level N** composes of basic PCP phases bordered by gate-type milestones.
- **Level N-1** composes of process steps in the form of project milestones horizontally located in so-called swim lanes depending on the responsibility for their fulfillment. Milestones represent critical process steps that must be completed to be able to move from one gateway to another.
  Form: Level N and N-1 are interpreted in the form of a process map.
- **Level N-2** is defined by sub-activities (also called work packages) that must be fulfilled to be able to close the process steps.
  Form: Process cards containing inputs, activities/ work packages, responsibilities, outputs and supporting PM tools and techniques.
- **Level N-3** is defined by auxiliary tools and techniques of PM.
  Form: Methods, instructions, manuals, rating tables, diagrams, etc.

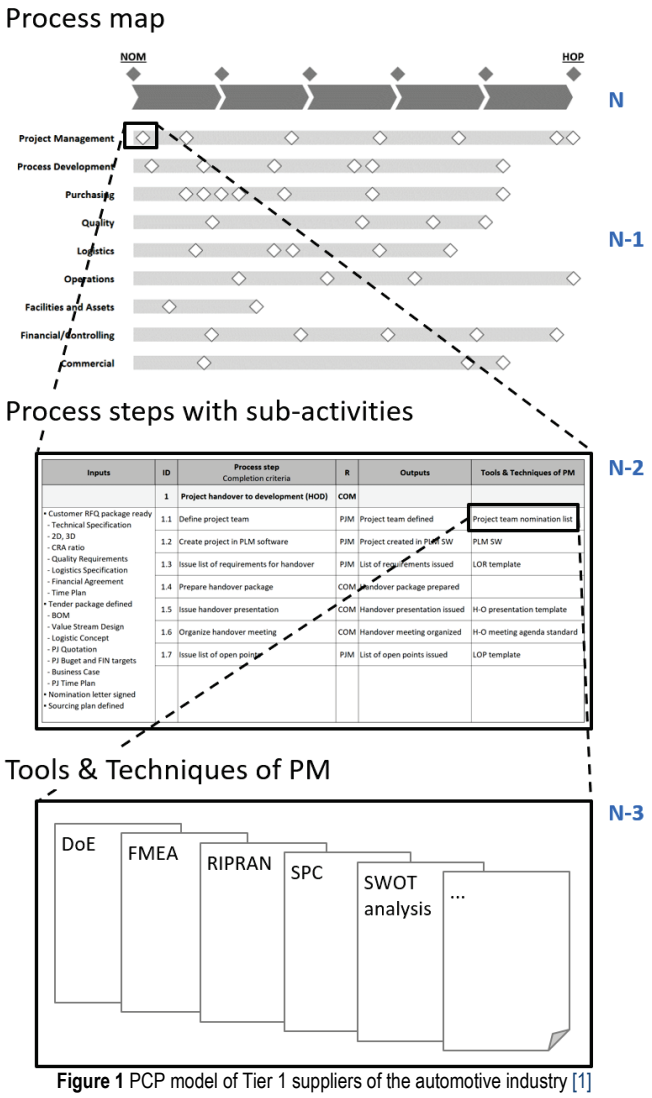


Figure 1 PCP model of Tier 1 suppliers of the automotive industry [1]

The definition of the form of the standard was followed by the proposal of its content.

Step 3: Defining the content of the PCP model of Tier 1 automotive suppliers

In order for the model to reliably fulfill its function, it is necessary not only to choose a suitable form, but above all to ensure its eligible content. The proposed procedure for ensuring an eligible content is presented in Tab. 1.

Table 1 Matrix of definition of eligible content of the PCP model

		Vertical level of PCP			
#	Activity / Method	N	N-1	N-2	N-3
1	A: Mining existing literary sources M: Systematic Literature Review	×	×	×	×
2	A: Defining Key Success Factors of PM in automotive industry M: Delphi method	×	×		
3	A: Defining/treating process steps M: Structured interview w. experts	×	×	×	×

After explaining the context of standardization this paper further focuses on description of activity #1 - SLR of sources

that have already dealt with the problematics and that can potentially serve as an input in all 4 levels of the PCP model.

3 SYSTEMATIC LITERATURE REVIEW OF SOURCES TREATING PROJECT MANAGEMENT OF AUTOMOTIVE INDUSTRY SUPPLIERS

Considering the limited scope of this paper, the research focuses exclusively on the investigation of PM standards applicable to Serial suppliers (Build to Print), who do not deal with product development and specialize in the development of production systems and the provision of serial deliveries (see research limitations below).

For successful project management, it is necessary to properly treat simultaneously the quality of the project processes and the quality of the project product [6]. Since the main business activity of Serial suppliers of the automotive industry is the development of production systems and the provision of serial deliveries, the basic prerequisite for the success of the PM of these companies is the mastery of procedures and tools that treat both the quality of the project management itself and the quality of the production systems that are developed and implemented within the project. Despite an extensive search in global databases, it was not possible to find a branch standard that dealt with the issue of project management of Serial suppliers of the automotive industry, treated simultaneously the quality of project processes and the quality of project product, and had the character of a comprehensive methodology. Due to the absence of a comprehensive solution, a new strategy was defined. Its essence was not to continue on the search for information sources that fully correspond to the assignment, but to look specifically for sources that treat the quality of the project's processes and, in particular, for sources that treat the quality of the project's product.

3.1 Sources Treating the Quality of Project Processes - General Project Management Standards

To treat the quality of the project's processes, the research focused on general PM standards.

3.1.1 Review of General Project Management Standards

Current best practices used in project management are summarized in internationally recognized general standards. They are based on decades of experience of thousands of PM professionals (project sponsors, project managers, project teams, academics, trainers, consultants) working in a wide range of industries and geographies, and are based on their established and proven project management best practices. General standards are designed to be applicable to any project regardless of its scope, type, organization, geography, or socio-technical environment. This is achieved by separating project management activities from the specific needs and requirements of individual fields, companies, or specific users. However, specialized aspects of any type of project are integrated with these standards and together they provide a secure overall framework for project work. Due to their properties, general standards are understood and recognized by the general public and form the proven basis of any project [7].

During the research, it turned out that in terms of the number of certifications granted, the most popular general standards of PM worldwide are:

- PMBOK Guide by Project Management Institute,
- IPMA ICB4 by International Project Management Association, and
- PRINCE2 by Axelos PRINCE.

PMBOK Guide

As it emerged from the research of the publication PMBOK Guide [8], the PMBOK standard is based on a total of five main process families, ten knowledge areas, forty-seven processes and their interrelationships. All processes and process steps have defined inputs, outputs and transformation tools (actions, methods, techniques). In its entirety, this PM standard contains a set of 132 process tools and techniques, including their description and functionality. The PMBOK Guide is a de facto set of project tools and methods designed for all project phases. The advantage of the standard is the complexity, detail and clarity given by the clear structure. The PMBOK Guide standard is process-oriented and based on managerial practice, focusing on best practices that are applicable to most projects. In project management, the standard does not provide the user step-by-step guidance on how to manage a project throughout its life cycle, but serves as a rich reference guide for those who want to have the widest possible range of analytical techniques and recommendations at hand. It is intended for experienced project managers who do not need step-by-step guidance through all phases of the project life cycle [7, 8, 12].

IPMA ICB4

As shown by the review of the publication ICB4 [9] and as its name Individual Competence Baseline implies, ICB4 is by nature a standard that defines the competencies (abilities, knowledge, skills) required for successful project management. The ICB4 project management competence standard distinguishes and describes in detail three competence areas that make up the so-called Eye of Competence. The competence areas mentioned are contextual, behavioural and technical (29 in total). Behavioural competencies define the personal and interpersonal skills required for successful project management, technical competencies define the technical aspects of project management, and contextual competencies refer to those competencies that form the context of project management. Within each of these domains, the standard defines the appropriate knowledge and skills that are required for successful management in that area. The focus is then on the ability of appropriate application by specific personalities. So there is a lot of room for creativity and one's own opinion. The ICB standard therefore does not dictate processes, but recommends certain process steps that need to be appropriately applied to a specific project situation.

Unlike the previous process concept of the PMBOK standard, the concept of the IBC standard is competence. The standard is therefore not focused on the exact form of the defined processes and their specific application, but focuses on the person of the project manager and members of the project team and on what knowledge, skills and abilities - competences - they should have in order to successfully manage projects [9, 10, 12].

PRINCE2

The review of the standard Managing Successful Projects with PRINCE2 [11] showed that it is based on seven principles, consists of seven processes and describes seven topics. It is possible to adapt the standard within a specific project, but it is necessary to take into account the principles that are the backbone of the entire standard. Individual processes can be greatly simplified and each of them has many possibilities of use according to the specifics of the project. However, the principles remain and ensure that the project is a project in a controlled environment. The review also revealed that PRINCE2 is not a project management standard in the sense of the previous two PMI or IPMA standards. It is more of a guide or project management methodology. Unlike the previous two analysed standards, the PRINCE2 methodology focuses on WHAT, WHEN, WHO and WHY need to be done. What, on the other hand, cannot be found in this methodology is the answer to the question of HOW to do it. Thus, detailed coverage of project management tools and techniques is lacking. The PRINCE2 methodology has approximately 40+ techniques in its entirety, which it refers to, however, only 2 techniques are described in more detail - Quality control and Product planning. For comparison, the PMBOK standard contains a total of 132 tools, which it elaborates in detail. The methodology explains the absence of tools and techniques by the fact that there is a whole range of techniques for planning and controlling projects, which are described in detail in separate books, so it is not necessary to repeat it again in the PRINCE2 manual. PRINCE2 also does not address management skills as one of the key factors in competent project management. The concept of the PRINCE2 methodology is thus different from other PM standards not only in terms of the insight into PM issues, but also in the way it is processed and the use of different terminology, which is chosen with regard to the environment (state administration) in which it was developed [11-14].

3.1.2 Common Attributes of General Project Management Standards

As it emerged from the evaluation of General project management standards, everyone views project management from a different angle. Nevertheless, they exhibit common attributes that characterize the standards and distinguish them from PM standards with a different scope of common agreement. These attributes can be summarized in the following two categories defined based on the contribution of the assessed resources to the needs of this research:

Advantages (+):
- They treat the quality of project processes,
- They are universal, applicable to any project,
- Specific aspects of any project are integrated with these standards,
- They represent best practices of the field.

Disadvantages (–):
- They do not treat the quality of the project product,
- They do not treat specific aspects of particular fields,
- They do not offer a comprehensive guide to pm - they represent either methodology or tools,
- They have a considerable scope,
- Adaptation to a specific project requires a high administrative burden.

3.1.3 Evaluation of General Project Management Standards and their Validity for this Research

There is no clear answer to the question of whether the PRINCE2 Manual, PMBOK Guide, or ICB4 is better for PM of automotive suppliers. None of the standards offers a balanced, universal guide to project management, each treats a different aspect of project management, and each has its strengths and weaknesses. Based on the research and evaluation of the attributes of individual standards, the following recommendations can be made:

1) The PRINCE2 methodology can be recommended as the main information source suitable for creating a robust foundation (skeleton) of the planned methodology for the area of treatment of the quality of project processes. PRINCE2 is the only one of the analysed PM standards that has the nature of a methodology, it guides the user step by step through all phases of the project life cycle, it is strictly formal, it contains a detailed description of roles and competences, it can be adapted to the nature and scope of a specific project and its use is possible even without prior knowledge of the user. It is a guide in the entire life cycle of the project and gives very detailed instructions on WHAT to do, WHEN to do it, WHO should do it and WHY.
2) What, on the other hand, cannot be found in this methodology is the answer to the question of HOW to do it. For that reason, it is advisable to combine the methodology with another standard that describes these tools and methods in detail and thus appropriately complements the main shortcoming of the PRINCE2 methodology.
3) The PMBOK standard can be recommended as a supplementary information source that will serve as a reference manual containing a wide range of proven PM tools and techniques (132) for all phases of the project, including their detailed description and description of functionalities.
4) The IPMA standard, which defines the competencies of the project manager required for successful project management, is of no fundamental importance for the creation of an industry-specific PM methodology of automotive industry suppliers.

See Tab. 2 for a summary.

Table 2 Summary table presenting the resulting recommendations for sources treating the quality of project processes

	PCP level treated			
Recommended literary sources	N	N-1	N-2	N-3
PRINCE2 (Methodology)	×	×	×	
PMBOK (Tools & Techniques of PM)				×

The selected general project management standards PRINCE2 and PMBOK have been selected as globally recognized and proven best practices by PM professionals from around the world. Although the standards approach project management from two different perspectives, these are two of the most recognized methods of global project management, which fit together well and their combination appropriately covers the needs of this research.

However, since by their very nature general PM standards do not treat the quality of the project's product and do not take into account the specific needs of different fields (a tax for their universality), in the following part the review will focus on identifying, analysing and evaluating relevant information sources that meet these attributes. The review will be performed in the branch databases of the automotive industry [15-19].

3.2 Sources Treating the Quality of the Product of the Project - Branch Standards of Project Management of the Automotive Industry

During the research, branch standards created and issued by trade groups of the automotive industry were identified as a suitable source for treating the quality of the production systems of suppliers of the automotive industry. Automotive industry business groups bring together car manufacturers and major suppliers of the industry, and their purpose is to develop topics of common interest, share practical experience and define rules facilitating cooperation between all levels of the supply chains of the industry. A search of information sources revealed that the following groups are among the world's most important trade associations of the automotive industry: AIAG (USA), ANFIA (IT), FIEV (FR), JAMA (JP), SMMT (UK), or VDA (DE). Considering the number of business groups, it was decided to limit the search at this point to an analysis of branch standards issued by the AIAG group. AIAG is the most popular of the listed business groups worldwide, has a wide field coverage, and its publications are used with confidence by car makers and their suppliers all over the world.

AIAG Group

The Automotive Industry Action Group (AIAG) is a global non-profit organization of car manufacturers and their suppliers founded in 1982 in North America. The company was founded by the progressive managers of Chrysler, Ford Motor Company and General Motors (the Big Three). Currently, AIAG has over 4,000 members, which include OEM, such as GM, Honda, Nissan, Stellantis, Toyota, or Volkswagen, and important suppliers of this industry, such as Adient, Bosch, Continental, Lear Magna, or ZF [20].

The goal of AIAG is to increase prosperity in the automotive industry by improving business processes and activities that are part of the supply chain. Under the auspices of the AIAG, the gathering of ideas on the basis of specific problems of the supply chain of the automotive industry takes place in the form of an open forum, and a suitable solution is sought together. People at all levels of management participate in it. The result of the mutual cooperation are the branch standards of the automotive industry that help to align the procedures of suppliers of different subject matter (products, services) and scope of supply and different supplier levels. By standardizing the definitions of terms and their content, the standards support the consistency of work in the supply chain, create transparency and mutual understanding between its individual participants, they can be adapted to the needs of the respective project and, through proven methods and tools, they help to ensure the agreed quality of the start-up, deliveries in the pre-production stages and even deliveries during series production [20].

In the databases of the AIAG group, 80+ publications divided into the following categories were discovered:

- Quality publications,
- Supply chain publication, and
- Publication of social responsibility.

The following selection criteria were defined to select suitable information sources. The publication had to:

1) Be universal, applicable regardless of the subject of supplies and the nature of production systems,
2) Treat the quality of production systems,
3) Be applicable in the PCP phase.

The following publications met the defined criteria:

- Advanced Product Quality Planning (APQP),
- Production Part Approval Process (PPAP),
- AIAG & VDA FMEA Handbook,
- Measurement System Analysis (MSA),
- Statistical Process Control (SPC),
- The Costs of Poor Quality Guide (COPQ),
- Layered Process Audit (LPA) Guideline, and
- Effective Problem Solving Guide.

3.2.1 Review of Branch Standards of Project Management of the Automotive Industry

From the review of selected publications of the AIAG group, it emerged that APQP has the unique nature of a comprehensive methodology tailored to the needs of automotive industry. The remaining publications can be characterized as well-known, proven tools and techniques of project management of industrial enterprises, which can also be well applied in the automotive industry. Since these tools and techniques are covered in detail in separate books, there is no need to repeat them again in this paper. For that reason, only the APQP methodology will be described in the following part.

Advanced Product Quality Planning

As declared by AIAG, APQP is part of the five Core Tools for effective quality management with PPAP, FMEA, MSA, and SPC being the other core tools. Advanced Product Quality Planning [22] is a manual providing guidelines designed to produce a product quality plan, which will support the development of a product or service that will satisfy the customer. The primary goal of product quality planning is to facilitate communication and collaboration between engineering activities. As such, it requires the engagement of a cross-functional team (CFT) that includes marketing, product design, procurement, manufacturing, and distribution. The objective is to ensure a clear understanding of the voice of the customer (VOC), and to translate it into requirements, technical specifications, and special characteristics. APQP consists of five phases. Within the individual phases, the methodology recommends the use of previously mentioned supporting tools, such as FMEA, Control Plan, SPC, or MSA, which increase the probability of success of development activities. For practical reasons, the five phases are translated into five main activities that correspond to the stages of the Product Creation Process (PCP). The contents of the individual phases are presented in the methodology through recommended decisive inputs and outputs. The outputs of the previous phase automatically become the inputs of the upcoming phase. Each input/output is then briefly characterized. The verification of the correctness of the procedure is enabled by predefined checklists of questions located in the appendix. As the book explains, the use of the APQP methodology leads to simplification of product quality planning, efficiency and improvement of development activities, significant cost savings and facilitation of communication with subcontractors [22, 23].

3.2.2 Common Attributes of Branch Standards of Project Management of the Automotive Industry

Similarly to the general PM standards mentioned in the previous text, the AIAG branch standards share common attributes that characterize the standards and distinguish them from PM standards with a different scope of common agreement. They are again structured into two categories based on their contribution to the needs of this research:

Advantages (+):

- They solve the needs of automotive industry and thus fulfil the essence of branch standards,
- They treat the quality of project products of the automotive industry,
- They represent the Voice of the Customer (VOC),
- They have a universal character and can be tailored,
- They represent best practices of the field.

Disadvantages (–):

- They treat the quality of project processes only to a very limited extent,
- They do not offer a comprehensive and balanced guide - they represent either methodology or tools,
- They are tools of PM of suppliers and complement this, but in no case replace it,
- They have only a recommendatory character.

3.2.3 Evaluation of Branch Standards of Project Management of the Automotive Industry and their Validity for this Research

After analysing all selected publications of the AIAG group and subsequent evaluation of their properties, the following conclusion were made:

1) The APQP publication can be recommended as the main source of information that will support the creation of the skeleton of the planned methodology for the area of treatment of the quality of project product. Similarly to PRINCE2 in the field of general PM standards, APQP is the only one of the analysed publications of the AIAG group to be of the nature of a complex methodology. It was tailored from the very beginning by the AIAG group to the needs of the automotive industry, it is a guide in the entire PCP phase, it defines the scope of planning participants, it can be adapted to the nature and scope of a specific project, and its use is possible even without prior expert knowledge of the user. The publication gives rough instructions on WHAT to do, WHEN to do it, WHO should do it and WHY.
   What, on the other hand, cannot be found in this publication is the answer to the question of HOW to do it. Although APQP recommends 11 analytical techniques that can be used in product and process quality planning and thus increase the probability of project success, it describes the analytical techniques very briefly. Similarly to PRINCE2, it refers to techniques as well-known tools detailed in separate books that do not need to be described again. As for the number of techniques, APQP mentions only 11 techniques, unlike, for example, the PMBOK standard, which lists 132. In addition, the methodology exhibits a low level of detail, overgeneralization of content, and a disproportionately small range compared to general PM standards. For comparison, while the PRINCE2 methodology describes PM procedures on 480 pages, the APQP methodology on 108 pages only. From the above reasons, it is advisable to combine the APQP methodology with another information sources that will appropriately complements its shortcomings.
2) The remaining publications of the AIAG group, such as PPAP, FMEA, MSA, SPC, COPQ, LPA and Effective Problem Solving Guide, can be characterized as widely known and proven tools and techniques of project management of industrial enterprises, which can also be well applied in the automotive industry. Their analysis confirmed that all of them can be confidently recommended as a supplementary information source that will suitably complement the deficiencies of the APQP methodology.
   See Tab. 3 for a summary.

Table 3 Summary table presenting the resulting recommendations for sources treating the quality of the project product

	PCP level treated			
Recommended literary sources	N	N-1	N-2	N-3
APQP (Methodology)	×	×	×	
PPAP, FMEA, MSA, SPC, COPQ, LPA, EPSG (Tools & Techniques)				×

From the findings of this part of the review, it can be conducted that the selected publications of the AIAG group represent appropriate information sources very well usable for the standardization of PM procedures of automotive industry suppliers. The resources address topics that support product and process quality assurance in the supply chain at the PCP stage, intertwining and complementing each other as appropriate. However, since the branch standards of the automotive industry do not sufficiently treat the quality of the project processes, it is recommended to combine them with the previously recommended general PM standards and thus create a suitable combination, which would appropriately cover both key areas of project management and provide a safe overall framework for processing the branch methodology of PM in the automotive industry.

4 THESES OF THE SYSTEMATIC LITERATURE REVIEW

The key findings resulting from this research can be summarized in the following theses:

- Product Creation Process (PCP) phase activities have a major impact on the quality and costs of future products and related production systems.
- In the PCP phase, most non-conformities arise from the entire product life cycle.
- Eliminating non-conformities in the PCP phase requires only a fraction of the costs necessary to eliminate non-conformities during the implementation and use of the product.
- The occurrence of non-conformities in the Development phase can effectively be prevented by the use of modern project management practices.
- For successful project management, it is necessary to properly treat both the quality of the project processes and the quality of the project product.
- The quality of project processes is treated by general project management standards.
- General PM Standards are comprehensive, universal and adaptable to any PJ, but they do not treat the quality of the project product and branch specifics.
- The quality of production systems of Serial suppliers of the automotive industry is treated by branch PM standards of the automotive industry.
- Branch standards of PM treat the needs of automotive industry and represent the Voice of the Customer, but they treat the quality of PJ processes to a very limited extent and do not offer a comprehensive guide.

5 CONCLUSION

This paper deals with the issue of standardization of project management of development activities of automotive industry suppliers. The subject of the research was defined as a systematic literature review of relevant information sources and the output recommendations which of the existing sources, why and for what purpose suppliers can use in the standardization process.

As described in the previous text, standardization of PM practices requires inclusion of procedures and tools that treat both the quality of project processes and the quality of the project product. In this research, the production systems of the suppliers of the automotive industry are considered to be the product of the project.

To treat the quality of project processes, general PM standards were identified as a suitable information source. General PM Standards are designed to be applicable to any project regardless of its scope, type, organization, geography, or socio-technical environment and they provide a secure overall framework for project work. As the main source of information for creating a skeleton of PM standard of suppliers of the automotive industry (PCP model level N to N-2) is recommended the methodology Managing Successful Projects with PRINCE2. As a supplementary information source with a wide range of proven project management tools and techniques is recommended the PMBOK Guide (level N-3). However, since by their very nature general PM standards do not treat the quality of the project product and do not take into account the specific needs of specific fields, it is recommended to combine them with the branch standards of the automotive industry. These will suitably complement their deficiencies.

To treat the quality of the project product and the specific needs of the automotive industry, the branch PM standards of the business group AIAG were identified as a suitable information source. As the main information source for creating a skeleton of PM standard of suppliers of the automotive industry (PCP model level N to N-2), the APQP methodology is recommended. As supplementary auxiliary tools and methods of project management for treating the quality of production systems of automotive industry suppliers (level N-3) the publications PPAP, FMEA, MSA, SPC, COPQ, LPA and Effective Problem Solving Guide are recommended.

The previous analysis showed that the recommended sources fit together very well and their proper combination will provide a secure overall framework for creating a competent methodology of project management of development activities of automotive industry suppliers that is so desperately lacking.

Acknowledgement

The authors acknowledge the financial support from the project SGS-2023-025 entitled "Environmentally sustainable production" resolved within the Internal Grant Agency of the University of West Bohemia in Pilsen.

6 RESEARCH LIMITATIONS

- Application sector
  - Automotive industry

- Serial suppliers without own product development
- Process orientation
- Project management (not program or portfolio)
- Scope of common agreement
- Branch methodology of the automotive industry
- Part of the Product Life Cycle to be treated
- Product Creation Process (PCP).

7 REFERENCES

[1] VDA-QMC. Product creation - Maturity level assurance for new parts. Germany: VDA, the 2nd Revised Edition, October 2009. 96 pages. ISBN 978-80-02-02522-1.

[2] VDA Volume 2. Quality Assurance for Suppliers – Production process and product approval (PPA). Germany: VDA, the 5th Revised Edition, November 2012. 56 pages. ISBN 978-80-02-02443-9.

[3] Tecklenburg, F. K. G. (2010). Design of Body Assemblies with Distributed Tasks Under the Support of Parametric Associative Design (PAD). Hertfordshire, 339 pages. *PhD Thesis*. University of Hertfordshire. https://reposit.haw-hamburg.de/handle/20.500.12738/2783

[4] Nenadál, J. (2018). *Management kvality pro 21. století*. Prague: Management Press. (in Czech)

[5] Articles. Volkswagenag.com. ©Volkswagen AG 2021. Last change 18.03.2021. Retrieved from: https://www.volkswagenag.com/de/news/2020/04/Volkswagen_supports_its_40_000_suppliers.html.

[6] Řeháček, P. (2013). *Projektové řízení podle PMI*. Prague: Ekopress, 123 pages. (in Czech)

[7] Jurová, M. (2016). *Výrobní a logistické procesy v podnikání*. Prague: Grada Publishing. (in Czech)

[8] Project Management Institute, (2017). *A Guide to the Project Management Body of Knowledge*. The 6th edition. Newtown Square, Pennsylvania, 756 pages.

[9] IPMA®. (2017). *Individual Competence Baseline for Project Management*. Version 4.0. Netherlands, Nijkerk: IPMA, 212 pages.

[10] Doležal, J. (2016). *Projektový management: Komplexně, prakticky a podle světových standardů*. Prague: Grada Publishing. (in Czech)

[11] Bennet, N. & AXELOS (2017). *Managing Successful Projects with PRINCE2*®. The 6th Edition. United Kingdom: TSO, the Stationery Office.

[12] Máchal, P., Kopečková, M., & Presová, R. (2015). *Světové standardy projektového řízení pro malé a střední firmy: IPMA, PMI, PRINCE2*. Prague: Grada. Manažer. (in Czech)

[13] Bentley, C. (2010). The essence of PRINCE2® Project Management Method. The 7th Revised Edition. England, 258 pages.

[14] Swanston, K. (2017). Project Management PRINCE2® in 30 minutes by Karen Swanston. Retrieved from: https://www.youtube.com/watch?app=desktop&v=8-Msk4ff8ew. Chanel ILX Group.

[15] Rastogi, A. (2019). PRINCE2 Vs PMP: How to choose the right credential for your Project Management career? Retrieved from: https://www.greycampus.com/blog/project-management/prince2-vs-pmp-how-to-choose-the-right-credential-for-your-project-management-career.

[16] Knowledge TRAIN, PRINCE2® vs the PMBOK® Guide: A comparison In: www.knowledgetrain.co.uk [online]. © 2005-2021 Knowledge Train®. Retrieved from: https://www.knowledgetrain.co.uk/project-management/pmi/prince2-and-pmbok-guide-comparison.

[17] PRINCE2, APM or PMI? | Choosing project management qualifications. In: www.findcourses.co.uk [online]. Last update: 29.8.2019. Retrieved from: https://www.findcourses.co.uk/inspiration/articles/choosing-a-project-management-qualification-prince2-apm-pmi-9204.

[18] Fiampolis, P. & Acaster, M. (2015). Optimizing project management with a combination of PRINCE2 and PMBOK. IV(XII-12-2015). Retrieved from: https://pmworldlibrary.net/wp-content/uploads/2015/12/pmwj41-Dec2015-Fiampolis-Acaster-opimizing-project-management-advisory.pdf.

[19] Roseke, B. (2021). IPMA vs. PRINCE2: How to Decide Between Them. Retrieved from: https://www.projectengineer.net/ipma-vs-prince2-how-to-decide-between-them/.

[20] AIAG. About AIAG. Retrieved from: https://www.aiag.org/about.

[21] Wab, G. & Wagner, R. (2017). *Projektmanagement in der Automobilindustrie*. The 5th Updated and Revised Edition. Wiesbaden: Springer Gabler, 402 pages. (In German)

[22] Advanced Product Quality Planning and Control Plan [online]. Detroit: AIAG, 2008. 108 pages. ISBN 1605341371.

[23] Stamatis, D. H. (2018). *Advanced product quality planning: the road to success*. Florida - Boca Raton: CRC Press/Tailor & Francis Group. https://doi.org/10.1201/9780429401077

[24] Production Part Approval Process (PPAP). The 4th Edition. Detroit: AIAG, 2006. 69 pages. ISBN 1605340936.

[25] AIAG & VDA FMEA Handbook. The 1st Edition. Detroit: AIAG, 2019. 238 pages. ISBN 1605343676.

[26] Measurement Systems Analysis MSA. The 4th Edition. Detroit: AIAG, 2010. 231 pages. ISBN 1605342114.

[27] Statistical Process Control (SPC). The 1st Edition. Detroit: AIAG, 2005. 221 pages. ISBN 1605341088.

[28] The Cost of Poor Quality Guide. The 1st Edition. Detroit: AIAG, 2012. ISBN 9781605342566.

[29] Layered Process Audit Guideline. The 2nd Edition. Detroit: AIAG, 2014. 29 pages. ISBN 1605343005.

[30] Effective Problem Solving Guide. The 2nd Edition. Detroit: AIAG, 2018. 425 pages. ISBN 1605343994.

Authors' contacts:

Filip Knapp, Mgr. Ing.
(Corresponding author)
University of West Bohemia in Pilsen,
Faculty of Mechanical Engineering,
Univerzitní 22, 306 14 Pilsen, Czech Republic
Tel.: +420 377638401, e-mail: fknapp@fst.zcu.cz

Michal Šimon, doc. Ing. PhD
University of West Bohemia in Pilsen,
Faculty of Mechanical Engineering,
Univerzitní 22, 306 14 Pilsen, Czech Republic
Tel.: +420 377638400, e-mail: simon@fst.zcu.cz

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230504194331

Subject review
Received: 2023-05-04, Accepted: 2023-05-08

Modelling Freight Allocation and Transportation Lead-Time

Aurelija Burinskiene*, Arunas Burinskas

Abstract: The authors have investigated sustainable environment delivery systems and identified transportation lead-time investigation cases. This research study aimed to increase freight delivery lead-time and minimize distance in transportation. To reach the goal, the paper's authors, after analysis of the hierarchy of quantitative methods and models, proposed the framework for modeling freight allocation and transportation lead-time and delivered a study that includes discrete event simulation. During the simulation, various scenarios have been revised. Following the simulation mentioned above analysis, around 3.8 % of distance could be saved during freight delivery if lead-time for transportation were revised by choosing five days criteria for modeling freight allocation. The savings depend on the number of received orders from different geographic locations.

Keywords: delivery system; discrete event simulation; environmental sustainability; freight transportation; lead-time

1 INTRODUCTION

Due to the effect on the environment, it is important to investigate the long lead-time for freight transportation compared to the short one. It began with the topic of environmental sustainability. In the longer period, more research was provided. It broadened our understanding of sustainability, much more than reducing carbon emissions in delivery logistics. In the pick of the topic's popularity, freight logistics was named green and/or gold, suggesting an improved reduction in CO_2 emissions and costs. The simulation of the freight logistics process allows us to see delivery system behavior, which gives higher efficiency.

The study's authors investigated delivery processes and identified wasteful critical ones. As a result, sustainable solutions for freight logistics were determined by applying the discrete event simulation, which was based on the analysis of scientific literature and the application of synthesis methods used in the article.

This study suggests a framework to determine the trade-offs between short and long freight delivery lead-times. The article's authors revise studies on sustainable environment delivery systems seeking to identify contemporary trends and provide investigations on the topic.

The research has a limitation in applying economy of scale when consolidating less than truckload cargo minimizes costs. However, short lead-time allows to gain higher customer satisfaction. Nevertheless, the results are valuable for experts looking for freight transportation distance savings and systematic efficiency improvement. The research's originality is using different lead-time criteria for distance waste elimination.

2 MODELLING FREIGHT ALLOCATION AND TRANSPORTATION LEAD-TIME

The concept of environmental sustainability got greater attention in 1980. The sustainability topic was initially promoted to society, primarily following concerns about environmental problems. To minimize these concerns, most authors investigate sustainability as a tool for reducing carbon emissions. Many authors investigate this type of sustainability and publish papers on it [1-3].

The development of co-modal, intermodal, and multimodal freight delivery systems has been identified as a means of reducing high costs and pollution in freight transportation. Efforts can yield partial solutions to undesirable consequences for the freight transport sector. But more this needs to be done, particularly by reducing pollution in the road haulage sector; this is a major concern for all governments worldwide.

2010-2019 CO_2 emissions from freight transport have risen in all regions except Europe, where they have fallen by 2%, thanks to smart fuel-saving regulations and smart, sustainable urban mobility initiatives. The fastest growth rates were in developing regions and Asia. Asia had the highest emissions in absolute terms by comparing 2019 with 2010: 41 percent in Asia, 27 percent in Africa, 13 percent in Oceania, and 6 percent in North America, etc. Energy demand and CO_2 emissions also continued to rise for all types of road transport means (lorries, mini-vans), particularly fast-growing in heavy goods road transport. Making the road transport sector sustainable also requires the involvement of the private sector and society.

For the modeling freight allocation and transportation lead-time, meta-heuristic methods could be implemented to optimize various processes. Among meta-heuristic methods, genetic algorithms are widely used. In addition to genetic algorithms, meta-heuristic methods include evolutionary algorithms, differential us evolutionary algorithms, and partial optimization. Genetic algorithms are used to plan vehicles. In the works of many scientists, problems with the optimization of transport routes are solved with the help of genetic algorithms.

Many algorithms are used to create transport routes, group them, and improve them. A gigantic and long route is split into shorter routes. However, meta-heuristic methods, including genetic, evolutionary, and differential evolutionary algorithms, are not widely used by scientists to search for the best transport route. In the first table, meta-heuristic methods are already used for which route planning cases are already used and for which cases are not.

Currently, classical algorithms are applied to route planning. With the help of these algorithms, various solutions are sought. One such solution is for the route planning of vehicles of defined and indeterminate capacity. One of the prerequisites for choosing a vehicle with a defined capacity in search of the best route is that the demand is known and fixed.

Meanwhile, a classic algorithm is used with separate delivery and pick-up routes for the delivery and collection of parcels. That is, in the beginning, delivery routes are planned, after which the pick-up (collection) routes are designed accordingly, such as an algorithm called VRPMDP. The transports tool, firstly, carries out routes during which the parcels are delivered to customers, then mixed routes, during which the parcels are delivered and collected from customers served at that time. Two vehicles can come to the customer at different time intervals.

Table 1 Application of meta-heuristic methods for modeling freight allocation and transportation lead-time

Methods	Modeling	Ongoing research includes/does not include
Genetic algorithms (GA)	[4-9]	**Includes:** identifying the time intervals allotted for loading and unloading when planning routes; fuzzy demand in case of route planning; route planning with multiple loading points; route planning for timely delivery JIT for cargo. **Does not** include: transport route planning with flexible delivery time intervals; route formation solutions covering several planning periods; cases of improvement of planned routes during transportation.
Evolutionary algorithms (EA)	[10-12]	**Includes:** historical demand route planning cases and route planning with named delivery time intervals. **Does not include:** cases of increasing the load on transport; the delivery and return of small consignments when planning routes; other cases.
Differential evolutionary algorithms (DE)		**Does not include:** loading and unloading flexible and fixed time windows for loading and unloading when planning routes; the use of vehicles of different capacities; route formation solutions covering different planning periods; other cases.
Partial optimisation (PSO)	[13, 14]	**Includes:** route planning of a vehicle of a defined capacity. **Does not include:** cases of planned route improvements; routing decisions involving the transport of different product groups.

There may also be adjustments during the execution of routes. If a new customer emerges from whom to pick up the cargo, the planned route must be improved to include a new case of picking up the cargo. In addition, there may be cases of cargo delivery and pick-up cancellations when the planned routes need to be adjusted accordingly.

In conclusion, it should be mentioned that heuristics can be used to form routes for the delivery and pick-up of goods. Using heuristics, which can effectively manage various restrictions, such as the number of freight receiving and delivering points, time windows, and the various uses of vehicles, can solve route optimization problems.

Therefore, optimization solutions for planning transport routes, when these involve the use of heuristical methods, contribute to the total number of kilometers planned, reducing pollution and protecting the environment.

3 COMPARISON OF LEAD-TIMES APPROACHES

Mainly in the literature, the authors focus on shorter delivery times and lead-times. The focus on demand and supply constraints in this area is also key, but different from delivery service management, waiting for freight here means the time the customer waits before receiving the required item. Delivery time has been researched in much of the supply chain management literature since the early 1990s.

Many studies have examined strategies for defining appropriate delivery lead-times to increase efficiency [15, 16] or the achievement of the target service levels [17]. Some comments were directed at shortening decisions to shorten delivery lead-times and reduce inconvenience, and theoretical models have explored the benefits of shortening delivery times. One such concept that supports shorter lead-times is JIT [18] (Tab. 2). Some authors have shown the benefits of shortened delivery lead-times, which are sales-driven and linked with demand aspects. Others specifically felt that the customer was sensitive to delivery lead-time.

Interim choices for consumers are decisions about estimating trade-offs between costs and benefits that may occur at different times. These choices are important in psychology, economics, business, and public policy. This flow of literature generally recognizes that consumers prefer shorter freight delivery delays as time for waiting could also mean costs. Thus, consumers accept delays only if a longer waiting time means higher future value. The first models of interim selection viewed that the value of time or its discount factor was fixed and had a positive impact. But contemporary research has shown that many anomalies can violate such principles. In particular, the time discount rate decreases with increasing waiting time. In addition, consumers are more sensitive to time to get lower-value results than higher-value results, which means that consumers are more sensitive to the timing of cheap products.

Delivery lead-time is the main factor for customers making purchase decisions, and the possibility of reducing it gives retailers the ability to improve their competitiveness. The online retail company competes hard by suggesting short lead-times for freight delivery services. Companies competed hard with traditional retailers in the first years of online retailing. However, online retailers have found prompt delivery an important factor in competition. This strategy was so successful that it drastically changed consumer expectations of acceptable delivery lead-time.

However, the concept of environmental sustainability was announced earliest among other sustainability directions, but for the reduction of CO_2 emissions, different solutions were researched. In 2017 [19] presented research focusing on the combination and the sourcing decisions examining the

problem of sustainable economic order quantity (S-EOQ) with the stochastic delivery timing and multimodal transport option. With S-EOQ, sequential solutions focus on order points, which can be affected by various factors, including the price per product, inventory costs, variability of delivery time, and CO_2 emissions. The authors presented an equation for the adjacent S-EOQ problem to have a wider understanding. They investigated the problem by using experiments and different scenarios to identify the inclusion of sustainability aspects in traditional models related to operations-oriented solutions, covering transport mode selection, their combinations, and supply solutions.

Table 2 Literature overview on lead-time concepts

Concepts	1950	1970	1990	2000	2010
JIT production & Quick response manufacturing (QRM) with short lead-time	[20]			[21]	[22]
Effect of lead-time on inventory		[23]	[24]		[25]
Lead-time sensitive demand				[26]	[27]
Sustainable EOQ (S-EOQ) with lead-time variability					[19]

In delivery logistics, conceptual frameworks toward sustainability are still important, and practical named problems are evident today. Both could give a better understanding of sustainability, including various delivery types but focusing on freight delivery by road. In the next section, the authors examine freight delivery lead-time aiming to produce more sustainable practices. The simulation will examine the practices and give guidelines on making freight delivery practices more sustainable.

4 THE STUDIES ON FREIGHT DELIVERY SYSTEM

Delivery is a physical process for freight transportation from a source to a destination. However, this activity does not include the modeling of freight allocation – it is the area where CO_2 emissions are generated, even if the client covers the delivery costs. The authors expand the concept of delivery by including therein decision-making approach. In case of CO_2 emissions reduction, clients shall be informed that a longer lead-time could be assigned to their orders.

In this research, the authors analyzed freight delivery services in different aspects: shorter and longer lead-time investigations before deciding on freight allocation for delivery.

Over time, researchers have been concerned about analyzing delivery systems. The approach to such systems has developed fundamentally new concepts and methods, mainly in mathematics and science. In operations research, applying such methods is premature but has great potential to analyze the freight transportation aspects. In principle, the authors investigate the system, whether resources are allocated efficiently, and whether there are appropriate conditions for optimal resource allocation to maximize output. For researching such problems, various methods were applied and defined in Tab. 3.

Table 3 The hierarchy of quantitative methods and models for investigating freight delivery lead-time

Types	Modeling technique	Solution methods	Authors investigating the topic
Mathematical programming methods	Single-objective	Bi-level linear programming (LP)	[28]
	Multi-objective	Multi-objective mixed integer linear programming (MILP)	[29]
		Fuzzy-goal programming	[30]
		Polynomial dynamic programming	[31]
		Queuing model	[32]
		Non-linear programming	[29]
Simulation methods	System dynamic (SD)		[33]
	Discrete event (DES)		[34]
Heuristic methods	Simple heuristic	Simulated annealing heuristics (SAH)	[35]
	Artificial intelligence (Al) techniques	Markov chains	[36]
		Object-oriented Petri nets	[37]
		Bayesian network modeling	[38]
		Fuzzy logic	[38]
		Artificial Neural network	[39]
		Grey system and rough sets	[40]
	Meta-heuristic	Genetic Algorithm (GA)	[41]
		Evolutionary Algorithm (EA)	[42]
		Differential evolution algorithm (DEA)	[43]
		Particle swarm optimization (PSO)	[44]
		Ant Colony Optimization	[45]
		Greedy Randomised Adaptive search procedure	[46]
Hybrid model	Hybrid simulation	SD-DES	[47]
Analytical model	Multi criteria decision making (MCDM)	Analytical hierarchy process (AHP)	[48]

To revise methods useful for problem-solving, a hierarchical view can be applied to identify the type and technique of the model and specific methods in a category. Five main categories are used, starting from the mathematical programming category, moving to model methods, revising heuristic and hybrid models, and finalizing with analytical

models. Modeling methods are different in nature and relate to one or several purposes.

Many studies focus on the application of optimizing approaches. In addition, different methods have been used in the delivery system, such as multi-objective mixed integer linear programming, system dynamics, discrete events, inequality of variation, fuzzy logic, Markov chains, etc. However, also we could identify methods that are not yet explored and applied in studies.

For freight application, several route methods are used, such as first transport planner plans route and then cluster clients ("route first-cluster second"), or does this in the opposite way – first cluster clients and after plans route ("cluster first-route second"). Several researchers delivered simulation papers focusing on the topic. For decision-making aiming to reduce CO_2 emissions, it is necessary to include the aspect of lead-time as one day, it could be just one client per geographic area, and the planner could not cluster together. Further on, the authors developed a framework that could help allocate freight with lead-time modification.

5 FRAMEWORK FOR MODELLING FREIGHT ALLOCATION AND TRANSPORTATION LEAD-TIME

During the day, many orders (freights) are placed and delivered to customers as a single one into a geographic territory. In this case, the company asks customers to specify the urgency of their order delivery, which is not optimal. If possible, enterprise delivers freights with as possible lower lead-time duration.

The delivery plan of freights is the schedule where each customer announces the delivery timing information. According to the schedule, 10 minutes per order and 3 minutes per pallet are usually selected. The authors have chosen the shortest path approach to develop a delivery schedule following the optimization results. The model covers only freight movements between the customers following the vector basis.

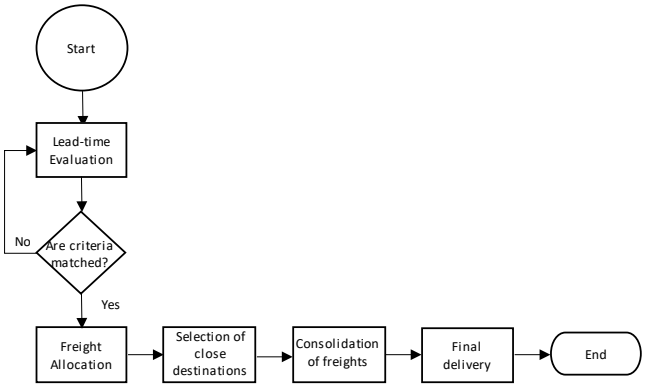


Figure 1 Adaptation of algorithm steps for freight transportation

In Fig. 1, the authors present algorithm steps. The first possibility shows the situation of lead-time evaluation. If there are orders from customers that match the same destination, their freights are allocated to the same group. But this doesn't represent the shortest distance choice. Under the next step, we must select freight that fits the closest destination criteria.

Let's assume that the customer usually places an order (n) which has its weekly volume (the number of transport units).

A freight's delivery routes are constructed each day for the next day's deliveries, but lead-time could be increased to reduce CO_2 emission for non-lead-time sensitive freights. The author suggests including steps helping to identify when freight has to be allocated to the route. The current freight allocation algorithm was modified to reach the goals.

The new framework includes several components (Fig. 1):

1) The lead-time evaluation;
2) Freight allocation with lead-time modification;
3) Selection and consolidation of freights.

The constructed strategic freights' delivery schedule is later adjusted daily (Fig. 2). According to the schema presented in Fig. 2, freight allocation is modified to reach the optimization. Each customer's delivery frequency and volume are provided and not investigated herein. In operational conflict cases, the priority status is allocated to the freights with critical needs.

Further on, the share among all freights delivered under the same truck is calculated as follows (1):

$$x_{mn} = \frac{q_{mn}}{\sum_{m=1}^{w} q_{mn}} \tag{1}$$

herein m – customer, w – all customers, q – volume in handled units, x – share of freight of m in the truck n (in percentage). The counter is the volume of a specific customer, and the denominator is the max number of pallets under the order or loading bill if it is received upfront the actual delivery. If the freight has no volume within the truck frame, its share will equal zero.

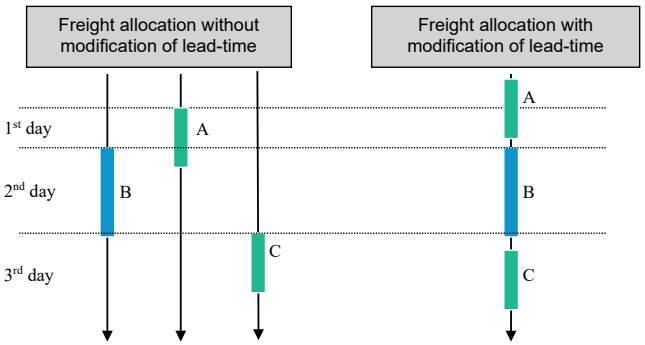


Figure 2 Schema of consolidation possibilities (A, B, C – different orders)

If all freights allocated to the truck, match its maximum capacity when all freight of the customers could be consolidated. The schema for consolidation is presented in Fig. 2.

In Fig. 2, there is the presentation of both frameworks:

1) Freight allocation without modification of lead-time (current freight allocation algorithm) (on the left);

2) Freight allocation with lead-time modification (the suggested framework) (on the right).

Let's assume that the customer places the order and which delivery time is one day. We have to add a waiting time of between 1 and 4 days to prolong lead-time. Such could be formulated by using Eq. (3).

The authors suggest a matrix for the distance parameter. Herein, the distance represents point-to-point (P2P) distances, which are vector or Euclidean based on a Steiner – Weber model.

The next step is to optimize delivery, which depends on freight (i.e., order) volume with limited unloading time.

The total freight distance S_i^n is calculated by the Eq. (2):

$$S_i^n = \sum_{j=1}^{y} d_{ij} \tag{2}$$

Herein d – the shortest distance, i – route, j – order, y – all orders in the route.

The total delivery lead-time T_t consists of 2 parts: T_o – physical delivery lead-time and T_g – waiting time (as specified in Eq. (3):

$$T_{tj} = T_{oj} + T_{gj};\ g = [1, v] \tag{3}$$

Herein T_t – total delivery lead-time, o – physical delivery lead-time in days, j – order, g – number of waiting days, v – maximum waiting limit in days.

The framework allows waiting time to be defined on an order level.

For simulation, various lead-time scenarios will be used to revise the results and benchmark them with scenarios where the waiting time equals zero.

For the case study, the three months of data are used from a 3PL company operating in the USA, which had 11 440 orders. The data source attributes are loading location, delivery location, order date, and quantity. The main relationship measured during the calculations is expressed as the duration of lead-time and the distance (Fig. 3).

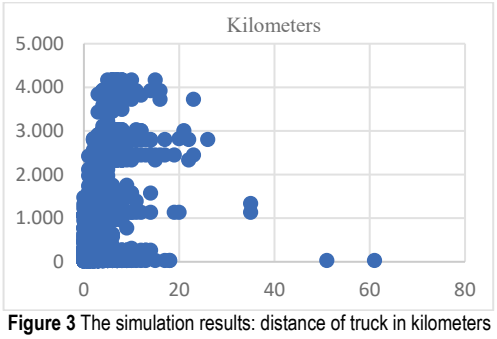


Figure 3 The simulation results: distance of truck in kilometers

As side indicators, the utilization level of trucks and warehouses and the CO_2 emission level will be determined. The scenarios of analysis will consider cargo consolidation between different periods, such as one day, two days, three days, four days, and five days. The authors estimated that 25 percent of freights are delivered as single ones.

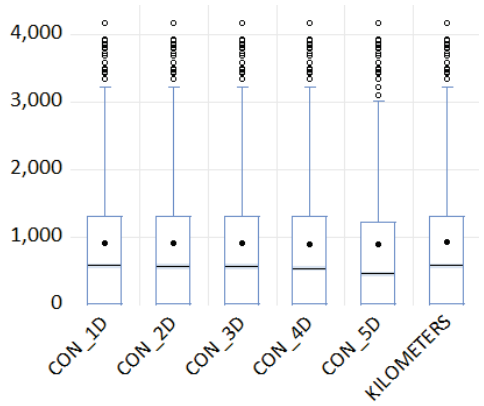


Figure 4 The simulation results: distance according to the lead-time scenarios

The authors formed a simulation model with Wolfram Matematica software, while results were processed with EViewer statistical software. Using a simulation package, various scenarios were compared to lead-time duration lasting from 1 to 5 days. In Fig. 4, the authors presented simulation results (see Fig. 4).

The benchmarking of simulation results for five days and the current scenario shows possible distance improvement. The authors figured out that the longest lead-time duration (i.e., five days) helps to save up to 3.8 % kilometers. Further, the authors suggest working out the freight delivery plan following geographical territory.

6 CONCLUSION

The theme of sustainability got special attention after 1980, focusing on environmental sustainability. However, the results of CO_2 emissions are still not as good in the road transport sector. Later, papers focusing on different lead-time approaches were investigated from 1950, when the short lead-time concept was implemented with JIT production implementation. The aspect of sustainability was not integrated into delivery studies, except in one study, which highlighted different transport modes and more sustainable ones but with longer lead-time duration.

The authors apply many qualitative methods and models for investigating delivery lead-time seeking to build a sustainable environment delivery system. However, using these methods results in resource efficiency and the implementation of freight transportation processes supporting higher efficiency.

The authors presented a case study where distance waste was eliminated and identified freight delivery possibilities during discrete event simulation. The results allow us to avoid waste and get a reduction in distance. This shows possibilities for organizing a delivery plan to have a more sustainable freight delivery system. The research results showed that to reduce distance activity, the company has to re-design the freight allocation process by adding waiting time into freight delivery. Following the scenarios-based simulation study, around 4% of the distance could be

minimized by modifying lead-time to five days. These improvements could be reached if the suggested framework is applied to the delivery management system. Most actions as the revision of customers' delivery plans and their implementation, are important to reach the expected results. A deeper understanding of the features of freight delivery is important to form and apply solutions to increase efficiency in freight transportation. Such an approach focuses on distance waste elimination in the area and allows it to reach benefits through the implementation of efficient resource usage.

Acknowledgements

The project received funding from the European Regional Development Fund (project No 01.2.2-LMT-K-718-03-0030) under the grant agreement with the Research Council of Lithuania (LMTLT).

7 REFERENCES

[1] Fichtinger, J., Ries, J. M., Grosse, E. H., & Baker, P. (2015). Assessing the environmental impact of integrated inventory and warehouse management. *International Journal of Production Economics, 170*, 717-729. https://doi.org/10.1016/j.ijpe.2015.06.025

[2] Caniato, F., Caridi, M., Crippa, L., & Moretto, A. (2012). Environmental sustainability in fashion supply chains: An exploratory case based research. *International Journal of Production Economics, 135*(2), 659-670. https://doi.org/10.1016/j.ijpe.2011.06.001

[3] Marchant, C. (2010). Reducing the environmental impact of warehousing. In McKinnon, A. (eds.), *Green logistics: improving the environmental sustainability of logistics*, 167-192. Replika Press Pvt Ltd, India.

[4] Pu, X., Lu, X., & Han, G. (2022). An improved optimization algorithm for a multi-depot vehicle routing problem considering carbon emissions. *Environmental Science and Pollution Research, 29*(36), 54940-54955. https://doi.org/10.1007/s11356-022-19370-0

[5] Agrawal, A. K., Yadav, S., Gupta, A. A., & Pandey, S. (2022). A genetic algorithm model for optimizing vehicle routing problems with perishable products under time-window and quality requirements. *Decision Analytics Journal, 5*, 100139. https://doi.org/10.1016/j.dajour.2022.100139

[6] Cosma, O., Pop, P. C., & Sitar, C. P. (2022). A two-level based genetic algorithm for solving the soft-clustered vehicle routing problem. *Carpathian Journal of Mathematics, 38*(1), 117-128. https://doi.org/10.37193/CJM.2022.01.09

[7] Olaniyi, O. S., James, A. K., Ibrahim, A. A., & Makanjuola, A. F. (2022). On the Application of a Modified Genetic Algorithm for Solving Vehicle Routing Problems with Time Windows and Split Delivery. *IAENG International Journal of Applied Mathematics, 52*(1), 1-9.

[8] Liu, Q., Xu, P., Wu, Y., & Shen, T. (2022). A hybrid genetic algorithm for the electric vehicle routing problem with time windows. *Control Theory and Technology, 20*(2), 279-286. https://doi.org/10.1007/s11768-022-00091-1

[9] Ahmed, Z. H., Hameed, A. S., & Mutar, M. L. (2022). Hybrid Genetic Algorithms for the Asymmetric Distance-Constrained Vehicle Routing Problem. *Mathematical Problems in Engineering, 4*, 18-25. https://doi.org/10.1155/2022/2435002

[10] Erbao, C. & Mingyong, L. (2009). A hybrid differential evolution algorithm to vehicle routing problem with fuzzy demands. *Journal of computational & applied mathematics, 231*(1), 302-310. https://doi.org/10.1016/j.cam.2009.02.015

[11] Tan, K. C., Cheong, C. Y., & Goh, C. K. (2007). Solving multiobjective vehicle routing problem with stochastic demand via evolutionary computation. *European Journal of operational research, 177*(2), 813-839. https://doi.org/10.1016/j.ejor.2005.12.029

[12] Tan, K. C., Chew, Y. H., & Lee, L. H. (2006). A hybrid multiobjective evolutionary algorithm for solving vehicle routing problem with time windows. *Computational Optimization & Applications, 34*(1), 110-115. https://doi.org/10.1007/s10589-005-3070-3

[13] Ai, T. J., and Kachitvichyanukul, V. (2009). Particle swarm optimization & two solution representations for solving the capacitated vehicle routing problem. *Computers & Industrial Engineering, 56*(1), 380-387. https://doi.org/10.1016/j.cie.2008.06.012

[14] Marinakis, Y. & Marinaki, M. (2013). Combinatorial neighborhood topology particle swarm optimization algorithm for the vehicle routing problem. In *European Conference on Evolutionary Computation in Combinatorial Optimization*, 133-144. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37198-1_12

[15] Lee, Y. H. & Kim, S. H. (2002). Production–distribution planning in supply chain considering capacity constraints. *Computer Industrial Engineering, 43*, 169-190. https://doi.org/10.1016/S0360-8352(02)00063-3

[16] Lim, S. J., Jeong, S. J., Kim, K. S., & Park, M. W. (2006). A simulation approach for production–distribution planning with consideration given to re-plenishment policies. *Int. J. Advanced Manufacturing Technology, 27*, 593-603. https://doi.org/10.1007/s00170-004-2208-2

[17] Bilgen, B. & Ozkarahan, I. (2007). A mixed-integer linear programming model for bulk grain blending and shipping. *International Journal of Production Research, 107*, 555-571. https://doi.org/10.1016/j.ijpe.2006.11.008

[18] Yang, J. S. & Pan, J. C. H. (2004). Just-in-time purchasing: an integrated inventory model involving deterministic variable lead time and quality improvement investment. *International Journal of Production Research, 42*(5), 853-863. https://doi.org/10.1080/00207540310001632448

[19] Lee, S. K., Yoo, S. H., & Cheong, T. (2017). Sustainable EOQ under lead-time uncertainty and multimodal transport. *Sustainability, 9*(3), 476. https://doi.org/10.3390/su9030476

[20] Lehekar, M. P. & Urne, M. (1950). Implemetation of Kanban. *International Journal of Trend in Scientific Research and Development, 1*(6), 1-4. https://doi.org/10.31142/ijtsrd4616

[21] De Treville, S., Shapiro, R. D., & Hameri, A. P. (2004). From supply chain to demand chain: the role of lead time reduction in improving demand chain performance. *Journal of operations management, 21*(6), 613-627. https://doi.org/10.1016/j.jom.2003.10.001

[22] Godinho Filho, M. & Saes, E. V. (2013). From time-based competition (TBC) to quick response manufacturing (QRM): the evolution of research aimed at lead time reduction. *The International Journal of Advanced Manufacturing Technology, 64*(5), 1177-1191. https://doi.org/10.1007/s00170-012-4064-9

[23] Das, C. (1975). Effect of lead time on inventory: a static analysis. *Journal of the Operational Research Society, 26*(2), 273-282. https://doi.org/10.1057/jors.1975.61

[24] Ben-Daya, M. A., & Raouf, A. (1994). Inventory models involving lead time as a decision variable. *Journal of the Operational Research Society, 45*(5), 579-582.

https://doi.org/10.1057/jors.1994.85
[25] Liao, T. W. & Chang, P. C. (2010). Impacts of forecast, inventory policy, and lead time on supply chain inventory—a numerical study. *International Journal of Production Economics, 128*(2), 527-537. https://doi.org/10.1016/j.ijpe.2010.07.002
[26] Yang, B. & Geunes, J. (2007). Inventory and lead time planning with lead-time-sensitive demand. *Iie Transactions, 39*(5), 439-452. https://doi.org/10.1080/07408170600838456
[27] Pekgün, P., Griffin, P. M., & Keskinocak, P. (2017). Centralized versus decentralized competition for price and lead-time sensitive demand. *Decision Sciences, 48*(6), 1198-1227. https://doi.org/10.1111/deci.12259
[28] Gupta, S., Chatterjee, P., Yazdani, M., & Gonzalez, E. D. S. (2021). A multi-level programming model for green supplier selection. *Management Decision, 12*, 24-30. https://doi.org/10.1108/MD-04-2020-0472
[29] Munoz-Villamizar, A., Santos, J., Montoya-Torres, J. R., & Velazquez-Martinez, J. C. (2020). Measuring environmental performance of urban freight transport systems: A case study. *Sustainable Cities and Society, 52*, 101844. https://doi.org/10.1016/j.scs.2019.101844
[30] Çifçi, G. & Büyüközkan, G. (2011). A fuzzy MCDM approach to evaluate green suppliers. *International Journal of Computational Intelligence Systems, 4*(5), 894-909. https://doi.org/10.1080/18756891.2011.9727840
[31] Eksioglu, S., Geunes, J., Palak, G., & Azadi, Z. (2014). Analyzing the Impact of Carbon Regulatory Mechanisms on Supply Chain Management (No. 2012-078S).
[32] Misni, F. & Lee, L. S. (2017). A review on strategic, tactical and operational decision planning in reverse logistics of green supply chain network design. *Journal of Computer and Communications, 5*(8), 83-104. https://doi.org/10.4236/jcc.2017.58007
[33] Carvalho, H., Duarte, S., & Machado, V. C. (2011). Lean, agile, resilient and green: divergencies and synergies. *International Journal of Lean Six Sigma, 3*, 17-28. https://doi.org/10.1108/20401461111135037
[34] Arnau, Q., Barrena, E., Panadero, J., de la Torre, R., & Juan, A. A. (2022). A biased-randomized discrete-event heuristic for coordinated multi-vehicle container transport across interconnected networks. *European Journal of Operational Research, 302*(1), 348-362. https://doi.org/10.1016/j.ejor.2021.12.035
[35] Kumar, R., Ganapathy, L., Gokhale, R., & Tiwari, M. K. (2020). Quantitative approaches for the integration of production and distribution planning in the supply chain: a systematic literature review. *International Journal of Production Research, 58*(11), 3527-3553. https://doi.org/10.1080/00207543.2020.1762019
[36] Kleywegt, A. J., Nori, V. S., & Savelsbergh, M. W. (2002). The stochastic inventory routing problem with direct deliveries. *Transportation science, 36*(1), 94-118. https://doi.org/10.1287/trsc.36.1.94.574
[37] Bevilacqua, M., Ciarapica, F. E., & Mazzuto, G. (2012). Modelling and performance analysis of a supply chain using timed coloured Petri nets. *International Journal of Business Performance and Supply Chain Modelling, 4*(3-4), 285-316. https://doi.org/10.1504/IJBPSCM.2012.050386
[38] Prakash, S., Kumar, S., Soni, G., Jain, V., & Rathore, A. P. S. (2020). Closed-loop supply chain network design and modelling under risks and demand uncertainty: an integrated robust optimization approach. *Annals of operations research, 290*(1), 837-864. https://doi.org/10.1007/s10479-018-2902-3
[39] Comert, S., Yazgan, H., & Turk, G. (2022). Hopfield neural network based on clustering algorithms for solving green vehicle routing problem. *International Journal of Industrial Engineering Computations, 13*(4), 573-586. https://doi.org/10.5267/j.ijiec.2022.6.002
[40] Agbo, A. A. & Zhang, Y. (2017). Sustainable freight transport optimisation through synchromodal networks. *Cogent Engineering, 4*(1), 1421005. https://doi.org/10.1080/23311916.2017.1421005
[41] Fahimnia, B., Davarzani, H., & Eshragh, A. (2018). Planning of complex supply chains: A performance comparison of three meta-heuristic algorithms. *Computers and Operations Research, 89*, 241-252. https://doi.org/10.1016/j.cor.2015.10.008
[42] Boonsothonsatit, K., Kara, S., Ibbotson, S., & Kayis, B. (2015). Development of a Generic decision support system based on multi-Objective Optimisation for Green supply chain network design (GOOG). *Journal of Manufacturing Technology Management, 2*, 15-24. https://doi.org/10.1108/JMTM-10-2012-0102
[43] Cui, L., Deng, J., Zhang, Y., Tang, G., & Xu, M. (2020). Hybrid differential artificial bee colony algorithm for multi-item replenishment-distribution problem with stochastic lead-time and demands. *Journal of Cleaner Production, 254*, 119873. https://doi.org/10.1016/j.jclepro.2019.119873
[44] Venkatesan, S. P. & Kumanan, S. (2010). Bi-criteria allocation of customers to warehouses using a particle swarm optimisation. *International Journal of Operational Research, 9*(1), 65-81. https://doi.org/10.1504/IJOR.2010.034361
[45] Goli, A., Tirkolaee, E. B., & Weber, G. W. (2020). A perishable product sustainable supply chain network design problem with lead time and customer satisfaction using a hybrid whale-genetic algorithm. In *Logistics operations and management for recycling and reuse* (pp. 99-124). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33857-1_6
[46] Hajipour, V., Tavana, M., Di Caprio, D., Akhgar, M., & Jabbari, Y. (2019). An optimization model for traceable closed-loop supply chain networks. *Applied Mathematical Modelling, 71*, 673-699. https://doi.org/10.1016/j.apm.2019.03.007
[47] Hrušovský, M., Demir, E., Jammernegg, W., & Van Woensel, T. (2018). Hybrid simulation and optimization approach for green intermodal transportation problem with travel time uncertainty. *Flexible Services and Manufacturing Journal, 30*(3), 486-516. https://doi.org/10.1007/s10696-016-9267-1
[48] Low, Y. S., Halim, I., Adhitya, A., Chew, W., & Sharratt, P. (2016). Systematic framework for design of environmentally sustainable pharmaceutical supply chain network. *Journal of Pharmaceutical Innovation, 11*(3), 250-263. https://doi.org/10.1007/s12247-016-9255-8

Author's contacts:

Aurelija Burinskiene, Associate professor
(Corresponding author)
Vilnius Gediminas Technical University,
Sauletekio Avenue 11, 10223 Vilnius, Lithuania
E-mail: aurelija.burinskiene@vilniustech.lt
aurelija.burinskiene@vgtu.lt

Arunas Burinskas, Associate professor
Vilnius University,
Sauletekio Avenue 9, 10223 Vilnius, Lithuania
E-mail: arunas.burinskas@evaf.vu.lt

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230505103326

Subject review
Received: 2023-05-05, Accepted: 2023-05-09

Determining Learning Outcomes Relevant for Logistics Higher Education on Sustainability and Industry 4.0

Tina Cvahte Ojsteršek, Simona Šinko*, Brigita Gajšek

Abstract: The ambitious goals of the European Union require companies to transform themselves into sustainable and smart. To do so, for example, in the logistics field, employees who will know how to implement green transport and logistics and how to establish smart systems based on a higher level of digital maturity are needed. This article lists 51 essential topics from Sustainability, Industry 4.0, Logistics 4.0, and Digitalisation areas, further supported by 127 unique learning outcomes gathered through a multi-stage international Delphi study including practitioners and academics. Experts from 6 countries participating in the study first identified the most important topics from the field and the connected learning outcomes in the following rounds, and consequently achieved consensus on the most important learning outcomes that the future workforce in the logistics sector should have. Supporting the logistics sector with a workforce educated based on this framework should help it reach the European Union's strategic goals.

Keywords: digitalisation; higher education; Industry 4.0; learning outcomes; logistics; sustainability

1 INTRODUCTION

Among the six defined priorities for 2019-2024 of the European Commission are a European Green Deal, digital transformation, and empowering people with the skills necessary for a new generation of technologies [1]. The goal of the European Commission is also to assist all industrial sectors in harnessing emerging technologies and facilitating the shift towards an intelligent Industry 4.0 industrial system [2]. In order to achieve the European Union's stated mission, companies will need to focus on sustainability, digitalisation, and Industry 4.0. The growing need to develop these areas is also reflected in the scientific literature, as briefly summarised below.

The heightened awareness of economic, environmental, and social issues has brought sustainability into the spotlight. As a result, customer purchasing behaviours have shifted towards sustainable products from companies that prioritise eco-friendliness, social welfare, and financial prosperity. This shift has motivated supply chains to prioritise sustainability initiatives and Industry 4.0 technologies [3,4].

Sustainable development refers to creating solutions that satisfy society's current needs without compromising future generations' ability to meet their own needs [5]. The three sustainability dimensions, environmental, economic, and social, are frequently discussed in the literature [6,7]. Achieving balance in these processes is crucial for businesses to be sustainable and prosperous. However, due to these dimensions' intricate nature and interdependence, attaining this balance and success can be challenging [8].

Industry 4.0 was frequently mentioned after 2010 as essential to Germany's Action Plan High-tech Strategy 2020 [9] and represents the 4th Industrial Revolution. Although Industry 4.0 has substantially impacted the industry since its emergence, especially in the manufacturing sector, the current priorities are focused on the digital transformation of the economy. In practice, companies have a low level of digital maturity, which is a prerequisite for their active participation in Industry 4.0.

Hermann et al. [10] defined Industry 4.0 as a set of technologies and ideas related to organising the value chain that emerged during the fourth industrial revolution. The core components of Industry 4.0, according to the authors, are Cyber-physical systems (CPSs), Internet of Things (IoT), Internet of Services (IoS), and Smart Factory, along with design principles such as interoperability, virtualisation, decentralisation, real-time capability, service orientation, and modularity. The fundamental goal of the Industry 4.0 movement is to provide factories and businesses with strategic tools for offering new types of services and business models, revolutionising the concept of the production chain, and progressively transforming traditional factories into smart factories [11,12]. Although only some companies are mature enough to implement Industry 4.0, there is already talk of Industry 5.0. Industry 5.0 is a new concept that aims to combine the strengths of machines and human experts to improve the manufacturing industry. It is expected to bring back the human touch to manufacturing by combining the critical thinking of humans with the high speed and accuracy of machines. Industry 5.0 will enable mass personalisation, increase manufacturing efficiency and improve production quality. Collaboration between people and machines will lead to high-speed production, where repetitive tasks are assigned to machines, and critical thinking tasks are assigned to people [13].

All concepts mentioned above are also relevant to logistics, which is the focus of this article. Sustainable logistics is a vital sub-component of the Green Supply Chain Management (GSCM) process, a popular topic in recent decades due to globalisation, market competition, customers' demands, and the exploration of new markets [14]. On the other side, the advancements in Industry 4.0 technologies have brought about benefits in logistics operations, resulting in the emergence of Logistics 4.0. This leads to improved efficiency in logistics activities, such as transportation, inventory management, and material handling [15].

To effectively implement sustainability, Industry 4.0, and digitalisation concepts in reality, competent employees are needed. Logistics students and other logistics participants need to understand and be able to implement sustainability, Industry 4.0 technologies and digitalisation to meet future challenges in logistics.

Logistics students or employees in the logistics sector often overlook sustainable development issues because they primarily focus more on the engineering aspects of logistics. However, this is not following modern development of the logistics field and Logistics 4.0 initiatives [16]. In order to get a sufficient number of logistics graduates that will be well-versed future sustainability leaders, HEIs need to improve learning experiences [17] and monitor the needs of the economy and adapt the curricula and learning outcomes that students need to acquire.

Recognising the significance of integrating sustainability subjects into education, including higher education, has been extensive [18]. As Leich and co-authors [19] claim, education is the starting point for helping students to engage with sustainability-related issues and become change-makers. Teaching only from an economic perspective or promoting only economic growth can potentially lead to unsustainable consumption patterns. Sustainability needs to be integrated into the educational process horizontally, especially content aimed toward systemic and holistic thinking based on the principles and approaches of sustainable development [20].

Besides understanding sustainability, it is important for logisticians to continually enhance their knowledge and stay up-to-date with recent innovations [21] and with concepts of Industry 4.0. From this point of view, it is necessary to update educational content with sustainability, Industry 4.0, and digitalisation content, and establish lifelong education [22]. Short educational programs accessible to logistics students and logistics sector employees would also be welcome.

However, an analysis of the literature on Industry 4.0 education and institutional innovation suggests that there are limited studies that systematically examine the crucial institutional factors required for an Industry 4.0 response toward educational innovation, change, or reform [23]. Boston Consulting Group [24] summarises the nine pillars of Industry 4.0 that need to be addressed in educational programs: advanced robotics, additive manufacturing, augmented reality, simulation, horizontal/vertical integration, industrial internet, cloud, cybersecurity, big data, and analytics. Romero-Gazquez and co-authors [22] developed a specific module for teaching the basic concepts of Industry 4.0, which includes the following thematic fields: autonomous robots, simulation, system integration, internet of things, cybersecurity, cloud computing, additive manufacturing, augmented reality, and big data. Some managerial and technological implications that present themselves when sustainability and circular economy are at the forefront were defined by Khan and co-authors [25] as digital platforms, e-learning solutions, artificial intelligence, and machine vision. Various models have been suggested for sustainability competencies, and the framework proposed by Wiek and co-authors [26] for key competencies in sustainability is one of the most frequently cited. Burndiers and co-authors [27] did a Delphi study to provide insights into a consensus statement on critical sustainability competencies. The research aimed to promote sustainability and related programs at higher education institutions across the United States.

Two critical questions should be addressed when deciding on a curriculum position or focus, independent of the field of study. The first is, "What are the main takeaways for students regarding knowledge and understanding from this course? What content counts, then?" The second question is, "How should it be taught?" [28]. Considering the first question and the definition of learning outcomes described below, it is essential to focus on learning outcomes that are important for sustainable logistics and Industry 4.0. Learning outcomes are statements about what the student should know, understand and be able to demonstrate after completing the learning process. Learning outcomes focus on the student's achievements and not only on learning content [29].

A scientific literature review was expected to not sufficiently describe which learning outcomes are relevant in connecting logistics with sustainability, Industry 4.0, and green, which was confirmed by the presented literature overview. This presents a research gap that needs to be overcome if educational efforts are to be designed to cater to the industry's needs and identified research highlights. The most significant contribution of the study is that it reveals learning outcomes that have been shown in practice to be necessary for working in Logistics 4.0 concerning sustainability. To the best of the authors' knowledge, this is the first time anyone has explored relevant learning outcomes in this area with a methodology that combines researchers, academics, and practitioners in a single study on an international level.

In order to find out which learning outcomes are essential for logistics students to cope successfully with future logistics challenges, a Delphi study was designed and undertaken with the final goal of identifying the topics and contributing learning outcomes referring to teaching sustainable logistics and Industry 4.0.

In the following chapters, the paper briefly summarises the main features of the Delphi studies that were followed in preparing the present study. A detailed description of the steps of the study itself follows. In the Results, we present which topics in logistics, sustainability, green, and Industry 4.0 experts consider important for students to listen to during their studies. The learning outcomes that emerged from the study as the most important are also presented. The paper concludes with findings and a conclusion.

2 METHODS AND METHODOLOGY

A Delphi study was designed to determine key and essential learning outcomes for logistics higher education on sustainability and Industry 4.0. The Delphi methodology is a valuable tool for researching the future of education as it can help to explore the underlying assumptions that lead to different judgments [30]. The Delphi method systematically combines opinions by a group of experts to reach a consensus on a complex problem [31] in an asynchronous way [32]. Structured anonymous communication among individuals is highlighted in the Delphi method [33]. It is valuable for scenarios where individual opinions must be gathered and consolidated to tackle a lack of consensus or an incomplete understanding of a subject [34].

The Delphi method is a group technique in which a researcher or research team brings together a panel of experts, presents questions, consolidates feedback, and

guides the group toward a consensus. Unlike traditional survey methods that aim to determine means and generalise results across a population through online, in-person, or mail surveys, the Delphi method involves an iterative process similar to a long series of focus groups [35]. The outcome of a Delphi study is usually evidenced by the agreement among experts in a specific area where there was previously none [36].

The typical procedure involved in the Delphi process is as follows [32]: (1) identifying and defining the problem or topic to be addressed in the process, (2) designing the process from an operational perspective, (3) selecting experts to participate in the process, (4) creating a questionnaire that initially allows for open-ended responses but becomes progressively more restricted in subsequent rounds, (5) conducting at least two rounds of the process, evaluating the results after each round, and adapting the questionnaire in between rounds, (6) using the results to achieve a consensus on the topic.

A different variation of the Delphi method involves creating a questionnaire that permits respondents to provide comments alongside their answers. The questionnaire remains constant throughout all rounds of the Delphi process, but each subsequent round incorporates a summary of the experts' comments and reasoning for their responses. This allows experts to review and revise their opinions based on new arguments presented in the succeeding rounds of the Delphi process. Typically, this variation is carried out over two rounds [37].

The Delphi panel's makeup should not rely on random sampling or statistical sampling methods. Instead, it should be a deliberate choice based on the research question and consist of experts who are knowledgeable and reputable in the field of Delphi [38]. Furthermore, the panel's composition should be kept confidential from the participating expert panelists, and direct communication between them should be avoided. Panelists and their views should be fully anonymised [37]. After the first round of the Delphi process, which typically includes open-ended questions, the responses should be analysed using content analysis to consolidate them. Aggregation and synthesis can then be employed to generate input for the second round [38]. Despite being mainly qualitative, the Delphi method can also incorporate some quantification, particularly in the second and subsequent rounds. This can be achieved by using 5-point or 7-point Likert scales to gauge experts' opinions on statements derived from the initial open-ended questions [32].

2.1 Delphi Implementation to Elaborate Learning Outcomes Relevant for Logistics Higher Education on Sustainability and Industry 4.0

The study presented in this article was implemented as a part of the Erasmus+ project: Sustainable Logistics4.0: Digital and green skills for boosting innovation and sustainability of the logistics sector. It includes four universities: Politechnika Poznanska (Poland), Gaziantep Universitesi (Turkey), Universidade de Aveiro (Portugal), and University of Maribor (Slovenia), and three companies Valuedo srl (Italy), ECQA (Austria) and Zerynth srl (Italy).

Based on the research aims, a multi-round Delphi study was designed that began with open ended questions and progressed to more concrete outcomes in subsequent rounds. It had a national and an international level. Hence, two rounds on each level were sufficient for the desired level of consensus among the included experts. The first step for each round was selecting potential experts, followed by the questionnaire design. After the questionnaire design was finished and confirmed by all project partners, the round of Delphi was conducted. After individual rounds that included open-type questions, the researchers read and analysed the open-ended responses independently. The research team then discussed the responses to modify and consolidate each contribution. When using the Likert scale, the results were analysed with the use of descriptive statistics. The objectives of a particular Delphi round are presented in Fig. 1. Objectives and detailed implementation of the study are described in more detail below.

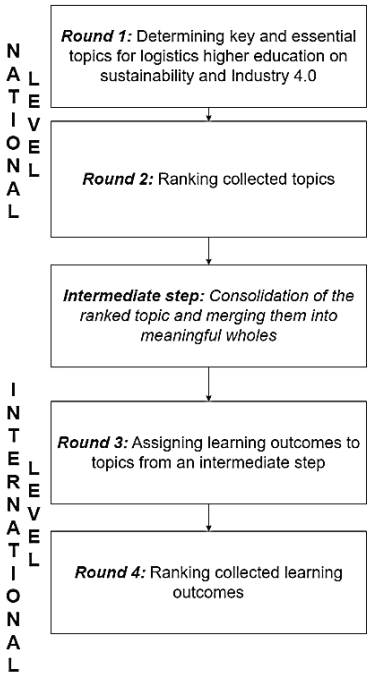


Figure 1 Delphi study process

2.2.1 Expert Panel

During the preparation phase, an international group of experts was gathered. The expert panellists were selected from academic and company representatives to ensure a broad consensus on the final results. Briedenhann and Butts [39] argue that including both professional and academic experts is essential for achieving a balance between different perspectives on knowledge and bridging the gap between research and professional communities in terms of knowledge sharing, communication, priorities, etc. The Delphi method does not require the expert panel to be representative regarding statistical sampling. Rather than the number of experts, the quality of the panel members determines their representativeness [35]. In addition, having a geographically diverse panel of experts was crucial for capturing a wide range of perspectives on the discussed topic.

Inclusion criteria mandated for the academics or teachers were that they had published at least two papers on logistics and/or sustainability in the last five (5) years or have taught a course from this field at least a high school level in the last five (5) years. Practitioners from companies were included in the case that they work for a company that provides logistics services or from companies that have logistically complex processes (e.g., a large production company). In all four rounds of the Delphi method, we included a sufficient number of experts (Table 1). In all four rounds, we included more practitioners than academics. Practitioners provided valuable insights into the needs of the economy. Participation in the Delphi study was completely anonymous.

Table 1 Number of experts included in expert panels

		Academics	Practitioners	Sum
National level	First round	32	36	68
	Second round	25	31	56
International level	First round	9	19	28
	Second round	7	9	16

2.2.2 Delphi Implementation

The research employed an online Delphi approach. The appeal of using the internet for communications and data collection was due to the time and cost savings it provided, which were particularly beneficial given the global nature of the study and the conventional constraints imposed on the Delphi methodology [40]. The Delphi approach used quantitative and qualitative questions, which have been demonstrated to improve research rigor [41].

2.2.3 National Level of the Delphi Study

Researchers from four universities individually performed the national level of the Delphi study. In this part, four expert panels were held in parallel: in Poland, Portugal, Slovenia, and Turkey. The process and questions for the panellists were the same for all panels to get quality input into the international round of the Delphi study. The questionnaires prepared were translated into the national languages. The recommendation given to all researchers was to prepare both an online and a Word version of the questionnaire and send both versions to participants to increase the chances of getting a response since some people prefer the modality of answering a questionnaire. The national-level expert panels have been implemented at different times - depending on the capacity to recruit panellists. Implementations occurred between 8th December 2022 and 15th January 2023.

The main idea behind the first round at the national level was to determine key and essential topics for logistics, sustainability, and Industry 4.0 and which will have to be taught in the future. Therefore, experts were asked open-ended questions. The analysis of the individual results was qualitative to extract all the topics of interest in a manner of potential teaching topics. In the second national round, the collected topics were ranked. Experts evaluated topics based on the importance of educating future logisticians. A 10-point Likert-type scale (not important at all – 1 score, essential – 10 score) was used to assess the participant's level of agreement. Each research group did a basic analysis of the results. Means and standard deviation were calculated for each topic regarding the importance ratings from the panellists. This presented the final conclusion of the national-level Delphi study.

All four inputs of the identified learning topics or course contents from the national level of Delphi were consolidated into a final list of proposed topics for the international round of Delphi. The consolidated topics were then input into the third round of Delphi, held at the international level of Delphi.

2.2.4 International Level of the Delphi Study

An international group of researchers performed the international round of the Delphi study. Panelists were experts from Slovenia, Poland, Portugal, Turkey, India, and Mexico. In Delphi round 3 (held between 3rd and 20th of February 2023), the experts were asked to propose the learning outcomes that students should achieve learning about a specific topic, where the input topics were the highest-rated topics from Delphi round 2 on the national level. Proposed learning outcomes by panelists were then consolidated and presented back to the panelists in the Delphi round 4 (held between 20th and 31st of March 2023). The panelists rated learning outcomes in Delphi round 4 according to the importance of learning outcomes. A 10-point Likert-type scale (not at all important – 1 score, very important/crucial – 10 score) was used to assess the participant's level of agreement. Means and standard deviation were calculated for each learning outcome regarding the importance ratings from the panellists. The final result of the Delphi on the international level is thus the list of essential learning outcomes for future logisticians.

3 RESULTS

The results are presented in two subsections, following the described Delphi methodology. First, the national study results are presented, and then the international study results present essential learning outcomes relevant to logistics higher education.

3.1 National Level of the Delphi Study

The consolidated list of topics from all four expert panels in national Delphi studies had 147 topics. Tab. 2 presents only consolidated and merged 19 topics that were found to be most relevant. The decision to present only the highest-rated topics was made because the list of topics was too extensive. As seen in the table, topics were divided according to the study areas of this article: digitalisation, Industry 4.0, and Sustainability. A specific topic for "green" was added because this is where most of the effort in the sustainability area is currently being invested. Another addition was the area of "basic knowledge" to see if the experts also proposed any topics related to the basic knowledge of logisticians. The most significant number of topics is from the field of

digitalisation (7), followed by green (5). Three topics each have been identified in the teaching areas of Sustainability and Industry 4.0.

Table 2 Topics in the areas of logistics, sustainability, and Industry 4.0 that experts consider the most relevant

Topic	M	SD	Teaching Area
Human-Centric Digital Transformation	9.5	2.07	I4.0
SAP system	9.5	1.08	D
Information management. Construction of information models. Data analysis.	9.5	1.56	B
Reverse and green logistics	9.47	1.07	G
Green packaging/packaging	9.37	0.95	G
Digitalisation of document flows and transition to paperless logistics operation	9.33	1.18	D
Collection and processing of large amounts of data and analytics, knowledge of databases, advanced use of data editing and analysis programs (e.g., Excel)	9.33	0.85	D
Digitalisation of business operations, ERP, programs and applications to support logistics activities	9.33	0.94	D
Green storage	9.31	1	G
Green transport	9.31	1	G
Circular economy and waste management	9.26	0.99	S
Autonomous warehouse equipment	9.26	0.99	I4.0
Optimisation in logistics (process optimisation, routes, supply chains) to achieve sustainability goals, optimisation tools	9.08	1.26	S
Green purchasing	9.05	0.97	G
Autonomous solutions in logistics (storage, transport). Autonomous vehicles and means of internal transport	9	1.48	I4.0
Digitalisation of logistics processes	9	1.78	D
Database	9	1.56	D
Big data in order to, for example, profile the image of the customer or predictive analyses to examine the supply chain in order to minimise threats	9	2.08	D
Quality control	9	1.56	B
Disruptions, Risk Analysis & Mitigation Strategies	9	1.74	B
Environmental Impact of Logistics Operations Tools	9	2.01	S

*B – basic logistics knowledge, D – topics related to the Digitalisation, G – topics related to Green, I4.O – topics related to the Industry 4.0, S – topics related to Sustainability

Other relevant topics (evaluated with a mean value between 8.0 and 8.99) were topics about environmental ethics (S), emission calculation methods (S), product lifecycle analysis (S), 17 sustainable development goals (S), circular economy and waste management (S), sustainable warehousing and logistics (S), sustainable operations management (S), internet of things (D), virtual and augmented reality (I4.0), cybersecurity (D), digitalisation of communication (D), electronic documentation systems (D), data management (D), machine learning (I4.0), automation and robotics (I4.0), IoT Technologies (I4.0), connected information systems and collaborative platforms (D), route optimisation (B), simulation tools and procedures for simulation of logistics processes (B), logistics planning (B), production planning (B), logistic centres (B), reverse logistics (B), modelling of business processes (B), legal regulations (B), goods tracking systems (RFID systems) (D), management of storage devices (I4.0), resource management (B), dark warehouse (I4.0), smart logistics (I4.0), work-life balance (S), labour law and respect for employee rights (S). Other identified topics (86) were evaluated with a mean value between 5 and 8. Among all identified topics, only one was considered insignificant with a score lower than 5 (M: 4.5, SD: 2.84): Cultural, economic, and political forces influencing environmental attitudes and decision-making based on understanding science and technology.

All defined topics were analysed and consolidated since some were listed several times from different textual records. The result was a shortened list of topics that were entered into the third round of the Delphi study performed on an international level. A final list included 51 topics that were considered essential for future logisticians. Topics were then divided into seven different thematic chapters (Tab. 3): (1) Industry 4.0 and Logistics 4.0 concepts, (2) Digitalization and Logistics 4.0 maturity models, (3) Sustainability in Logistics 4.0, (4) Economic sustainability in Logistics 4.0, (5) Social sustainability in Logistics 4.0, (6) Environmental sustainability in Logistics 4.0 and (7) Industrial Internet of things for green logistics.

Table 3 Final list of topics essential for future logistician

Thematic chapters	Addressed topics
Industry 4.0 and Logistics 4.0 concepts	• Industry 4.0 definition and potential for implementation as Logistics 4.0 • Autonomous warehouse equipment • Autonomous vehicles and mobile robots • Robots and cobots in logistics • Automatic truck unloading systems • Artificial intelligence and machine learning in logistics • Drone applications • Virtual and augmented reality
Digitalization and Logistics 4.0 maturity models	• Digital economy concept • Human-Centric Digital Transformation • Connected Information Systems and Collaborative Platforms (ERP and MES in logistics) • Digitalisation of logistics processes • Digital maturity • IT security and cybersecurity, blockchain • Digitalisation of document flows and electronic data interchange • Multi-criteria support for decision-making processes in logistics (e.g. route optimisation) • Goods tracking and tracing systems on the basis of different identification approaches • Data management • Digital twin technology
Sustainability in logistics 4.0	• 17 sustainable development goals • Circular economy • Regulatory and policy frameworks for sustainable logistics • Integration of social, environmental, and economic sustainability issues in logistics • Environmental ethics • Sustainable tangible fixed assets in logistics (land, buildings, equipment and spare parts, small inventory)
Economic sustainability in logistics 4.0	• Economic analysis of sustainable interventions in logistics • Low-carbon economy and climate-neutral economy/logistics • Triple Bottom Line

Table 3 Final list of topics essential for future logistician (continuation)

Thematic chapters	Addressed topics
Social sustainability in Logistics 4.0	• Social aspects of sustainability in logistics • Work-life balance • Corporate social responsibility in logistics
Environmental sustainability in Logistics 4.0	• Reverse logistics • Green and sustainable packaging • Green and sustainable transport • Green and sustainable storage and warehousing • Green and sustainable intralogistics • Emission and carbon footprint calculation methods • Life cycle analysis • Renewable energy sources in logistics • Waste management and recycling systems in logistics • Energy saving in logistics processes • Green and sustainable purchasing • Sustainable Operations Management • Green and sustainable city logistics
Industrial Internet of things for green logistics	• (Industrial) Internet of Things (IoT) and technologies (e.g., cloud computing) • Adopting IIoT in the logistics sector • Designing intelligent logistics systems • Simulation tools and procedures for logistics processes • Optimisation tools and procedures for logistics processes • Project management • Teamwork

3.2 International Level of the Delphi Study

Based on the final list of topics from the national level Delphi study, the panelists were asked to propose essential learning outcomes that future logisticians will need to achieve under each topic on the list to cope with a green and smart future. As is seen in Tab. 4, panelists determined 607 learning outcomes in Delphi round 3. Most learning outcomes were suggested for the thematic chapter Industry 4.0 and Logistics 4.0 concepts. The fewest were defined for the chapter Economic sustainability in logistics 4.0. After combining the same learning outcomes propositions, a list of 127 unique learning outcomes was defined and used in the Delphi round 4 on the international level. The list of learning outcomes was shortened to 79.08% reporting essential learning outcomes for future logisticians.

Table 4 Number of learning outcomes across different thematic chapters

Thematic chapter	Number of learning outcomes determined in Delphi round 3	Number of learning outcomes after consolidation
Industry 4.0 and Logistics 4.0 concepts	160	25
Digitalization and Logistics 4.0 maturity models	123	28
Sustainability in Logistics 4.0	62	13
Economic sustainability in Logistics 4.0	26	8
Social sustainability in Logistics 4.0	37	8
Environmental sustainability in Logistics 4.0	125	32
Industrial Internet of things for green logistics	74	13
Sum	607	127

Table 5 Average importance of learning outcomes devided between course modules

Course modules	Number of learning outcomes	M	SD
Logistics 4.0 in the smart society	21	7.79	1.41
Sustainability in Logistics 4.0.	61	8.06	1.34
IIoT: adopting 4.0 for green logistics	39	7.7	1.51
Green challenge and practical applications	6	8.3	1.5

Table 6 Higher-ranked learning outcomes per course module

Learning outcomes	M	SD
Logistics 4.0 in the smart society		
Is able to identify 4.0 technologies and their applicability to the Logistics 4.0 cases.	8.3	0.99
Can assess the maturity of a logistics department/company for Industry 4.0 using the Industry 4.0 maturity model.	8.2	1.25
Has knowledge of the requirements and infrastructure necessary for the implementation of different autonomous warehouse equipment	8.2	1.42
Can identify the benefits and risks of usage of different types of autonomous warehouse equipment in a warehouse and describe their functions.	8.2	1.09
Is able to describe the need for tracking and traceability along the supply chain and outline implementation alternatives.	8.2	1.2
Sustainability in Logistics 4.0.		
Can critically judge transport modes in sense of green and sustainable goals.	8.7	1.4
Can plan green and sustainable transport based on the comparison of alternatives.	8.7	1.49
Can locate and measure waste in a supply chain and establish its reuse/recycling.	8.6	1.11
Can learn and follow national/international legislations for sustainable logistics.	8.6	0.87
Can plan green and sustainable storage and warehousing by following a methodology.	8.6	0.94
IIoT: adopting 4.0 for green logistics		
Can plan digital technology (digital platforms, software, and other technologies) to automate and optimise the various logistics processes such as transportation planning, inventory management, and order fulfilment.	8.3	1.26
Can plan, collect, receive, organise, control and use data with the support of appropriate tools.	8.3	1.36
Can define what artificial intelligence and machine learning are and describe how they can be applied to improve logistics operation.	8.2	1.48
Can identify potential and possibilities of artificial intelligence and machine learning applications in logistics processes.	8.2	1.51
Is able to identify logistics processes that can benefit from digitalisation.	8.2	1.51
Is able to find solution providers and choose the most suitable one among them.	8.2	1.47
Green challenge and practical applications		
Implements teamworking.	8.9	1.22
Can manage human relations.	8.7	1.1
Can prepare a project plan (cost, time, resources, risk, changes management, quality, WBS, list of activities).	8.4	1.54
Can monitor and direct project implementation.	8.2	1.44
Knows multi criteria decision making tools and can apply at least one.	7.9	1.78

After consolidation, the learning outcomes were divided into four (4) course modules (Tab. 5). All 127 learning outcomes were retained. In Delphi round 4 on the international level, each learning outcome was rated on a scale from 1 (not at all important) to 10 (very important).

After the panel finished, the average learning outcomes' importance rating was calculated for all modules (Tab. 5). On

average, all learning outcomes scored as important. The result confirms that the determined learning outcomes are important for future logisticians and that appropriate consensus was reached among all of the Delphi panellists.

Tab. 6 lists five higher-ranked learning outcomes per course module. The lowest average score (6.8) received the learning outcome "can describe the structure of a digital twin for a specific logistics system" from course module IIoT: adopting 4.0 for green logistics.

4 CONCLUSIONS

The ambitious goals of the European Union require companies to transform themselves into sustainable and smart companies. Regarding sustainability, the highest goals are set in the area of greenness. Smart companies, however, will be based on digital maturity and implemented elements of Industry 4.0. The mentioned drivers have been particularly strengthening since 2010. During this time, however, not much has changed in higher education. We are witnessing pilot applications of topics from sustainability, greenness, and Industry 4.0 and exceptional practices. However, we cannot claim that the spirit of transition has already comprehensively encompassed the entire academic space.

There are few attempts to involve international groups of experts from different scientific disciplines in a focused thinking about adding sustainability and Industry 4.0 content to HE courses. This research has succeeded in doing so for the field of logistics.

In the scientific literature, there is much talk about which knowledge and competencies are necessary but almost nothing about learning outcomes, which define what the participant in education knows at the end of the educational process. In this regard, we view the study as original.

This article lists 51 essential topics from Sustainability, Industry 4.0, Logistics 4.0, and Digitalisation areas, further supported by 127 unique learning outcomes. Experts from 6 countries believe that the future workforce in the logistics sector should be able to use them in practice. Supporting the logistics sector with a workforce educated based on this framework should help it to reach the strategic goals of the European Union. However, simply adding these learning objectives to regular study programs will not be enough. The research also supports efforts to establish lifelong learning for logistics professionals who have acquired their education in past eras.

There are several limitations of this research. The number of participating experts decreased over the Delphi rounds. The main obstacle was the length of the questionnaires. Furthermore, we allow the possibility that we have omitted a topic or learning outcome. If more experts were included, it would very likely further improve the research results.

Acknowledgements

This research is funded by the Erasmus+ project Sustainable Logistics 4.0: Digital and green skills for boosting innovation and sustainability of the logistics sector [Slog4.0]

5 REFERENCES

[1] European Union. (n. d.). European Union priorities 2019-2024. Accessed: Apr. 03, 2023 on https://european-union.europa.eu/priorities-and-actions/eu-priorities/european-union-priorities-2019-2024_en

[2] European Parliament. (2015, September). Industry 4.0: Digitalisation for productivity and growth. Accessed: Apr. 03, 2023 on https://www.europarl.europa.eu/RegData/etudes/BRIE/2015/568337/EPRS_BRI(2015)568337_EN.pdf

[3] Ageron, B., Gunasekaran, A., & Spalanzani, A. (2012). Sustainable supply management: An empirical study. *Int. J. Prod. Econ.*, *140*(1), 168-182. https://doi.org/10.1016/j.ijpe.2011.04.007

[4] Toktaş-Palut, P. (2022). Analyzing the effects of Industry 4.0 technologies and coordination on the sustainability of supply chains. *Sustain. Prod. Consum., 30*, 341-358. https://doi.org/10.1016/j.spc.2021.12.005

[5] WCED - World Commission on Environment and Development. (1987). Report of the world commission on environment and development. Accessed: Apr. 03, 2023 on https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf

[6] Husgafvel, R., Pajunen, N., Virtanen, K., Paavola, I.-L., Päällysaho, M. Inkinen, V., Heiskanen, K., Dahl, O., & Ekroos, A. (2015). Social sustainability performance indicators – experiences from process industry. *Int. J. Sustain. Eng*., 8(1), 14-25. https://doi.org/10.1080/19397038.2014.898711

[7] Li, W., Alvandi, S., Kara, S., Thiede, S., & Herrmann, C. (2016). Sustainability Cockpit: An integrated tool for continuous assessment and improvement of sustainability in manufacturing. *CIRP Annals, 65*(1), 5-8. https://doi.org/10.1016/j.cirp.2016.04.029

[8] Yildiz Çankaya, S., & Sezen, B. (2019). Effects of green supply chain management practices on sustainability performance. *J. Manuf. Technol., 30*(1), 98-121. https://doi.org/10.1108/JMTM-03-2018-0099

[9] Federal Ministry of Education and Research. (2010). Ideas. Innovation. Prosperity. High-Tech Strategy 2020 for Germany. Ideas innovation prosperity. Bonn: Postfach.

[10] Hermann, M., Pentek, T., & Otto, B. (2016). Design Principles for Industrie 4.0 Scenarios. In *The 49th IEEE Hawaii International Conference on System Sciences (HICSS)*, Koloa, HI, USA, 3928-3937. https://doi.org/10.1109/HICSS.2016.488

[11] Mourtzis, D., Siatras, V., Angelopoulos, J., & Panopoulos, N. (2020). An Augmented Reality Collaborative Product Design Cloud-Based Platform in the Context of Learning Factory. *Procedia Manuf., 45*, 546-551. https://doi.org/10.1016/j.promfg.2020.04.076

[12] Thames, L. & Schaefer, D. (2016). Software-defined Cloud Manufacturing for Industry 4.0. *Procedia CIRP, 52*, 12-17. https://doi.org/10.1016/j.procir.2016.07.041

[13] Maddikunta, P. K. R., Pham, Q. V., Prabadevi, B., Deepa, N., Dev, K., Gadekallu, T. R., ... & Liyanage, M. (2022). Industry 5.0: A survey on enabling technologies and potential applications. *Journal of Industrial Information Integration*, *26*, 100257. https://doi.org/10.1016/j.jii.2021.100257

[14] Isaksson, K., Björklund, M., Evangelista, P. & Huge-Brodin, M. (2011). The challenge and adoption of green initiatives for transport and logistics service providers. *The 16th Annual LRN Conference 7-9 September*, Southampton, United Kingdom.

[15] Strandhagen, J. O., Vallandingham, L. R., Fragapane, G., Strandhagen, J. W., Stangeland, A. B. H., & Sharma, N. (2017). Logistics 4.0 and emerging sustainable business models. *Adv. Manuf.*, 5(4), 359-369. https://doi.org/10.1007/s40436-017-0198-1

[16] Mulder, K. F., Segalàs, J., & Ferrer-Balas, D. (2012). How to educate engineers for/in sustainable development: Ten years of discussion, remaining challenges. *Int. J. Sustain. High. Educ*., *13*(3), 211-218. https://doi.org/10.1108/14676371211242535
[17] Dyer, G. & Dyer, M. (2017). Strategic leadership for sustainability by higher education: the American college and university presidents' climate commitment. *J. Clean. Prod.*, 140, 111-116. https://doi.org/10.1016/j.jclepro.2015.08.077
[18] Cortese, A. D. (2003). The critical role of higher education in creating a sustainable future. *Plan High Educ,* 31, 15-22.
[19] Leicht, A., Heiss, J., & Byun, W. J. (2018). *Issues and trends in education for sustainable development*. Paris: United Nations Educational, Scientific and Cultural Organization.
[20] Lukman, R. K., Omahne, V., el Sheikh, L. T., & Glavič, P. (2021). Integrating Sustainability into Logistics Oriented Education in Europe. *Sustainability, 13*(4), 1667. https://doi.org/10.3390/su13041667
[21] Kovács, G. & Spens, K. M. (2011). Trends and developments in humanitarian logistics – a gap analysis. *Int. J. Phys. Distrib. Logist. Manag.*, *41*(1), 32-45. https://doi.org/10.1108/09600031111101411
[22] Romero-Gazquez, J. L., Canavate-Cruzado, G., & Bueno-Delgado, M.-V. (2022). IN4WOOD: A Successful European Training Action of Industry 4.0 for Academia and Business. *IEEE Trans. Educ*., *65*(2), 200-209. https://doi.org/10.1109/TE.2021.3111696
[23] AlMalki, H. A., & Durugbo, C. M. (2023). Evaluating critical institutional factors of Industry 4.0 for education reform. *Technol. Forecast. Soc. Change, 188*, 122327. https://doi.org/10.1016/j.techfore.2023.122327
[24] Boston Consulting Group. (2016). *Sprinting to Value in Industry 4.0: Perspective from and Implications for U.S. Manufacturers.* Accessed: Apr. 03, 2023 on https://www.bcg.com/publications/2016/lean-manufacturing-technology-digital-sprinting-to-value-industry-40 https://www.slideshare.net/TheBostonConsultingGroup/sprinting-to-value-in-industry-40
[25] Khan, S. A., Laalaoui, W., Hokal, F., Tareq, M., & Ahmad, L. (2022). Connecting reverse logistics with circular economy in the context of Industry 4.0. *Kybernetes*. https://doi.org/10.1108/K-03-2022-0468
[26] Wiek, A., Withycombe, L., & Redman, C. L. (2011). Key competencies in sustainability: a reference framework for academic program development. *Sustain. Sci., 6*(2), 203-218. https://doi.org/10.1007/s11625-011-0132-6
[27] Brundiers, K. et al. (2021). Key competencies in sustainability in higher education—toward an agreed-upon reference framework. *Sustain. Sci.*, *16*(1), 13-29. https://doi.org/10.1007/s11625-020-00838-2
[28] Scarff Seatter, C. & Ceulemans, K. (2017). Teaching Sustainability in Higher Education: Pedagogical Styles that Make a Difference. *Can. J. High. Educ., 47*(2), 47-70. https://doi.org/10.47678/cjhe.v47i2.186284
[29] Dias, D. (2020). *Learning Outcomes in European Higher Education: The International Encyclopedia of Higher Education Systems and Institutions*. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-8905-9_317
[30] Cross, V. (1999). The same but different: a Delphi study of clinicians' and academics' perceptions of physiotherapy undergraduates. *Physiotherapy*, *85*(1), 28-39.
[31] Linstone, H. A. & Turoff, M. (1975). *The Delphi method: techniques and applications*. London, United Kingdom: Addison-Wesley.
[32] Profillidis, V. A. & Botzoris, G. N. (2018). *Modeling of Transport Demand*. Elsevier.
[33] Dalkey, N. & O. Helmer. (1962). An experimental application of the Delphi method to the use of experts. *Manag. Sci., 9*, 458-467. https://doi.org/10.1287/mnsc.9.3.458
[34] Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group Techniques for Program Planning: A Guide to Nominal and Delphi Processes*. Glenview, IL: Scott, Foresman and Co.
[35] Powell, C. (2003). The Delphi technique: myths and realities: Myths and realities of the Delphi technique. *J. Adv. Nurs., 41*(4), 376-382. https://doi.org/10.1046/j.1365-2648.2003.02537.x
[36] Sackman, M. (1975). *Delphi critique*. Lexington: Lexington Books.
[37] Williamson, K. (2002). *Research Methods for Students, Academics and Professionals: Information Management and Systems* (Topics in Australasian Library and Information Studies). Chandos Publishing.
[38] Olsen, A. A., Wolcott, M. D., Haines, S. T., Janke, K. K., & McLaughlin, J. E. (2021). How to use the Delphi method to aid in decision making and build consensus in pharmacy education. *Curr. Pharm. Teach. Learn., 13*(10), 1376-1385. https://doi.org/10.1016/j.cptl.2021.07.018
[39] Briedenhann, J. & Butts, S. (2006). The application of the Delphi technique to rural tourism project evaluation. *Curr Issues Tour, 9*(2), 171-190. https://doi.org/10.1080/13683500608668246
[40] Hung, K. & Law, R. (2011). An overview of Internet-based surveys in hospitality and tourism journals. *Tour. Manag., 32*(4), 717-724. https://doi.org/10.1016/j.tourman.2010.05.027
[41] Hasson, D., Theorell, T., Wallén, M. B., Leineweber, M. C, & Canlon, B. (2011). Stress and prevalence of hearing problems in the Swedish working population. *BMC Public Health, 11*(1), 130. https://doi.org/10.1186/1471-2458-11-130

Authors' contacts:

Tina Cvahte Ojsteršek, PhD
University of Maribor, Faculty of Logistics,
Mariborska cesta 7, 3000 Celje, Slovenia
E-mail: tina.cvahte@um.si

Simona Šinko
(Corresponding author)
University of Maribor, Faculty of Logistics,
Mariborska cesta 7, 3000 Celje, Slovenia
E-mail: simona.sinko@um.si

Brigita Gajšek, PhD
University of Maribor, Faculty of Logistics,
Mariborska cesta 7, 3000 Celje, Slovenia
E-mail: brigita.gajsek@um.si

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230505103326

Subject review
Received: 2023-05-05, Accepted: 2023-05-09

Logistics 4.0: Challenges, Opportunities and Threats

Maciej Bielecki

Abstract: Climate change, the main accelerator of which has been consumerism and the industry driving it, soon will force companies to make changes that will intensify their activities in the natural environment. The Covid-19 pandemic, Due to their specificity, has disrupted global supply chains on an unprecedented scale. In addition, the progress of the analysis supported by the assumptions and tools of Industry 4.0 (I4.0) opens unlimited development opportunities. These changes also have a significant impact on Logistics, making the concept of Logistics 4.0 (L4.0) increasingly popular. The scope of defining L4.0 varies from the use of individual I4.0 tools to the presentation of complex models. This paper presents a literature review using the STAR method on the L4.0 concept to identify the scope of its determinants and possible perspectives for the development of logistics in the context of I4.0 and the challenges mentioned.

Keywords: Industry 4.0; Logistics 4.0; STAR

1 INTRODUCTION

The subject of L4.0 is a consequence of the emergence of the concept of I4.0, which was presented in 2011 as a strategic initiative of the German government in terms of the development of the sphere of high technologies and industrialization. It can be assumed that it is therefore a logical continuation of expanding the I4.0 concept to include other functional areas of the company.

The very concept of I4.0 often occurs in parallel with the term the 4th Industrial Revolution (4IR). It significantly changes the existing paradigms in the field of innovation, which by combining technologies blurs the boundaries between the physical, biological, and digital worlds [1]. Visible growth in collecting and storing data about everyone and everything, their processing, supported by ubiquitous telecommunications processes, opens new development opportunities not only for industry, but also for societies [2].

Federico (et al.) indicated 21 new technologies, which they defined as the I4.0 technological lever and included: Internet of Things (IoT), Smart Products (SP), Smart Machines (SM), Cyber-Security & Blockchain (CSB), Artificial Intelligence (AI), Automation (At), Big Data Analytics (BDA), Cloud Technologies (CT), Machine to Machine Communication (M2M), Radio Frequency Identification (RFID), Sensors Technologies (ST), Digitalization (Dt), Robotics (Rb), Omni Channel (OC), Optimization Systems (OS), Business Intelligence (BI), 3D-Printing (3Dp), Mobile Apps (MA), Enterprise Resources Planning (ERP), Augmented Reality (AR) and Nanotechnology (Nt). Roblek (et al.) under 4.0 also pointed to the technologies Internet of Services (IoS), Cyber-Physical Systems (CPS), Human to Human (H2H) and Human to Machine (H2M) communication via the Internet, Integrated Customer Relationship Management (ICRM) – CRM and social media (SocM) [3]. Zheng (et al.) added, Additive Manufacturing (AM), Virtual Reality (VR), Simulation and Modeling (S&M) and Visualization Technology (VT) combined with 3D-Printing (3-DP) [4]. Extending this issue with Autonomic Vehicles (AV) or the increasingly developing drone technology, it can be said that technologically the world has become ready to change the existing paradigms in the sphere of managing functional areas of the organization, including logistics.

The presented I4.0 technologies show great application potential. However, it should be noted that their application possibilities are currently limited only by the technological imperfection of 3-DP and the operating paradigms applicable in enterprises, based mainly on the second and third industrial revolution. Perhaps the third decade of the 21st century is the moment when you can question the existing, selected laws of industry functioning in favor of the opportunities created by I4.0 technologies.

2 MATERIAL AND METHODS

In the era of such a significant digitization of literature, scientific and research work has been greatly facilitated in terms of access to scientific publications. Virtually anywhere in the world where there is access to the Internet, a researcher can browse and analyze many publications. And it is this number of publications that is often a big problem. It involves the proper selection and analysis of large collections of publications, which would be impossible in the aspect of a traditional literature review. A valuable method that makes it possible to solve this problem is a literature review.

There are many types and models of literature review. Booth, Sutton and Papaioannou [5] proposed, for example, the SALSA (Search, Appraisal, Synthesis, Analysis) model, which can take various forms depending on the types of literature review.

Table 1 Selection of the type of literature review [5]

SALSA Stages	STAR	SYSTEMATIC
SEARCH	Comprehensive research of recent literature	Exhausting
APPRAISAL	Lack	Possible
SYNTHESIS	Narrative or tabular	Narrative or tabular
ANALYSIS	Current state of knowledge, indications for future research, as well as their limitations	Current state of knowledge, practical recommendations

From the point of view of the topic of the discussed issue, as well as the research questions, both the State-of-The-Art review and the Systematic search and review approach can be applied here (Tab. 1).

In the article, it was decided to use the STAR research method, as it corresponds to the subject of the issue and the purpose of the presented work. The STAR method allows for a review of the latest, at the same time setting further research directions [6].

While reviewing the literature using the STAR method, the primary research scheme proposed by Cooper was used, consisting of the following steps:

- Formulating the research problem.
- Data collection:
- Selection of databases in which publications will be searched.
- Specify keywords and search restrictions.
- Identification of the collection of publications.
- Selection of publications within a separate set [7].
- Data evaluation.
- Analysis and interpretation.
- Presentation of the results.

The research problem was formulated using five basic research questions presented below.

1) How many publications are currently available that have the term L4.0 in the title?
2) In what scientific areas is the concept of L4.0 defined?
3) How is the concept of L4.0 defined?
4) What general models of L4.0 appear in the analyzed literature?
5) What challenges, opportunities and threats are faced by L4.0?

To answer the first four research questions, a literature review using the STAR method, the theoretical basis of which was described above. The answer to the fifth research question will be an original analysis and interpretation resulting from the STAR literature review.

The second stage was data collection. Among the deliberately selected databases in which the search for publications was started, there were two most popular bibliometric and abstract databases, i.e., Scopus (SCPS) and Web of Science (WoS). As part of the database search process, the following criteria were formulated:

- The word "Logistics 4.0" should appear in the title of the publication.
- The time horizon of the searched databases was 8 years (since 2015, when the first publications with keywords according to the Scopus and Web of Science databases began to appear).
- Type of sources: scientific journals, electronic books, proceedings papers.
- The publication language is English.
- Publications had full access under Open Access, Full Text Finder, or Full Text Open Access.

The collected database of publications made it possible to proceed to the next stages, i.e., data evaluation, analysis, and interpretation, which will be presented in the next chapter.

3 RESULTS

The evaluation and analysis of the collected data was carried out in terms of the criteria listed in Chapter 2. Then, an additional data evaluation and analysis sheet was created, the task of which was to indicate quantitatively the titles containing the key word in the analyzed databases, separately for publications in scientific journals and books, and for publications published under Proceedings Papers. The total number of publications identified in bibliometric and abstract databases is presented in Tab. 2.

Table 2 Number of publications in scientific journals and books containing the word "Logistics 4.0" from 2015 to 2022

	2015	2016	2017	2018	2019	2020	2021	2022	Total
WoS	1	0	1	1	5	8	11	20	47
SCPS	0	0	1	0	4	13	19	25	62

Table 3 Scope of publications in scientific journals and books containing the word "Logistics 4.0" from 2015 to 2022

SCPS			WoS
Engineering	28	39	Engineering Industrial, Manufacturing, Electrical Electronic, Multidisciplinary, Mechanical, Chemical, Civil, Environmental
Business, Management and Accounting	23	26	Management, Business, Operation Research Management Science
Computer Science	17	9	Computer Science Information Systems, Artificial Intelligence, Interdisciplinary Application, Theory and Methods, Telecommunications
Environmental Science	8	27	Environmental Sciences, Studies, Green Sustainable Science Technology,
Decision Sciences	10	1	Statistics Probability
Social Sciences	10	1	Psychology Applied,
Economics, Econometrics and Finance	7	1	Economics
Materials Science	7	0	Materials Science Multidisciplinary,
Physics and Astronomy	6	5	Physics Applied,
Energy	5	0	No Similar
No Similar	0	3	Transportation, Transportation Science Technology,
No Similar	0	7	Chemistry Multidisciplinary, Analytical,
No Similar	0	4	Others Instruments Instrumentation, Ergonomics, Mathematics,

Another element of the research was to specify the thematic scope of the discussed publications. For this purpose, not fully compatible scope from the analyzed databases were used, e.g., in the SCPS database there was the thematic scope of Engineering, while in the WoS database the thematic scope of Engineering was additionally defined, e.g., Engineering Industrial or Engineering Manufacturing. In this way, related thematic SCPS were assigned, also showing those areas that had no equivalents, e.g., Energy.

This meant that in the SCPS database, there was a publication that was classified in the Energy thematic scope, while in the WoS database, there was no such publication. In a similar way, thematic areas related to Transportation, Chemistry, Ergonomic or Mathematics appeared in the WoS database, the equivalents of which could not be found in the Scopus database.

It should also be noted that each of the titles could be assigned to several thematic areas, therefore the data presented in Tab. 2 do not coincide quantitatively with the data presented in Tab. 3. The results of literature research also show that the largest number of titles containing a keyword appear in the areas (in the case of the WoS database, the cut-off point of 80/20 appears here - the Pareto principle) and Environmental Science (in the case of the SCPS database, the distribution is slightly more linear, and the areas of Decision Science, Social Science or Economics should be added). The thematic scope of publications in scientific journals and books with the title "Logistics 4.0" is presented in Tab. 3.

Table 4 Number of publications in proceedings papers containing the word "Logistics 4.0" from 2015 to 2022

	2015	2016	2017	2018	2019	2020	2021	2022	2023	Total
WoS	1	1	4	7	25	17	5	12	1	73
SCPS	0	1	1	5	17	12	7	10	2	55

Table 5 Scope of publications in proceedings papers containing the word "Logistics 4.0" from 2015-2022

SCPS			WoS
Engineering	30	48	Engineering Manufacturing, Industrial, Electrical Electronic, Civil, Multidisciplinary,
Computer Science	30	46	Computer Science Artificial Intelligence, Interdisciplinary Applications, Theory and Methods, Software Engineering, Information Systems, Telecommunications and Cybernetics
Business, Management and Accounting	19	30	Operation Research Management Science, Business, Management,
Decision Sciences	20	1	Statistics Probability
Social Sciences	2	10	Education Educational Research, Education Scientific Disciplines, Social Science Interdisciplinary
Environmental Science	5	7	Environmental Sciences, Studies, Green Sustainable Science Technology,
Economics, Econometrics and Finance	0	1	Economics
Materials Science	1	0	No Similar
Physics and Astronomy	2	0	No Similar
Energy	1	0	No Similar
No Similar	0	5	Transportation Science Technology, Transportation,
No Similar	0	3	Chemistry Multidisciplinary, Analytical,
No Similar	0	7	Automation Control Systems, Robotics
Other Mathematics, Medicine, Earth, and Planetary Science	13	2	Others Ergonomics

In the case of publications included in Proceeding papers, a slightly higher number of publications was identified in the case of the WoS database and a slightly smaller number of publications in the case of the SCPS database.

As in the case of the first group of publications, the most important subject areas include Engineering, Computer Science, Decision Science, Business and Management (and the names assigned to them from the Web of Science database), which is shown in Tab. 5.

Due to the optimization of the volume of the presented publication, in the next part of the work it was decided to use only publications available as scientific articles and available books – proceeding's papers were rejected. In total, we managed to gain access to 52 publications, of which 16 titles were repeated in both databases. This means that a total of 36 publications were included in the analysis. In the SCPS database, 4 unique titles were identified in the WoS database, similarly, in the WoS database, 16 unique titles were identified in the SCPS database.

The next step in implementing the research goals set before the article was a critical review of the collected articles in terms of defining the concept of Logistics 4.0.

From the conducted analysis, it was proposed to identify the following contexts for defining logistics 4.0:

I. Using Industry 4.0 tools in logistics.
II. Logistics 4.0 and the issue of managing organizations and economic aspects.
III. Creating new Logistics 4.0 models or definitional and conceptual assumptions for these models (Fig. 1).

As can be seen from Fig. 1, the largest number of titles of publications with L4.0 in the title refer to the economic management context, approx. 42%. In second place are publications that draw attention to the use of I4.0 tools in various areas and logistics processes. Articles focusing on conceptual-definitional and model aspects are only in third place.

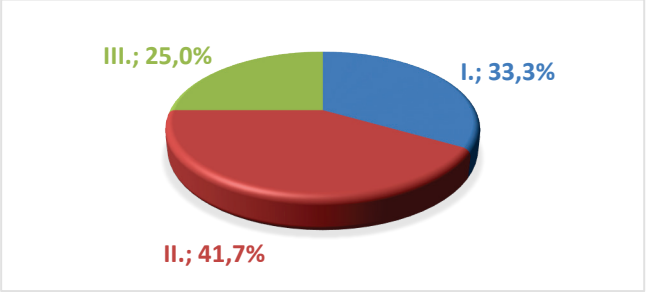


Figure 1 Definition contexts of the concept Logistics 4.0

The first context of defining logistics referred to the use of single or multiple I4.0 tools in logistics, logistics concepts, processes, logistics phases. Publications related to:

A. Single or multiple I4.0 tools used in logistics, e.g., blockchain.
B. I4.0 tools in the context of existing logistics concepts (Lean, Agile, LeAgile, Resilient, Green, Total Logistics Management, and others).

C. Use of I4.0 tools in specific logistics processes (e.g., transport, packaging, storage, order processing, inventory management).
D. The use of I4.0 tools in specific logistics phases (procurement, production, distribution, returns and disposal) – Fig. 2.

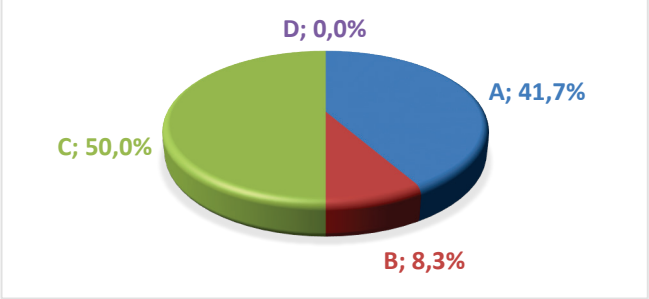


Figure 2 Context of defining logistics relating to the use of single or multiple I4.0 tools in logistics, logistics concepts, processes, logistics phases.

Fig. 2 shows that the largest number of analyzed publications relate to the use of single or multiple I4.0 tools in general logistics. Attention is drawn here to the technology used and its possibilities of application in logistics [8, 9] or very specific areas of it [10, 11] rather than to specific logistics processes [12] or logistics phases.

The second analyzed group consisted of articles relating to the broadly understood spheres of management and economics. Publications related to:

E. Management of functional areas of the organization in the context of the use of individual I4.0 tools - including, for example, social capital, outsourcing.
F. Absorption of knowledge and innovation.
G. Economic issues, current socio-economic trends e.g., sustainability and circular economy, and the use of I4.0 tools in specific industries – Fig. 3.

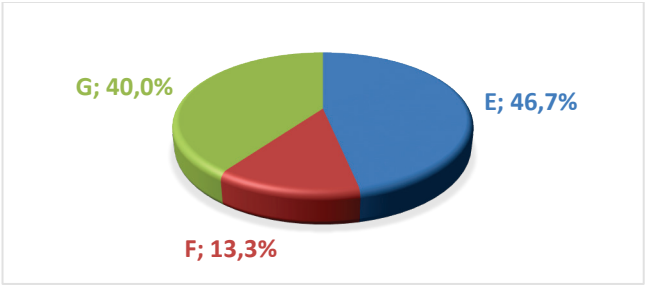


Figure 3 Context of defining logistics articles relating to the broadly understood spheres of management and economics.

As can be seen from Fig. 3, the largest number of publications in this group, i.e., about 47%, is occupied by publications relating to the spheres of organization management, e.g., social capital [13]. The next group consists of articles with the word L4.0 in the title referring to economic issues, economic and social trends, and specific industries [14] - 40%. The smallest group are publications on the absorption of knowledge and innovation, e.g. [15].

The third analyzed group of publication titles were those that referred to model and definitional contexts. In this way, works related to:

H. Models of Logistics 4.0.
I. Logistics 4.0 measurement models.
J. Logistics 4.0 maturity models.
K. Definition of Logistics 4.0 – Fig. 4.

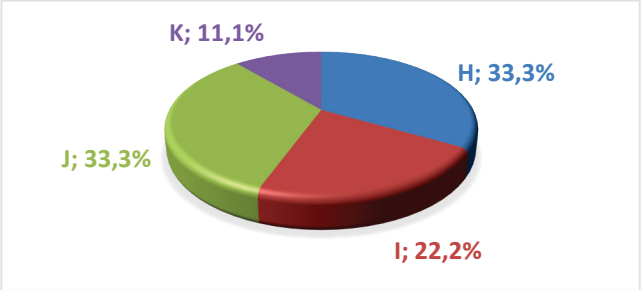


Figure 4 Context of defining logistics articles relating to model and definitional contexts.

The largest number of publications containing the L4.0 keywords applies equally to the general L4.0 models [16] and the L4.0 maturity models [17] – over 66% in total. L4.0 measurement models [18] and definitional issues [19] account for approximately 22% and 11%, respectively.

The literature review shows that the very concept of Logistics 4.0, in most cases, means issues of logistics management using selected I4.0 tools. Models whose aim would be to capture the entire L4.0 system, understood as a deliberately designed and organized cyber-physical system that implements broadly understood logistics issues, appear much less frequently.

To answer the fifth research question, three articles (from the searched group of publications containing the keyword Logistics 4.0 in the title) referring to the issue of the model approach to L4.0 were analyzed in detail.

The first of them by Strandhagen (et al.) [20] rightly combines the issue of logistics, the assumptions of I4.0 and the issue of sustainable development. It indicates eight basic trends in Logistics 4.0, which the authors include:

- Individualization.
- Servitization.
- Accessibility.
- Autonomy.
- Global network.
- Digitalization.
- Green logistics/circular economy.
- Sharing economy/collaboration.

Winkelhouse and Grosse [21], in turn, distinguish three basic elements within the L4.0 conceptual framework, i.e.:

- Technological blocks containing I4.0 technologies and digital transformation technologies.
- External changes related to the change of paradigms of mass production to mass personalization and individualization of products,
- Shortening of product life cycles,
- Issues of sustainable development and globalization.

As part of the inductive refinement of the L4.0 framework, they divide technological blocks into three basic stages related to information, i.e.: generation, handling and

use, in which selected technologies or I4.0 tools are used (e.g., IoT, CPS, CT, SocM) or Mobile Systems (MS). In turn, external changes were also divided into three groups, but slightly different: customer, market, and environment. On this basis, the L4.0 model, based on the existing paradigms of time, quality, cost, implements in individual phases and logistics processes, with the involvement of social capital based on socio-technical interactions, knowledge, and resources as well as system operators and administrators, as well as technological blocks - activities aimed at to cope with external changes [22].

The issues of sustainable development in the L4.0 model were also discussed by Bag's (et al.) [23]. They divided the L4.0 capabilities into technological, organizational, and environmental capabilities, and on this basis, they attempted to assess the company's performance. They have shown in their research that the impact of technological and environmental capabilities is stronger than the impact of organizational capabilities. Building L4.0 resources should focus more on dynamic technological and environmental capabilities rather than on organizational capabilities. Thus, the pressure of managers on the above-mentioned areas allows you to develop the capabilities of L4.0, and thus improve the agility, flexibility, and responsiveness of the organization [24].

In addition to the critical literature review presented in the paper, it is worth noting that Federico (et al.) [25] presented the theoretical structure of the Supply Chain 4.0 concept (SC4.0), which in a way refers to the theoretical constructs of the L4.0 models. They distinguished four basic pillars of SC4.0, which included leadership and social capital, technological levers based on I4.0 solutions, SC4.0 considering the concept of End to End (E2E) and defining measurements and requirements for processes. All this should contribute to the strategic results regarding the agility, flexibility or responsiveness of the organization mentioned by Bag's (et al.).

As you can see, the presented models have many common features, such as the use of I4.0 tools or the proper use of human capital. It seems, however, that they omit at least a few key elements that may pose serious challenges for enterprises wishing to follow the assumptions of the L4.0 concept, which will be discussed in the next chapter.

4 DISCUSSIONS

The presented answers to the first four research questions allow to indicate the current trends in the development of logistics and supply chains. The answer to the fifth research question requires a critical analysis of the publication in the context of the challenges, opportunities and threats that face the assumptions of the L4.0 concept.

When analyzing the collected literature, it should be noted that it seems to omit several key aspects that can be categorized in two areas: geo-political-social and technological. These issues, at the beginning of the third decade of the 21st century, are clearly felt from the point of view of practical aspects of logistics and supply chains in the world. What's more, these phenomena have not occurred on such a scale in recent decades, which is why they deserve attention even more. Geo-political and social issues include:

- The ongoing Covid19 pandemic, which certainly changes its face over time, thus hindering the predictability of economic phenomena.
- The ongoing conflict in Ukraine and other armed conflicts destabilizing global supply chains.
- Natural disasters and dynamic weather changes resulting from the deteriorating condition of the natural environment.
- The issue of the I-Gen (Internet Generation) society entering adulthood, which is very different from the previous X, Y and Z generations [26], the development of Society 5.0 (S 5.0) and Evergetics [27].

When it comes to technological aspects, the basic elements include:

- Considering the consequences of the introduction of 3-DP for logistics and supply chains.
- Legal regulations regarding the functioning of SF and its technology, their taxation, etc.
- The issue of access to AI – ChatGPT.
- Critical analysis of the use of thinking robots (perhaps also sentient robots due to artificial intelligence - which was brilliantly touched upon by Nobel Prize winner Ishiguro in his novel Klara and The Sun [28]).

As you can see, in addition to the opportunities presented by the development of civilization, including IV IR, there are several challenges and threats that should already be the subject of scientific research. In the realm of Geo-political and social issues, these include, for example:

- The presentation of business models no longer based on global supply chains (which was strongly verified by Covid).
- Exploring mediating conditions in areas of the world that are permanently threatened by armed conflicts, with a view to avoiding them, and building models for the functioning of economies in such crises,
- Identification of factors that allow for the simultaneous implementation sustainable development and the circular economy goals with existing business models.
- Social changes resulting from the large-scale use of information technology (IT) by people, profoundly changing the perception of knowledge, creativity, but also the very behavior of the I-Gen and younger generations in the context of, for example, education or the labour market.
- The conditions for the implementation of production and logistics activities under conditions of severely limited resources, turbulent climate change and high product personalization.

In the technological area, these include, for example:

- The use of additive manufacturing (e.g., 3-DP) on a mass scale, which may lead to changes in existing paradigms in the areas of logistics, production and quality stemming resulting from of the 2nd IR and 3rd IR.

- The use of IoT or the Internet of Everything (IoE) technologies, which will soon replace a large part of the processes at the interface of human-machine communication with M2M communication.
- The impact of augmented and virtual reality on human perception and the brain, especially when tools such as augmented reality glasses are used, for example, for logistics processes.
- The security of bigdata, which is set to increase through widespread digitalization, and the potential risks posed by criminal activity in the bigdata sphere.
- The effects of the application of AI, which through the aggregation of knowledge can generate ideas and projects hitherto unheard of (the question is also whether they are also safe for humans).

There are many areas where science and practice meet in the case of L4.0. The available technologies, allow fundamentally change the current industrial, economic, or social reality. However, these changes are too slow for the increasing number of challenges that are emerging, because they require a fresh, new perspective. An approach that in many cases will have to break away from existing laws, assumptions, and industrial paradigms in favor of completely new, often revolutionary solutions.

5 CONCLUSIONS

The presented review of the literature using the STAR method allowed to answer the research questions by indicating both the subject of publications related to the L4.0 issues and to indicate the existing achievements in this field. The main fields of knowledge in which the concept of L4.0 is situated were defined, and publications were selected in terms of definitional contexts. Finally, the L4.0 models presented in the analyzed publications are briefly discussed. Space was also left for further research of publications appearing as part of the Proceedings Papers.

Areas that have not yet been included in the literature analysis described have been indicated. This is probably due to the unlimited possibilities offered by the I4.0 concept and the need to have multidisciplinary knowledge that would enable an attempt to build this type of comprehensive models - logistics of the future – L4.0.

In such disciplines as economics or management sciences, it is very difficult to clearly indicate what an opportunity is, what a threat is, and what an opportunity for development is, because in general a given issue has its multi-contextual nature. Therefore, the presented article should be the beginning of a discussion on the modification of the existing production, logistics or quality paradigms towards the future development of I4.0, based on which logistics and supply chains would develop. Development that would be sustainable at the same time, caring primarily for the planet and the well-being of people.

It is also important to note the limitations of the research. The first limitation is the fact of selecting articles that only contained the word Logistics 4.0 in the title. Certainly, research containing the word Logistics 4.0 not only in the title, but also in the keywords and in the abstracts would have been much more detailed. The second limitation was the fact that only bibliometric-abstract databases such as WoS and SCPS were searched. This may in some ways leave out articles or conference proceedings that are not included in these databases. A third limitation is the incompatibility of the subject areas of the two databases, which significantly hinders the research.

Further research directions will develop in expanding the scope of keywords (e.g., Logistics 4.0, L4.0, Supply Chain 4, Supply Chain Management 4.0, etc.) and the scope of searches (title, keywords and abstract). In further research, it is also planned to extend the searched databases to identify comprehensive models that consider the characteristics of I4.0 in supply chains and logistics processes.

6 REFERENCES

[1] Kodama, F. (2018). Learning Mode and strategic Concept for the 4th Industrial Revolution. *Journal of Open Innovation: Technology, Market, and Complexity, 32*(4), p. 1. https://doi.org/10.3390/joitmc4030032

[2] French, A. M., Jain, H., Larsen, K. R, Risius, M., & Shim, J. P. (2021). The 4th Industrial Revolution Powered by the Integration of AI, Blockchain, and 5G. *Communication of the Association for Information Systems, 49*(10), p. 267. https://doi.org/10.17705/1CAIS.04910

[3] Roblek, V., Mesko, M., & Krapez, A. (2016). A Complex View of Industry 4.0. *SageOpen, 6*(2), 1-3. https://doi.org/10.1177/2158244016653987

[4] Zheng, T., Ardolino, M., Bacchetti, A., & Perona, M. (2021). The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review. *International Journal of Production Research, 59*(6), 1923-1924. https://doi.org/10.1080/00207543.2020.1824085

[5] Booth, A., Sutton, A., & Papaioannou, D. (2012). *Systematic approaches to a successful literature review*. Los Angeles: Sage, 24-26.

[6] Wondimu, N. A., Visser, U., & Buche, C. (2022). Interactive Machine Learning: A State-of-the-Art Review. https://arxiv.org/abs/2207.06196. https://doi.org/10.48550/arXiv.2207.06196

[7] Rowley, J. & Slack, F. (2004). Conducting a Literature Review. *Management Research News, 27*(6). https://doi.org/10.1108/01409170410784185

[8] Kim, E., Kim, Y., & Park, J. (2022). The Necessity of Introducing Autonomous Trucks in Logistics 4.0. *Sustainability, 14*, 3978. https://doi.org/10.3390/su14073978

[9] Tubis, A. & Poturaj, H. (2021). Challenges in the implementation of autonomous robots in the process of feeding materials on the production line as part of Logistics 4.0. *LogForum, 17*(3), 411-432. https://doi.org/10.17270/J.LOG.2021.611

[10] Chen, Y-T., Sun, E., Chang, M-F., & Lin, Y-B. (2021). Pragmatic real-time logistics management with traffic IoT infrastructure: Big data predictive analytics of freight travel time for Logistics 4.0. *International Journal of Production Economics, 238*, 108157. https://doi.org/10.1016/j.ijpe.2021.108157

[11] Di Nardo, M., Clericuzio, M., Murino, T., & Sepe, Ch. (2020). An Economic Order Quantity Stochastic Dynamic Optimization Model in a Logistic 4.0 Environment. *Sustainability, 12*, 4075. https://doi.org/10.3390/su12104075

[12] Perotti, S., Bastidas Santacruz, R. F., Bremer, P., & Beer, J. E. (2022). Logistics 4.0 in warehousing: a conceptual framework of influencing factors, benefits and barriers. *The International Journal of Logistics Management, 33*(5), 193-220. https://doi.org/10.1108/IJLM-02-2022-0068

[13] Lagorio, A., Cimini, C., Pirola, F., & Pinto, R. (2021). A taxonomy of technologies for human-centered logistics 4.0. *Applied Sciences, 11*(20). https://doi.org/10.3390/app11209661

[14] Zoubek, M., Simon, M., & Poor, P. (2022). Overall Readiness of Logistics 4.0: A Comparative Study of Automotive, Manufacturing, and Electronics Industries in the West Bohemian Region (Czech Republic). *Applied Science, 12*(11). https://doi.org/10.3390/app12157789

[15] Stachowiak, A., Adamczak, M., Hadas, L, Domański, R., & Cyplik, P. (2019). Knowledge Absorption Capacity as a Factor for Increasing Logistics 4.0 Maturity. *Applied Sciences, 9*(24). https://doi.org/10.3390/app9245365

[16] Strandhagen, J., Vallandingham, L., Fragapane, G., Strandhagen, J., Stangeland, A., & Sharma, N. (2017). Logistics 4.0 and emerging sustainable business models. *Advances in Manufacturing, 5*, 359-369. https://doi.org/10.1007/s40436-017-0198-1

[17] Zoubek, M. & Michal, S. (2021). A framework for a logistics 4.0 maturity model with a specification for internal logistics. *MM Science Journal*, 4264-4274. https://doi.org/10.17973/MMSJ.2021_03_2020073

[18] Dallasega, P., Woschank, M., Sarkis, J., & Tippayawong, K. Y. (2022). Logistics 4.0 measurement model: empirical validation based on an international survey. *Industrial Management & Data Systems, 122*(5), 1384-1409. https://doi.org/10.1108/IMDS-11-2021-0694

[19] Amr, M., Ezzat, M., & Kassem, S. (2019). Logistics 4.0: Definition and historical background. *Novel Intelligent and Leading Emerging Sciences Conference (NILES) - IEEE*, 1, 46-49. https://doi.org/10.1109/NILES.2019.8909314

[20] Strandhagen, J. O., Vallandingham, L. R., Fragapane, G. et al. (2017). Logistics 4.0 and emerging sustainable business models. *Advances in Manufacturing, 5*, 366-367. https://doi.org/10.1007/s40436-017-0198-1

[21] Winkelhaus, S. & Grosse, E. (2020). Logistics 4.0: a systematic review towards a new logistics system. *International Journal of Production Research, 58*(1), p. 21. https://doi.org/10.1080/00207543.2019.1612964

[22] Winkelhaus, S. & Grosse, E. (2020). Logistics 4.0: a systematic review towards a new logistics system. *International Journal of Production Research, 58*(1), p. 34. https://doi.org/10.1080/00207543.2019.1612964

[23] Bag, S., Gupta, S., & Luo, Z. (2020). Examining the role of logistics 4.0 enabled dynamic capabilities on firm performance. *The International Journal of Logistics Management, 31*(3), 607-628. https://doi.org/10.1108/IJLM-11-2019-0311

[24] Bag, S., Gupta, S., & Luo, Z. (2020). Examining the role of logistics 4.0 enabled dynamic capabilities on firm performance. *The International Journal of Logistics Management, 31*(3), p. 620. https://doi.org/10.1108/IJLM-11-2019-0311

[25] Frederico, G. F., Garza-Reyes, J. A., Anosike, A., & Kumar, V. (2019). Supply Chain 4.0: Concepts, Maturity and Research Agenda. *Supply Chain Management*, p. 18. https://doi.org/10.1108/SCM-09-2018-0339

[26] Twenge, J. (2018). *iGen: Why Today's Super-Connected Kids Are Growing Up Less Rebellious, More Tolerant, Less Happy--and Completely Unprepared for Adulthood--and What That Means for the Rest of Us*. Atria Books, Simon & Schuster, Reprint edition, New York.

[27] Skobelev, P. & Borovik, S. (2017). On the Way from Industry 4.0 to Industry 5.0: From Digital Manufacturing to Digital Society. *International Scientific Journal Industry 4.0, 2*(6), 308-309.

[28] Ishiguro, K. (2022). *Klara and the Sun*. Faber & Faber, London.

Author's contacts:

Maciej Bielecki, PhD
Department of Logistics and Innovations,
University of Lodz,
Ul. Rewolucji 1905 r. 37/39,
90-214 Lodz, Poland
maciej.bielecki@uni.lodz.pl

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230518082456

Subject review
Received: 2023-05-18, Accepted: 2023-05-22

PQCDSM-Logic in Maintenance (TPM) and Mountaineering

Stefan Schmidt*, Benjamin Stefan Godwin Schmidt

Abstract: TPM is the foundation for JIT (Just in Time) and Lean Manufacturing and forms the basis of JIT or on-time delivery. The goal of TPM is to improve equipment effectiveness and optimize equipment performance, namely PQCDSM (Productivity, Quality, Cost and Delivery, Safety and health, environment, and Morale). Many producers have tried to transform their production system to a JIT or Lean production system with the aim of increasing productivity and quality, but thus far with little success. This contribution shows how trekking and climbing tours can be used to illustrate the application of PQCDSM-Logic in mountaineering and how this can be transferred to logistics and maintenance practice. The background is the author's decades of experience with expeditions, trekking and climbing tours, TPM implementations and interviews with numerous experts. There are many similarities between the application of PQCDSM-Logic in mountaineering and in logistics and maintenance practice, which will help both in operational practice in industry and in high mountain tours, especially regarding safety in a changing environment. Presented is the extrapolation from mountain climbing to TPM and the importance of leadership for a successful (summit climbs and the like) transformation of the production system to a JIT or Lean production system.

Keywords: environment; equipment; leadership; logistics; PQCDSM; mountaineering; safety; Total Productive Maintenance

1 INTRODUCTION (LITERATURE REVIEW)

As early as the 1960s and 1970s, some manufacturing sectors in the West were already being outperformed or partially eliminated by Japanese competitors (optics, watchmaking, shipbuilding, and consumer electronics). After the first oil crisis in 1973 and, in particular, the second oil crisis in 1978, the Japanese automotive industry was also perceived to be a serious competitor. With great effort, especially through the implementation of rationalization and automation measures, Western manufacturers tried to become competitive again in the 1980s.

The fundamental basis of Japanese success was little understood. Although there were already some descriptions of Japanese production systems, this had little impact on industrial practice, e.g., the 1981 publications on the *Productivity Machine in Japan* by Hall [1] and *Japanese Factories Work* by Hayes [1] and on the *Toyota Production System* by Shingo 1981 [3], Monden 1983 [4] and Ohno 1988 [5]. Other publications on this topic included *Japanese Manufacturing Techniques* by Schonberger 1982 [6], *The SMED System* by Shingo 1985 [7], *Japanese Just-In-Time production system* by Reda 1987 [8], *and Total Productive Maintenance* by Nakajima 1988 [9], which were also without significant resonance in the manufacturing industry.

It was not until the publication of *The Machine that Changed the World* in 1990 [10] that Western manufacturers had a significant wake-up call and subsequently made frantic efforts to convert their production to a just-in-time production system. However, only few were successful. A boom of translations of Japanese books on production methods into German took place, such as *Six Sigma* 1990 [11], *Kaizen for Quick Changeover* 1995 [12] and *Introduction to TPM* 1995 [13], but after the Japanese real estate bubble burst, manufacturers returned to their traditional processes.

When the Japanese car company Toyota replaced Ford as the world's second-largest automaker in 2003 [14], the automotive industry finally woke up from its deep sleep. This was followed by hectic activity, with many companies trying to change their production system to Lean production or derivatives of TPS with TPM methodology and excellent manufacturing as their goal. As a result, the last two decades have been filled with numerous attempts to apply these concepts in companies. However, this has only been partially successful and has, in fact, often completely failed.

The main reason why most companies fail in implementing Lean is presented in this paper. In 1985, Shigeo Shingo stated that "60 to 70% of success comes from people" and 30 years later this still had not changed [15]. The results of a study conducted decades ago by Harvard University in the US on 400 different companies showed why most transformations had either partially or completely failed. The most important reason for failure was the lack of comprehensive leadership. Other reasons included the lack of an effective decision-making process as well as a lack of direction and vision in executing the transformation [15].

Numerous studies show the reasons for the failure of Lean. They revealed that Lean implementation failures mainly fall into two common failure mode categories, involving ‘Top Management Leadership’ and ‘People/Employee’ issues. "Almost all Lean experts agree that the main reasons for implementation failures involve senior management and ownership". Leaders are, after all, ultimately responsible for everything that happens within a company [16].

The track record for a successful implementation of the Lean methodology is spotty at best. Some studies state that failure rates for Lean programs range between 50 percent and 95 percent [17].

Achanga et al. [18] describe four main key factors for the implementation of Lean manufacturing in an SME environment: leadership and management, finance, skills and expertise, and the inherent culture of the organization. “Leadership and management commitment are the most critical factors in determining the success of a transition to Lean”. According to one study, the success or failure of Lean transitions strongly depends on the company's approach and on whether it has developed its own Lean philosophy [19].

The track record for implementing Lean is poor. According to studies, more than half of all Lean implementations fail because of corporate culture: No strategy, e.g., no clear responsibilities, no metrics, and no executive commitment. Lean requires top-down leadership [20]. The success of Lean implementation depends on top management commitment and leadership, the attitude of the employees, resources, and organizational culture [21].

For years, one major problem with Lean implementation, as indicated by the numerous surveys and studies on the topic, has been its overwhelmingly high failure rate. The results of Lean implementations are discouraging. For example, a 2007 Industry Week study found that nearly 70% of all US plants were using Lean, but only 2% were meeting their goals. There are many other literature reviews and case studies about the high failure rate of Lean and, after more than 30 years, it is still not getting any better [22].

The successful implementation of Lean can bring enormous benefits to any company, but in practice the failure rate is very high. An analysis of existing literature from the last 20 years and further analysis of studies from the last 10 years show: leadership from top management, the organizational culture of the company, effective communication, and knowledge and mindset remain the most important factors for a successful Lean implementation [23]. As with any major organizational change, it is necessary for senior management to be committed to reducing headwind and minimizing employee resistance [24].

Only few companies implement Lean manufacturing correctly and there are also a number of difficulties that inhibit the success of these initiatives, e.g., logistical and planning challenges as well as behavioural concerns. Some common challenges include, on the one hand, lack of support from senior management and, on the other, inadequate employee training, limited workforce, lack of appropriate tools, failure to monitor progress (insufficient performance data), problems arising from the implementation of multiple changes at once, and the difficulties caused by replacing established work practices (cultures) [25].

A study on the implementation of Lean in small to medium-sized enterprises (SMEs) reveals that the real problem for achieving Lean success was not a lack of commitment on the part of top-management, but rather their ignorance regarding what they should commit to, i.e., it was a knowledge problem rather than a commitment problem [26]. For SMEs with limited resources, management knowledge is especially important when implementing Lean. There are two categories of insufficient manager Lean knowledge in terms of failure factors [26]:

- Level 1: Ignorance, i.e., a complete lack of knowledge, of the following: Benefits of Lean and how to implement it as well as how Lean concepts have developed.
- Level 2: Erroneous knowledge of Lean implementation, i.e., unawareness that: Lean is not purely a manufacturing tool, belief that Lean will be of no benefit outside of mass production, is for process improvement only, is implemented by process engineers and consultants, i.e., and there is no need for 'me', the manager, to be involved.

It is not helpful to simply blame top management for these types of failures. However, the references given are only a very small selection on the topic of "no top management commitment". It remains to be seen whether "no top management commitment or "top management ignorance with either a complete lack of knowledge about Lean or erroneous knowledge about the implementation of Lean" are more damaging [26].

No commitment or no idea - Catch-22 or debacle? There are real reasons for failures and their causes. However, top management is the key person and the one who makes the decisions and ultimately bears overall responsibility.

2 TOTAL PRODUCTIVE MAINTENANCE (TPM) AND PQCDSM

Total Productive Maintenance (TPM) is a management system, which was developed by Seiichi Nakajima [9, 13] in Japan and implemented for the first time in 1971 to optimize operational processes through the creative participation of all employees. TPM is the foundation for the JIT (Just in Time) production system, which is dedicated to producing the right part at the right time and of the right quality. Reliability and the synchronising of production is indispensable. TPM involves productive maintenance, which is the responsibility of all employees and performed in small group activities, and equipment maintenance, which is performed on a company-wide basis. A dual goal of TPM is zero breakdowns and zero defects, which form the basis of JIT or on-time delivery.

With increasing robotization and automation, the greater the shift of the production process from workers to interconnected machines and the larger the role played by the equipment itself. The purpose of TPM is to enhance the effectiveness of the equipment and maximize its output by optimizing the Key Process Indicator's (KPI's) PQCDSM: *Productivity, Quality, Cost, Delivery, Safety (and health, environment), and Morale.* PQCDSM is dependent on the condition of the equipment and can be individually adapted to suit the requirements of the respective organization implementing it. To achieve this, they must first define and set targets of each of the areas of PQCDSM by considering all losses and bottlenecks that could potentially affect plant performance. TPM also aims to eliminate the 'big losses' to achieve an increase in Overall Equipment Effectiveness (OEE) [9, 13, 7, 12]. This is the only indicator that effectively combines the measures Availability, Performance, and Quality of a site.

Suzuki cites, for example, PQCDSM (Productivity, Quality, Cost, Delivery, Safety, Morale) improvements for early TPM implementers in Japan, see Table 1 [27]. Successful TPM implementations are driven by leadership commitment and support, high motivation and involvement of employees, sufficient resources, and communication [12, 13, 28].

A PQCDSM-process board of an automobile manufacturer is shown in Fig. 1 [28].

TPM Excellence Awards of JIPM, who, since 1994, has been evaluating and honouring enterprises or factories

throughout the world with their award for "Operational Excellence and Maintenance Excellence", with the aim of encouraging further progress and development. In the TPM Award Assessment Achievement Sheet for the TPM Excellence Award outside Japan, the PQCDSM categories are represented according to the TPM start or last award, current status, and targets of the company [29].

According to Nakajima, the application of TPM must be tailored to the individual needs of the company. Each company must develop its own action plan according to its problems, industry, type, and condition of equipment [9].

Table 1 PQCDSM-improvements for early TPM implementers in Japan [27]

- **P – Productivity:**
  Net productivity up by 1.5 to 2.0 times.
  Number of equipment breakdowns reduced by 1/10 to 1/250 of baseline.
  Overall plant effectiveness 1.5 to 2.0 times greater.
- **Q – Quality:**
  Process defect rate reduced by 90%.
  Customer returns/claims reduced by 75%.
- **C – Cost:** Production costs reduced by 30%.
- **D – Delivery:** Finished goods and Work in Progress (WIP) reduced by half.
- **S – Safety.**
  Elimination of shutdown accidents.
  Elimination of pollution incidents.
- **M – Morale:** Employee improvement suggestions up by 5 to 10 times.

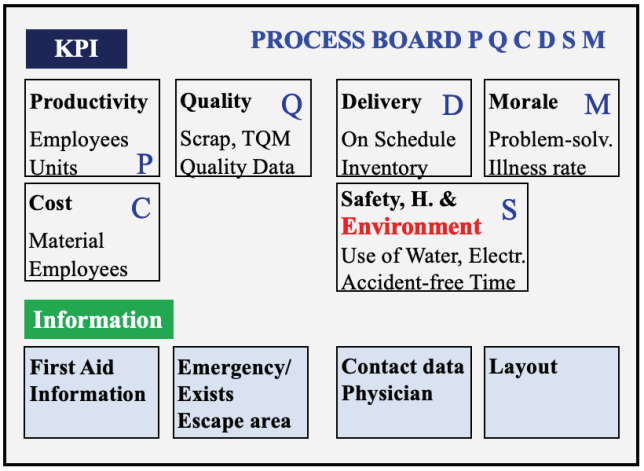


Figure 1 Process Board P Q C D S M an automobile manufacturer [28]

3 MOUNTAINEERING IN A CHANGING ENVIRONMENT AND PQCDSM

3.1 Mountaineering in a Changing Environment

The great outdoors in the high mountains is associated with potential risks and danger: rockfalls, avalanches, crevasses on glaciers, landslides, extreme cold, and dramatic weather changes. These can be minimized by appropriate equipment, cautious route selection, and adequate knowledge of the risks [30, 31, 32].

Mountains influence weather and weather phenomena in mountainous regions can be quite spectacular and often extreme. In recent decades, conditions and weather patterns have changed, sometimes significantly, especially in the high mountains of Asia, South America, and Africa. Landslides and regions vulnerable to landslides with loose, steep sandy gravel slopes; the possibility of floods and heavy rains in the semi-arid cold steppe that can wash away complete villages, the risk associated with heavily swollen river crossings; exposed ridges, or slippery ground; the increased danger of crevasses and ice breakage; crossing areas that are abandoned by climate refugees because of drought, cold spells, rock falls, minor earthquakes, and volcanic eruptions – all phenomena that the author has experienced himself [33, 34].

There are, on the one hand, known risks that can be successfully managed with proper planning, and, on the other, completely unexpected hazards (unknown unknowns) such as the October 2014 cyclone in Nepal that brought with it blizzards, unusually strong snowstorms and avalanches, a disaster that killed at least 43 people in the Annapurna/Thorong La pass area alone [35]. The unavoidable transportation to reach the tour start can be a risk, e.g., travelling along rocky roads that resemble the bed of a mountain river or having to fly to airports where landings are extremely hazardous, e.g., Lukla's airport in Nepal, called "the most dangerous airport in the world", due to the many accidents with numerous fatalities among passengers and crews [36]. Once at the destination, a cook, helpers, porters and/or pack animals are required, whose quality is vital for the successful outcome of an expedition?

For assessable risks, the necessary equipment is required: clothing, mountaineering boots, rope, crampons, ice axe, climbing harness, carabiner, helmet, tent, etc. In the high mountains, altitude sickness can occur from an altitude of 2000 to 2500 meters, if the ascent is too fast [37, 38, 39, 40]:

- HYPOXIA (oxygen deficiency) massively impairs judgment, slows down all reflexes and clouds the sense of reality (Hypoxia symptoms vs oxygen saturation [37] Table 2).
- Acute Mountain Sickness (AMS) occurs mainly at altitudes above about 2000 m and is the most common form of altitude sickness (leading symptom headache).
- High Altitude Pulmonary Edema (HAPE), 2/3 of all HAPE cases occur between 3000 and 4500 m above sea level (leading symptom sudden loss of performance).
- High Altitude Cerebral Edema (HACE), although is rarer than HAPE, is often fatal, and symptoms can progress very rapidly to coma and death (leading symptom ataxia).
- Combined high altitude edema: HAPE and HACE very often occur simultaneously. The course of this combination is very often fatal.

Good acclimatization is therefore essential and oxygen saturation should be monitored by a pulse oximeter throughout the tour [34, 37, 38, 39, 40].

Besides a lack of acclimatization, another risk factor is a person's individual vulnerability. During an expedition in 2017, a team member had to be flown out by rescue helicopter from a high plateau at 5200 meters due to his very poor general condition and oxygen saturation of only 51 percent. From 50 percent O_2, a rescue helicopter is used in Nepal, if

it can be requested with a satellite phone and the assumption of payment by an insurance company is proven. The author himself was able in 2018 to descend very quickly from 4900 meters to lower elevations after a night with 54 to 49 percent O_2 [34, 38]. This shows very clearly the safety aspect and the logistics required in mountaineering. With the support team, all equipment, and a food supply for several weeks, including beyond the envisaged tour end, must be provided, as replenishment is not possible.

Table 2 Hypoxia symptoms versus oxygen saturation [37]

Arterial Oxygen Saturation Levels / Hypoxia Symptoms Versus Altitude			
Altitude	O_2 Saturation	Stages of Hypoxia	Symptoms
0 – 10,000 ft 0 – 3048 m	98 – 90 %	Indifferent	• Decrease in night vision
10,000 – 15,000 ft 3048 – 4572 m	90 – 80 %	Compensatory	• Drowsiness • Impaired judgement, coordination, efficiency
15,000 – 20,000 ft 4572 – 6096 m	80 – 70 %	Disturbance	• Impaired flight control • Impaired handwriting, speech • Decreased coordination
20,000 – 25,000 ft 6096 – 7620 m	70 – 60 %	Critical	• Circulatory and central nervous system failure • Convulsions • Cardiovascular collapse • Death

3.2 Mountaineering in a Changing Environment

The TPM safety, health and environment pillar is particularly significant. "Ensuring equipment reliability, preventing human error, and eliminating accidents and pollution are the key tenets of TPM" [27, p. 323] and implementation of the TPM safety, health and environment pillar focuses on identifying and eliminating safety and environmental issues. According to the Heinrich Principle (Heinrich 1980), for every 500,000 safety incidents, there are 300 'near misses', 29 injuries, and 1 fatality (see Fig. 2). In studying industry accidents, Heinrich found that 88 percent of accidents were caused by unsafe actions of people, 10 percent were caused by unsafe physical conditions, while he considered the remaining 2 percent to be "acts of God" [41, p. 63], see Fig. 2. The Heinrich pyramid was further developed in 1966 by Frank E. Bird from an analysis of 1.7 million accident reports from nearly 300 companies.

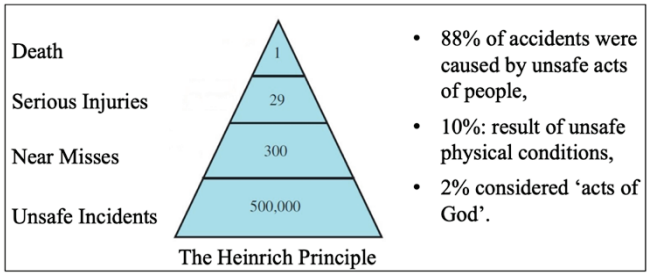


Figure 2 The Heinrich Principle of accidents [41, p. 64]

Environmental safety is increasingly becoming a focus of TPM implementation. Production management in the 21st century will not be effective if environmental issues are ignored. An environmental management system (EMS) that integrates environmental issues into manufacturing systems must be established. In other words, environmental safety goes beyond simply eliminating accidents. In today's manufacturing environment, environmental safety includes reducing energy consumption, eliminating toxic waste, and reducing raw material consumption. The environmental management system is thus an integral part of production management and should be implemented via TPM. This consists of environmental education, product, and equipment development, implementing improvements to reduce environmental aspects, and considering environmental impact. It is considered appropriate to develop these issues along the traditional TPM pillars [41, p. 65-66].

Safety, health, and environment have a significant importance both in mountaineering and as a pillar of TPM. How the most beautiful and remote high mountains are polluted shall not be further discussed here [33, 34]. Table 3 shows the experience during tours/expedition in high mountains with the application of PQCDSM logic in mountaineering and the possible impact on maintenance practice. [33, 34].

Table 3 PQCDSM-Logic - Mountaineering facts [33, 34]

P	– Excellent performance of support team (guide, cook, helper, porters, horse handler). – Public transport – a nightmare – Hindu society
Q	– Excellent quality of support team. – Excellent meals three times a day, even at elevations of up to nearly 5850 m.
C	– Total costs, no savings on life-saving equipment – Trip, Equipment, Energy, Insurance, Satellite telephone, Pressure bag.
D	– Overall organization – nearly everything on time, – But no plane because of bad weather, adventure public transport
S	– River crossings, cavasses, glaciers, insurance, falling, rock falls, crampons, – High altitude sickness, rescue helicopter, exposed crest, slippery ground, pulsoximeter – Waste management, mismeasurement – Safety is the maintenance of peace of mind
M	– Excellent moral of support team and of climbers (snowstorm etc.)

4 MOUNTAINEERING, MAINTENANCE AND LEADERSHIP

4.1 Maintenance and Mountaineering

According to Nakajima [9], equipment maintenance means equipment health. In other words, each person is responsible not only for his own health, but also for the health of the equipment he uses. There are many similarities between the application of PQCDSM logic in mountaineering and in logistics and maintenance practices that will help in operational practice and high-altitude tours, especially in terms of safety.

An expedition around remote Dhaulagiri and the ascent of Thapa Peak 6076 m in 2017 and the ascent of Mera Peak

2018 m demonstrate the application of PQCDSM logic in mountaineering and the transfer to logistics and maintenance practice, what do and what to consider (see Tab. 4) [28, 34], Fig. 3 shows the mountaineering route to Mera Peak 2018 [34].

Table 4 Maintenance & Mountaineering: and the transfer to logistics and maintenance practice [28, 34]

There are no minor things. - *"It's the little things that make the big things possible", J. W. Marriot.*
You are responsible, no one else is to blame. - *Defined tasks and responsibility.*
There are risks, known and unknown. - *Implement Risk management, emergency plan.*
There is always an unexpected first time. - *Risk management, but learn, to avoid same error a second time.*
Meticulous preparation always. - *Proper SOP, process description, training, maintenance manuals.*
Common sense (existence doubtful). - *Think about things from the ground – Aristoteles quoted that common sense is not so common.*
Strange / "clever" actions. - *Think ahead, safety first, no dangerous shortcuts.*
Communication saves lives - Silence kills. - *Digital communication speeds up, define reporting line.*
Supplier/Support. - *Spare parts available/at supplier, reliable supplier.*
Knowledge, skills, training, education, talent, good base. - *E. g. execute recertification, training.*
Exercise, exercise, exercise. - *Train evacuation in plants (to seldom done), instruct process sheets.*
Appropriate health information, visual, O_2 measure. - *Emergency/ first aid information. e. g. on process board.*
Equipment – know how to use it, train (pressure bag). - *Proper instructed and practiced.*
Saving money on equipment, higher risks follow. - *Equipment wears down, risk of break downs, savings damages occur.*
Adopt timetable and speed to abilities and elevation / environment. - *Provide buffer time.*
Know/Look for early warning signs. - *Early warning systems, monitoring process parameter e. g. vibration.*
Prepare known unknowns. - *Safety budget in case of emergency; alternative plan e.g., supply & production.*
Do not try to be clever / take unnecessary risks. - *Always Consider consequences.*
Hang on when it hurts, know when to stop, 110/120 bpm. - *Improve Resilience, consider cost sunk fallacy.*

Safety is the highest priority in mountaineering. The goal is to climb to the summit, but even more important is to return safely. Logistics are also crucial for successful mountain climbing. This also applies to companies. The ultimate goal is to sustain the company over the long term with reliable supply chains.

Figure 3 Mountaineering Route to Mera Peak 2018 [34]

4.2 Mountaineering and its Significance for Corporate Leadership

The fundamental basis of Japanese success was at first poorly understood and then largely ignored. Fig. 4 shows the first publication in Germany in 1982 on the way of production in Japan [43]. The EU-Japan Centre for Industrial Cooperation, a joint venture established in 1987 by the European Commission and the Japanese government (METI), promotes cooperation between the EU and Japan, e.g., training programs and Lean master classes on Lean management [44] and 'Lean in Europe visits' [45].

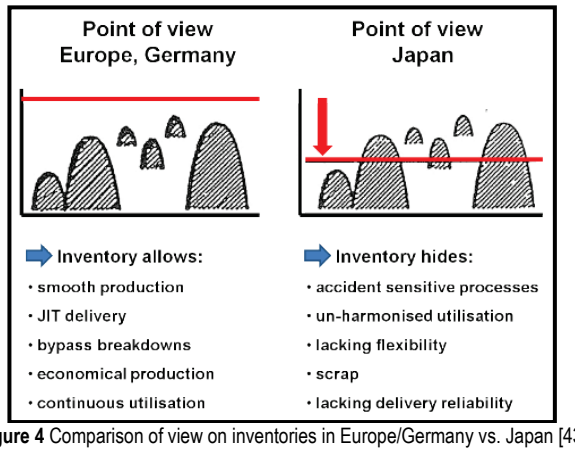


Figure 4 Comparison of view on inventories in Europe/Germany vs. Japan [43]

As noted previously, the numerous attempts over the past three decades in the West to transform companies to Lean production have only partially succeeded or often failed. For example, in over 30 years, the exceedingly high number of efficiency and quality campaigns by a well-managed automobile manufacturer have not led to any great success [46]. Leadership is the basis for excellence [47] and, as already

mentioned, the main reason for this lack of success is an absence of leadership commitment - Lean requires top-down leadership: no commitment/no ideas - Catch-22 or debacle?

Excellence in production is a goal for the industry with leadership as the foundation. Expedition and climbing leadership require important leadership skills. Reflections on my experiences in expeditions and climbing tours made me aware of the parallels between my high mountain tours (since 1973) and my industrial practice with Japanese production methods/Lean (since 1982). Both high mountain tours and the transition to Japanese production methods/Lean require planning, stamina, goal setting, commitment, calculation, teamwork, conviction, and endurance [33, 34, 38].

High mountain tours require an achievable goal. Common to high mountain tours and industry is the setting of a vision and the provision of support to employees by mountain/company leaders in achieving that vision. Most important, of course, is the safety aspect: sustainable survival of the company and of the climbers:

> *"In principle, the mountaineer's work is simple: "To win the game he has first to reach the mountain's summit, "said George Mallory, who took part in Britain's first three attempts on Everest in the 1920s. "But, further, he has to descend in safety" [48].*

Company leaders often use examples from mountaineering or sports, or they hire the relevant people to motivate their employees, in order to achieve their objective - world class manufacturing, which is synonymous with the victory of reaching the summit or a championship win. All require teamwork, a shared goal, and the staying power to overcome difficulties. As with mountain climbers, an organization's vision must be linked to a clear route to the summit or long-term goal. Once the vision and route are established, what happens next depends largely on the complementary actions of leaders.

My numerous expeditions and climbs highlighted key powerful characteristics and skills for leading such tours [33, 34, 38]. These qualities also fit very well for managers in Industry:

- Humility and partnership: mountaineers are especially aware that they may be exposed to forces of nature.
- Stamina, resilience, and determination: in the endurance sport of mountaineering, a goal can be achieved in hours, days, weeks or even months.
- Adaptability: they adjust their leadership style to rapidly changing conditions.
- Social intelligence and empowerment: they quickly establish positive interactions with stakeholders and provide a supportive space for growth and development.
- Trust builders: they help their clients to build trust in themselves and their support teams.
- Risk awareness and overcoming obstacles: act skilfully and safely under uncertain conditions, be well prepared, use checklists and overcome barriers.
- Big picture thinking: have a holistic view of the mission and be able to plan alternatives.
- Recognize and integrate what you have learned: learn from the most difficult experiences and share these insights with others.
- Practice patience, be grateful, and share your success.

There are now several works on this topic from business schools/universities, companies, and consultants. The Wharton School has developed Nano Tools for Leaders® with some similar criteria, i.e., lead like a mountain guide by adopting the strengths of mountain climbers [49].

Lessons from Everest show the role of collaborative leadership in a crisis situation and draw insights for business leaders. A collaborative leader creates a safe, clear, and cohesive environment for the group to work in, masters cross-boundary skills, including utilizing diversity in the group, and inspires the group through vision and character [50].

Mountaineering is like business leadership, in that leading a business or leading a rope team are both highly personal experiences. Success might be a result of superior planning, or it might also be just pure luck. As is the case for most mountaineering endeavours, the safety and success of a business rely on the strength of a team [51]. In fact, many of the required leadership skills in business can be developed while climbing mountains, namely, the necessity of practicing the art of patience, persistence, gratitude and sharing the success. [52].

Nine observations on leadership and teamwork from Mt. Everest shows there is power in having a clear goal with a hard deadline; that you must not be taken in by first impressions; that it is important for morale to frequently express your gratitude to your team, and that you should be aware that, although technology has its uses, it is only useful until it starts to detract from the mission. Some mountaineering teams are equipped with smartphones, cameras, smart watches, GoPro videos, and Wi-Fi and although this connects the team members to the outside world, it detracts from their connection to each other and their shared experiences. In this situation, technology can even be harmful as it brings with it distractions and misinformation and ultimately risks [53].

5 CONCLUSIONS

In mountaineering, everyone depends on each other for success and for their survival, and, to achieve this, tour leaders set their team a clear, definite goal. However, most companies do not have such a clear concept regarding their goals, even though creating one is the key to successful leadership. Therein lies the challenge. Lean production systems and the new development towards home offices require an adapted leadership culture [54, 55, 56].

Of the many similarities between the application of PQCDSM logic in mountaineering and in logistics and maintenance practice, safety should be given particular emphasis in both operational practice and high mountain tours. The most important goal in mountaineering is a safe return, while in industry it is the survival of the enterprise. However, the life span of companies has been decreasing dramatically for years. Furthermore, the endurance aspect, in particular, must be given far greater consideration.

The corporate life span has shortened dramatically in recent decades. The average length of time that US corporations remain in the S&P 500 stock market index has reduced from 33 years in 1965 to only 20 years in 1990. By 2026, the average length of stay is projected to be only about 14 years, and it is expected that about half of the 500 companies currently represented in the S&P will be excluded from the index within the next ten years [57]. Of the companies founded in Germany in 2015, the survival rate after five years was only 37.1 percent [58]. Due to this dramatically increased rate of change in the economic and technological environment, half of the large companies in the United States have disappeared since 2000 [59] and, according to a study by the John M. Olin School of Business at Washington University, it is estimated that about 40 percent of today's Fortune 500 companies will no longer exist by 2025 [60].

In addition to the lack of executive commitment described in the introduction, the lack of long-term direction for the transition to a Lean production system is the primary cause for failure. This is shown clearly by the over 30 years of failed efficiency and quality campaigns of automotive manufacturers [46] as well as by the results of VDI working groups (Association of German Engineers) [54, 55]. Although the change to Lean production is generally initiated relatively well, it is hardly followed up by top management, with the result that any achieved success will fade away over the next years. This happens not only once, but several times in wave movements over three decades.

For example, a company wins the business award of EFQM (European Foundation for Quality Management) for their achievements, but after some time all their success has disappeared due to a complete lack of attention by management. The benefits of the Continuous Improvement Process (CIP) were certified as a complete success by Lopez' office at Volkswagen (VW) in 1994, as it had led to a 21% increase in productivity and saved several billion Deutsche Mark, the German currency at that time. The goals for 1994 were 5000 to 6000 CIP workshops and the training of 800 CIP instructors, who would be responsible for the training of employees at VW and its suppliers [61, 62]. However, in spite of the good intentions and optimism, nothing survived from the implementation of Lopez's reforms at VW in the 1990s [46].

Mountain safety and long-term sustainable business development ensure survival in a hostile nature and competitive business environment, respectively. However, both require endurance, as no mountain can be climbed without endurance, and no sustainable change process can occur in businesses without long-term endurance. This is not a quick fix project, but a very long-term process lasting 10, 15 or even 20 years. Endurance is required over and over again.

The importance of endurance, or long-term development, is also shown by the quote from Guan Zhong (720-645 BC) [63]:

It takes ten years to grow a tree, but a hundred years to develop people.

十年□ 木, 百年□ 人.

"The best investment for one year is to grow crops; the best investment for ten years is to plant trees; the best investment for a lifetime is to educate people. What you gain from the growth of one year is grains; what you gain from the growth of ten years is trees; what you gain from the growth of a hundred years is talented high-quality people."

The meaning of this quote is that it takes time to educate people and develop them into productive members of our society. It is not a work such as growing crops or trees, where you can see the pay-off in a year or ten years. But it is a rewarding effort in which our society needs to invest.

But as long companies continue to set their direction in the form of quarterly reports, changes, such as the transition to a Lean production system, will be predominantly accompanied by many setbacks, failures, and patchwork.

In our fast-moving times, short-term profit beats the long-term, sustainable development of people & society!

6 REFERENCES

[1] Hall, R.W. (1981). *Driving the Productivity Machine: Production Planning and Control in Japan*. American Production and Inventory Control Society, Falls Church, VA.

[2] Hayes, R. H. (1981). Why Japanese Factories Work. *Harvard Business Review*, 57-66.

[3] Shingo, S. (1989). *A Study of the Toyota Production System from an Industrial Engineering Viewpoint (Produce What Is Needed, When It's Needed)*. Productivity Press, Cambridge, MA, (in Japanese 1981).

[4] Monden, Y. (1983). *Toyota Production System: Practical Approach to Production Management*.

[5] Ohno, T. (1988). *Toyota Production System – Beyond Large-Scale Production*. Productivity Press, Cambridge, Massachusetts.

[6] Schonberger, R. J. (1982). *Japanese Manufacturing Techniques: Nine Hidden Lessons in Simplicity*. Free Press, New York.

[7] Shingo, S. (1985). *A Revolution in Manufacturing: The SMED System*. Productivity Press, Portland.

[8] Reda, H. M. (1987). A review of Kanban—the Japanese Just-In-Time production system. *Engineering Management International, 4*(2), 143-150. https://doi.org/10.1016/0167-5419(87)90019-6

[9] Nakajima, S. (1988). *Introduction to TPM: Total Productive Maintenance*. Productivity Press.

[10] Womack, J. P., Jones, D. T., & Roos, D. (1990). *The Machine that changed the World*. Macmillan Publishing Company, New York.

[11] … (1990). Six Sigma. Verlag Moderne Industrie, Landsberg, Germany.

[12] Sekine, K. & Arai, A. (1995). *Kaizen für schnelles Umrüsten*. German edition, publisher and introduction S. Schmidt, Verlag Moderne Industrie, Landsberg. (in German)

[13] Nakajima, S. (1995). Introduction to TPM. Productivity Press, Portland, OR, German edition, publisher and introduction S. Schmidt, translation from English by Isabelle Gräfin Grote and S. Schmidt, Campus Verlag, Frankfurt/New York.

[14] … (2004). Toyota überholt Ford. *Frankfurter Allgemeine Zeitung*. https://www.faz.net/-gqe-ol2n, Accessed on 09 April 2023.

[15] Durin, M. (2023). Why are most companies failing with Lean implementation? Kaizen Institute China (ed.). https://kaizen.com/insights/why-are-most-companies-failing-with-lean-implementation/, Accessed on 09 April 2023.

[16] Kallage, R. (2006). Lean implementation failures. *Stamping Journal*. https://-www.thefabricator.com/stampingjournal/article/shopmanagement/lean-implementation-failures, Accessed on 09 April 2023.
[17] Thelen, M. (2011). Why Lean Manufacturing Fails. isixsigma (ed.) https://www.isixsigma.com/lean-methodology/why-lean-manufacturing-fails/, Accessed on 09 April 2023.
[18] Achanga, P., Shehab, E., Roy, R., & Nelder, G. (2006). Critical success factors for lean implementation within SMEs. *Journal of Manufacturing Technology Management, 17*(4), 460-471. https://doi.org/10.1108/17410380610662889
[19] Miina, A. (2012). Lean Problem: Why Companies Fail with Lean Implementation? *Management, 2*(5), 232-250. https://doi.org/10.5923/j.mm.20120205.12
[20] Turbinde, D. (2013). 10 Reasons Why Lean Projects Fail. I.B.I.S. Inc. (ed.), *Industry Week*. https://ibisinc.com/blog/10-reasons-why-lean-projects-fail/, Accessed on 09 April 2023.
[21] Jadhav, J. R., Mantha, S., & Rane, S. B. (2014). Exploring barriers in lean implementation. *International Journal of Six Sigma, 5*(2), 122-148. https://doi.org/10.1108/IJLSS-12-2012-0014
[22] Ledbetter, P. (2020). Why Do So Many Lean Efforts Fail? *Industry Week*. https://www.industryweek.com/operations/continuous-improvement/article/21144299/why-do-so-many-lean-efforts-fail, Accessed on 09 April 2023.
[23] Touhidul Islam, A. S. M. (2020). Lean Fails a Lot, Even Today – Are Organizations Taking Care of All Success Factors to Implement Lean? *The IUP Journal of Operations Management, 19*(2), 29-50. Available at SSRN: https://ssrn.com/abstract=3661146, Accessed on 09 April 2023.
[24] … (2021). 7 Reasons Why Companies Fail at Lean Manufacturing. Redviking (ed.), https://www.redviking.com/blog/7-Reasons-Why-Companies-Fail-at-Lean-Manufacturing_AE160.html, Accessed on 09 April 2023.
[25] Christiansen, B. (2022). Guest Post: 4 Tips for Avoiding Difficulties When Implementing Lean Manufacturing. Six Sigma Daily. https://www.sixsigmadaily.com/4-tips-for-avoiding-difficulties-when-implementing-lean-manufacturing/, Accessed on 09 April 2023.
[26] Pearce, A., Pons, D., & Neitzert, T. (2018). Implementing lean—Outcomes from SME case studies. *Operations Research Perspectives*, 5, 94-104. https://doi.org/10.1016/j.orp.2018.02.002
[27] Suzuki, T. (1994). *TPM in Process Industries (Step-By-Step Approach to TPM Implementation*. Productivity Press (ed.), Taylor & Francis Group, LLC, Boca Raton, FL, p. 4.
[28] Schmidt, S. (2017). Bei BMW die PQCDSM-Logik zur Strukturierung von KPI anwenden (PQCDSM at BMW). Internal document BMW AG (ed.), https://plaza.bmwgroup.net/docs/DOC-56487, Accessed on 09 April 2023.
[29] … TPM Award Assessment Achievement Sheet. Japan Institute of Plant Maintenance (ed.). Available at Achievement Sheet (Sample), Accessed on 09 April 2023.
[30] Schubert, P. (2019). *Sicherheit und Risiko, Band 1*. Bergverlag Rother GmbH, Oberhaching.
[31] Hoffmann, M., Bolesch, R., & Hofmann, G. (2021). *Alpin-Lehrplan 6: Wetter und Orientierung*. Bergverlag Rother GmbH, Oberhaching/Germany. (in German)
[32] Thillet, J.-J. & Schueller, D. (2017). *Wetter im Gebirge*. Bergverlag Rother GmbH, Oberhaching. (in German)
[33] Schmidt, S. (1998). Logbook early mountaineering tours and expeditions in the 70s to 90s (Alps, Japan, Africa, Iran, Himalaya, USA, Mexico). Internal document Health and Safety department of premium automotive manufacturer, Munich.
[34] Schmidt, S. (2020). Logbook mountaineering tours and expeditions 2010 to 2020 (Cordilleras, Himalaya, Alps, Karakoram, Hindukush). Internal document Health and Safety department of premium automotive manufacturer, Munich.
[35] Simon Wang, S.-Y., Gillies, R. R., Fosu, B., & Singh, P. M. (2015). The Deadly Himalayan Snowstorm of October 2014: Synoptic Conditions and Associated Trends. *Bulletin of the American Meteorological Society, 96*(12), 89-94. https://doi.org/10.1175/ BAMS-D-15-00113.1
[36] … (2023). Why Is Lukla Nicknamed the World's Most Dangerous Airport. *Simply Flying – The leading independent voice for aviation news and insight* (ed.), https://simpleflying.com/lukla-most-dangerous-airport/, Accessed on 16 April 2023.
[37] … (2016). Slow Onset Hypoxia. *Code 7700* (ed.), https://www.code7700.com/hypoxia_slow_onset.htm, Accessed on 09 April 2023.
[38] Schmidt, S. (2018). Höhenkrankheit: Grundlagen und eigene Erfahrungen (Altitude sickness: basics and own experiences). Internal document Health and Safety department of premium automotive manufacturer, Munich, 1-18. (in German)
[39] Feddersen, B. & Ausserer, H. (2016). *Bergmedizin Expeditionsmedizin pocket: Von Tagesausflug bis Himalaya-Expedition!* Börm Bruckmeier, Grünwald/Germany. (in German)
[40] Dehnert, C. & Bärtsch, P. (2017). Akute Bergkrankheit und Höhenhirnödem. *Höhenmedizin,* 74(12). (in German) https://doi.org/10.1024/0040-5930/a000954
[41] Pomorski, T. R. (2004). *Total Productive Maintenance (TPM) Concepts and Literature Review*, 62-66. https://tpmindonesia.files.wordpress.com/2010/01/01-tpm-book1.pdf, Accessed on 09 April 2023.
[42] Manivannan, S., Schmidt, B. S. G., & Schmidt, S. (2012). KPI's – Werkzeuge für exzellentes Lean Management. *Technologie & Management, 1*, 12-19.
[43] Schmidt, S. (2017). Strategies for improving operational performance. Presentation at first EU-JAPAN Lean Summit, The EU-Japan Centre for Industrial Cooperation (EUJC), Brussel, 1-18.
[44] … *Lean Management Training programmes, Business services, round tables, help desks*. EU-Japan Centre for Industrial Cooperation Brussel (ed.). https://www.eu-japan.eu/lean-management, Accessed on 09 April 2023.
[45] Schmidt, S. (2019). *Lean in Europe - visit 26 - BMW Group Plant Berlin*. EU-Japan Centre for Industrial Cooperation, Brussel (ed.), https://www.eu-japan.eu/events/lean-europe-visit-26-bmw-group-plant-berlin, Accessed on 09 April 2023.
[46] Schmidt, S. (2020). *Over 30 years of failed efficiency and quality campaigns of car manufacturers*. Internal document premium automotive manufacturer, Munich.
[47] Schmidt, S. (2015). Operational Excellence – Way to Sustainable World-Class Level. *Presentation at MOTSP 2015, 7th International Scientific Conference Management of Technology Step to Sustainable Production*, Brela, Makarska, Croatia.
[48] Walsh, D. (2019). *What Climbing Expeditions Tell Us about Teamwork?* Stanford Graduate School of Business (ed.), https://www.gsb.stanford.edu/insights/what-climbing-expeditions-tell-us-about-teamwork, Accessed on 09 April 2023.
[49] Maxwell, C. (2015). *Lead Like a Guide: Adopt the Strengths of Mountaineers*. The Wharton School, The University of Pennsylvania (ed.), https://executiveeducation.wharton.upenn.edu/thought-leadership/wharton-at-work/2015/07/lead-like-a-guide/, Accessed on 09 April 2023.
[50] Digenti, D. (2021). Lessons from Everest: The role of collaborative leadership in crisis. Leverage Networks, Inc. (ed.), *The Systems Thinker™, 12*(2). https://thesystemsthinker.com/lessons-from-everest-the-role-of-collaborative-leadership-in-crisis/, Accessed on 09 April 2023.

[51] Eaton, R. (2017). How mountaineering is like business leadership (and how it's not). Canadian Association of Management Consultants CMC-Canada (ed.), *Consult Magazine*. https://www.cmc-canada.ca/consult/how-mountaineering-is-like-business-leadership-and-how-its-not.
[52] Alvernaz, A. (2019). *5 Leadership Skills You Can Develop in The Mountains*. Trello, Inc. (ed.), https://blog.trello.com/leadership-skills-you-can-develop, Accessed on 09 April 2023.
[53] Timmermann, L. (2018). Nine observations on leadership and teamwork from Mt. Everest. Informa Connect Limited, London (ed.), https://informaconnect.com/nine-observations-on-leadership-and-teamwork-from-mt-everest/, Accessed on 09 April 2023.
[54] … (2012). Lean production systems – Basic principles, introduction, and review, VDI Guideline VDI 2870 part 1. Verein Deutscher Ingenieure (ed.). *Lean production systems - List of methods*, VDI Guideline 2870 part 2 (2013), Verein Deutscher Ingenieure (ed.).
[55] … (2017). Lean production systems – Lean leadership, VDI Guideline VDI 2871 part 1. Verein Deutscher Ingenieure (ed.). *Lean production systems - Lean leadership - List of methods*, VDI Guideline 2871 part 2 (2021), Verein Deutscher Ingenieure (ed.).
[56] Schmidt, S. & Schmidt, B. S. G. (2022). A Working from Home (WFH) 2022 - Best Practices Implementation and Case Study of Company Agreement. Tehnički glasnik, 16(4), 572-580. https://doi.org/10.31803/tg-20220511094430
[57] … (2019). Unternehmen werden in Schnitt nur 9 Jahre alt. *Wirtschafts Kurier*. (in German) https://www.wirtschaftskurier.de/artikel/unternehmen-werden-im-schnitt-nur-9-jahre-alt.html, Accessed on 09 April 2023.
[58] Braun, S. (2022). Gründungen und Unternehmensschließungen - Überlebensrate von Unternehmen. Institut für Mittelstandsforschung IfM Bonn (ed.) (in German) https://www.ifm-bonn.org/statistiken/gruendungen-und-unternehmensschliessungen/ueberlebensrate-von-unternehmen, Accessed on 09 04 2023.
[59] Kroker, M. (2016). Digitale Transformation: 40 Prozent der Fortune-500-Firmen verschwinden in nächster Dekade. Handelsblatt GmbH (ed.), Kroker's Look @ IT. (in German) https://blog.wiwo.de/look-at-it/2016/08/24/digitale-transformation-40-prozent-der-fortune-500-firmen-ver schwinden-in-naechster-dekade/, Accessed on 09 April 2023.
[60] Ioannou, L. (2014). A decade to mass extinction event in S&P 500. CNBC LLC. (ed.), *The Exponential Future of Finance*. https://www.cnbc.com/2014/06/04/15-years-to-extinction-sp-500-companies.html, Accessed on 09 April 2023.
[61] Schmidt, S. & Schmidt, B. S. G. (2018). Review of 35 years of Lean Production – outlook development to lean, agile, and IoT. *Presentation at MOTSP 2018, 10th International Scientific Conference Management of Technology Step to Sustainable Production*, Primošten, Croatia.
[62] López De Arriorúa, J. I. (1994). Der Nutzen von KVP – López: KVP voller Erfolg. *Die Welt*.
[63] Patil, A. (2012). *It takes ten years to grow a tree, but a hundred years to cultivate people*. Berkshire Publishing Group (ed.), China Connect University. https://www.berkshirepublishing.com/chinaconnectu/2012/02/06/it-takes-ten-years-to-grow-a-tree-but-a-hundred-years-to-cultivate-people-十年树木-百年树人十年树木-百/.

Authors' contacts:

Stefan Schmidt, Dipl.-Ing., Dipl.-Wirtschafting.
Technology Management and Management Training/VWI/VDI,
Fritzstrasse 41, D-82140 Olching, Germany
Stefan.Schmidt-1@t-online.de

Benjamin S. G. Schmidt, MA
Ludwigs Maximilian University,
Geschwister-Scholl-Platz 1
80539 Munich, Germany

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230518185836

Subject review
Received: 2023-05-18, Accepted: 2023-05-22

Use of Green Industry 5.0 Technologies in Logistics Activities

Maja Trstenjak*, Miljenko Mustapić, Petar Gregurić, Tihomir Opetuk

Abstract: Industry 5.0 is a human-centred concept of industrial development towards the sustainable and resilient system presented by the European Union which aims to become the global both innovation and industrial leader. It should overcome the barriers of the previously presented Industry 4.0. This paper presents the research conducted in the 112 Croatian manufacturing companies, dealing with their awareness level of the Industry 5.0, as well as the use of green and digital elements in logistics activities. The results have shown that the awareness of the digital concept of both Industry 4.0 or 5.0 remains low, but the companies are more open towards the implementation of the green elements than the digital ones, with the potential for future development recognized.

Keywords: ergonomics; Industry 4.0; Industry 5.0; green industry; green logistics; logistics 4.0; sustainability

1 INTRODUCTION

Industry 5.0 is a strategy presented by the European Union to overcome the shortcomings of Industry 4.0 implementation. Regarding the many unexpected distributions of the market industry must be resilient, sustainable, and human-centred [1]. Green transition, therefore, remains imperative, with a human worker placed again in the centre of the system to improve efficiency and productivity with a special contribution to general society [2]. Physical worker remains therefore a valuable resource, but the jobs now are improved to increase safety and avoid the decrease of productivity due to monotonous repetitive tasks. Therefore, new skills and knowledge is required, to provide the optimal transition towards the new system. Lack of needed knowledge and skills was previously recognized as one of the most common barriers in the implementation of Industry 4.0, included in control, optimization, and decision-making processes rather than manual work [3]. Industry 4.0 technologies in the beginning were subject to availability and required a very high investment cost, but with time, their availability on the market has increased, while the trend of high variability and fast customization of products remained, which requires a design of systems of high flexibility and modularity with unclear predictable benefits in the future [4]. Therefore, Industry 5.0 focuses on developing a human-centered, sustainable, and resilient production system, which should answer the market demands, but also the unpredictable local and global events in society that might occur and affect the production in a negative manner [5]. Sustainability is encouraged to be achieved by using green technologies following the principles of circular economy, which also implies the use of the renewable energy resources in the production, but also a recyclable and re-useable materials, resources, and tools [6].

In this paper, the research will be conducted to get an overview of the current state of the Croatian manufacturing industry regarding the use of the green principles in the logistics activities such as warehousing, transport and picking. The implementation and awareness of green and digital elements will be examined and compared to the previous researches, not only in Croatia, but also on the global level. Based on the research results, the potential for the future development of the Croatian manufacturing companies will be presented.

2 IMPACT OF INDUSTRY 4.0 AND 5.0 ON GREEN LOGISTICS

The Web of Knowledge platform was browsed, as the most relevant database of published scientific work. The searched terms were "Green Logistics", "Logistics 5.0" and "Industry 4.0" or "Industry 5.0". The timespan was set from 2011 to 2023 since Industry 4.0 dates back to 2011. The research field was limited to "engineering, industrial", "engineering, manufacturing" and "engineering, multidisciplinary".

The green initiative in the manufacturing processes can often be used as a marketing asset, while many works in the literature deal with the relation between the actual implementation of green technologies within the manufacturing companies and the corporate brand strategy. Environmental awareness should certainly be part of the business strategy, but alongside continuous integration and evaluation of green elements [7]. Local governance regulations and taxes have an influence on the green and digital technologies implementation in the logistics processes [8]. The price of conventional vehicles, such as a carbon tax, may lead to both an increase and a decrease in environmental performance [9].

Green technologies enable significant energy savings, especially with the use of renewable energy sources. Energy management of warehousing tends to be one of the key factors in the future development of logistics activities of manufacturing companies. Industry 5.0 technologies enable the monitoring of energy consumption in real-time, as well as its optimization.

With adequate energy management, sustainable and green development can be achieved, which significantly impacts warehouse building and its management [10]. Different material handling activities have different constraints in this matter, while adopting smart automatic

picking systems by Industry 4.0 standards increases energy efficiency. In this case, managerial strategies also play an important role in the adoption of available equipment to increase warehouse productivity at negligible costs [11]. When comparing manual and fully automated warehouses, it is suggested that an energy balance should be established for the material handling equipment, energy consumption for building maintenance (heating, cooling, lighting, etc.), and energy generated by the photovoltaic system on the roof. Part of the energy consumed is noticed to be spent on maintenance activities, especially in the case of facilities with a low degree of automation [12, 13]. Use of advanced optimization methods have a positive impact on energy optimization, such as genetic algorithms, hybrid simulated annealing and tempering algorithm [14] two-step optimization model based on integer programming or Mixed Integer Linear Programming [15]. This can generate higher investment costs in the beginning but enables the sustainability of the system in the future [16].

One of the key goals in the EU, but also in global environmental strategies, is the reduction of the carbon footprint. The reduction of the carbon footprint enables the decrease of overall costs and leads to economic sustainability. 10% of worldwide CO_2 emissions derive from logistical supply chains, while 20% of the overall logistical cost relates to the amount of energy required for heating, cooling, and lighting as well as material handling equipment [17]. Therefore, optimal managerial decisions should be made, for both economic benefits and environmental impacts [18].

One of the motivational factors in implementing green warehousing is social responsibility, while one of the biggest barriers was local law and regulation. Therefore, it is suggested that top management be the key initiator of green technologies implementation in the warehouse. Moreover, waste reduction through green management can improve employees' living conditions and productivity, by Industry 5.0 human-centric and sustainable standards [19].

Green transportation methods and vehicles can have a positive impact on energy efficiency and reduction of the carbon footprint but there are several challenges for the internal and external transportation system to remain efficient and achieve sustainability of a system.

Transport is referred to as the fastest-growing source of greenhouse gas emissions. It should provide environmental safety, new customer relationships, and product experience [20].

The performance of electric vehicles has numerous constraints such as battery performance, technological advances, and energy management, so routing challenges must be considered [21].

Routing optimization is also shown to be very useful in improving the environmental impact [22, 23], while the decisions on supply lead time, reorder quantities, and storage equipment also have an impact on costs and emissions [24].

Using biofuels is another way to reduce environmental impact in transport, but it also has certain limitations. The main risks can be such as lack of investor confidence in biofuel developments (the highest score); energy or fuel security issues; negative public perception of biofuels (equal second highest); increased food prices; high barriers to entry into the fuel market; and misdirected agricultural expansion or land use (equal fifth highest) [25].

Proper human interaction with technology within the human-centric systems by Industry 5.0 standards [26].

Collaborative robots in the logistics sector enable cost savings, as well as the reduction of CO_2 emissions [27], but the workplace should be designed to create a safe environment for the human worker [28].

Automatic warehousing systems are one of the green technologies and their implementation is influenced by perceived advantage, cost, technological turbulence, business partner influence, firm size, firm scope, and operational performance, especially in SMEs [29].

3 RESEARCH DESIGN

The goal of this research was to get a general insight into the current state, potentials, and challenges in the implementation of green and digital logistics activities of Croatian manufacturing companies. Therefore, a survey was created based on the theoretical framework and research results from the previously published works, explained in Chapter 2.

For the purposes of the research structuring, the following research questions were set as a guidance:

RQ1: *What is the awareness level of the Croatian manufacturing companies with Industry 4.0 and 5.0 elements and its implementation into logistics activities?*

RQ2: *What is the awareness level of the green digital elements in Logistics in the Croatian manufacturing companies?*

RQ3: *What is the perception of the implementation and use of the green and digital elements of logistics in Croatian manufacturing companies?*

The research questions lead to definition of target group of the research, which was Croatian manufacturing companies. The companies included in the research were found in the digital repository of the Croatian Chamber of Economy, which is publicly available. This was a starting point for the web search of contact of the companies. Therefore, the target group was manufacturing companies, not exclusively specialized for warehousing or transport, but with warehousing and transport as a standard part of their organizational system. The data was collected through an online questionnaire, structured in the Google Forms online application, and sent to 952 active manufacturers with available contact. The results were received from a total of 112 participants, which makes a total response rate of 11.76%. Since the survey was sent to manufacturing companies with no limitations of their industrial field, there are various types of participants, but one of the most common was from the metal machining industry (27,6%) and process industry (18%).

The survey was structured to have four parts. In the first part, basic information about the manufacturer was collected,

such as the company size, years of active presence on the market, number of participant's work experience years within the current company, and if the green technologies and the use of renewable energy sources are part of the corporative brand and strategy. The second part of the survey was related to the current level of digitalization and the general use of digital technologies. Familiarity with Industry 4.0 or 5.0, the current level of digitalization within the company, challenges in the digitalization process as well as the current use of green technologies, renewable energy sources, and interest in green technologies implementation was examined. In the third part, the participants had to answer questions regarding digitalization and green technologies implementation in logistics activities. The data was collected in March 2023, while the statistical analysis was provided by IBM SPSS v27 software.

The data will be analyzed with z-test, t-test (with Bonferroni correction), and the Pearson correlation coefficient.

A z-test [31] is used to determine whether two population means are different when the variances are known and the sample size is large ($n > 30$). A z-test is used in hypothesis testing to evaluate whether a finding or association is statistically significant or not. In particular, it tests whether two means are the same (the null hypothesis). A z-test can only be used if the population standard deviation is known and the sample size is 30 data points or larger. Otherwise, a t-test will be employed.

For Null hypothesis H0: $\mu = \mu_0$ vs alternative hypothesis H1: $\mu \neq \mu_0$, two-tailed test is used.

A t-test [32] is a statistical test that is used to compare the means of two groups. It is used in hypothesis testing to determine whether two groups are different from one another. The 95% confidence interval is considered. This is the range of numbers within which the true difference in means will be 95% of the time.

The Bonferroni test is a type of multiple comparison test used in statistical analysis. When performing a hypothesis test with multiple comparisons, eventually, a result could occur that appears to demonstrate statistical significance in the dependent variable, even when there is none. The Bonferroni test is a statistical test used to reduce the instance of a false positive. The Bonferroni test, also known as "Bonferroni correction" or "Bonferroni adjustment" suggests that the p-value for each test must be equal to its alpha divided by the number of tests performed.

The Pearson correlation coefficient (r) [33] is the most common way of measuring a linear correlation. It is a number between –1 and 1 that measures the strength and direction of the relationship between two variables.

4 RESULTS

As shown in Fig. 1, more than half of the companies which participated in the survey are not yet familiar with Industry 4.0. Among those familiar with the concept, only 27.7% of the companies have already implemented certain element of Industry 4.0 or 5.0 (Fig. 2) while 28.6% did not. Size of the company is shown to be influential on the familiarity with the Industry 4.0 or 5.0 and the significant difference in familiarity has been proven between the middle and large companies compared to micro and small companies ($p < 0,05$).

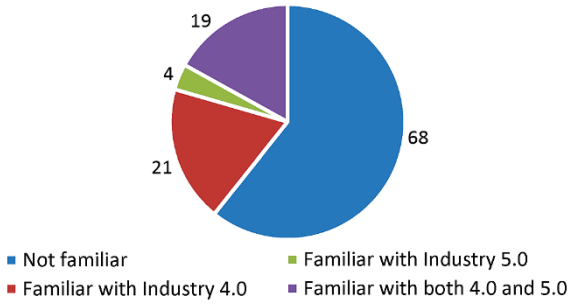


Figure 1 Familiarity with Industry 4.0 and 5.0

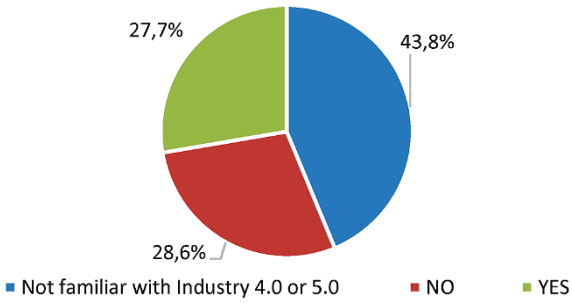


Figure 2 Implementation of Industry 4.0 or 5.0 elements (in general)

The companies have rated their interest for the digitalization with an average grade of 3.47 in which 1 represents no interest at all interest and 5 very high interest, while their approximation of the current digitalization level is 2.82, where 1 represents no elements of digitalization present in the company and 5 represents that all the processes in the company are digitalized. Regarding the use of green technologies, the average grade is 2.53 with a significant difference noticed between the companies with and without green corporate strategy. Similar difference is found regarding the interest for the implementation of renewable energy resources, which was rated with 3.48.

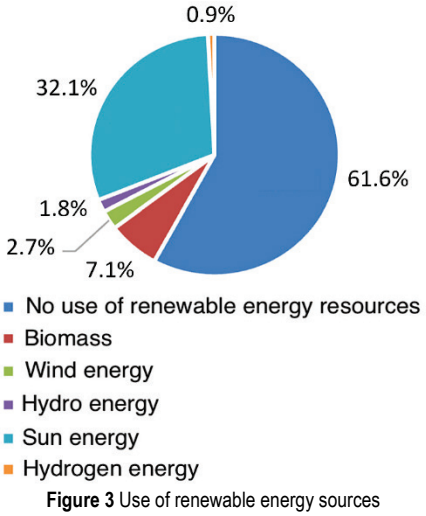


Figure 3 Use of renewable energy sources

As shown in Fig. 3, majority of the companies (61.6%) do not use renewable energy sources in their processes while most common used renewable energy source is Sun energy 32.1%.

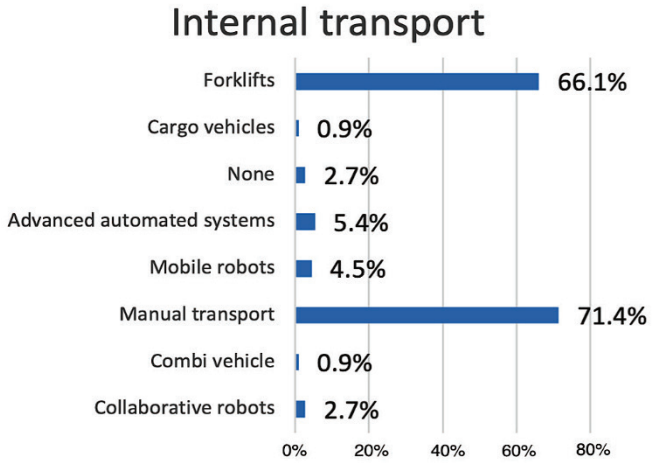


Figure 4 Use of renewable energy sources

In internal transport, most companies rely on manual transport (71.4%) and the use of forklifts (66.1%). Less than 5% of the companies use robots while only 5.4% use advanced automated transport systems.

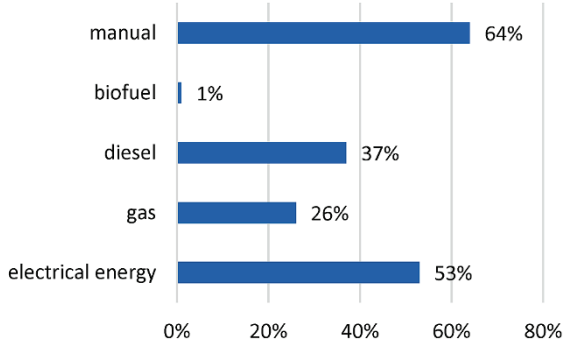


Figure 5 Internal transport drive

Results in Fig. 5 reveal that the most common drive of internal transport is manual, followed by electrical energy (53%). Only 1% of the companies use biofuel as the drive for internal transport vehicles. The need for the automatization of internal transport was rated with 2.64 where 1 represents that there is no need for the automatization and 5 that there is a very high need for automatization. There is a significant difference noticed ($p < 0.05$) regarding the company size and the presence of a green corporate strategy in the company. Similarly, the need for the implementation of renewable energy sources in internal transport is rated with 2.98, with a significant difference noticed groups those with and without green corporate strategy. The influence of the human in the internal transport is rated with 4.03, where 1 represents no influence at all and 5 very high influence.

Since Industry 5.0 is a human-centred system, with a special emphasis on human safety and well-being, the safety methods and measures in the internal logistics activities were examined. Most companies use protective footwear and clothing, while only 25% have implemented sensors for stopping the machine in cases of potential danger. The average grade of the safety of the human worker in the working area is rated 4.0.

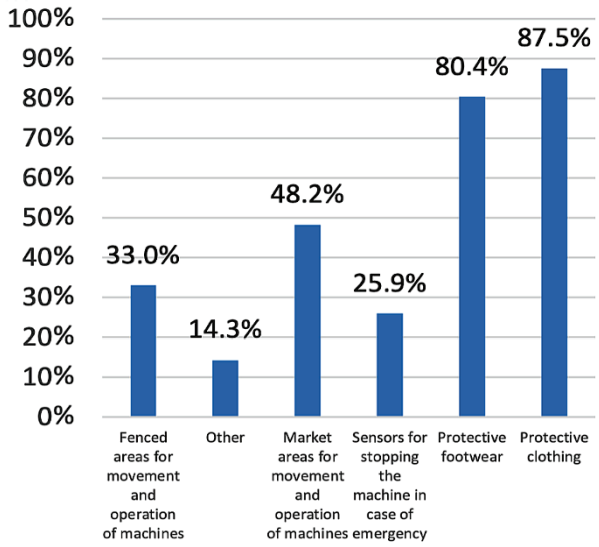


Figure 6 Safety methods and measures in internal logistics

Regarding reusable containers, pallets, and other material handling equipment, 79.5% of the companies have stated that they use it on a regular basis, with no significant differences noticed between the groups. The level of resource recycling is rated at 3.40 where a significant difference is noticed between the companies with and without green corporate strategy. The monitoring of the energy efficiency of the internal transport vehicles is rated at 2.48.

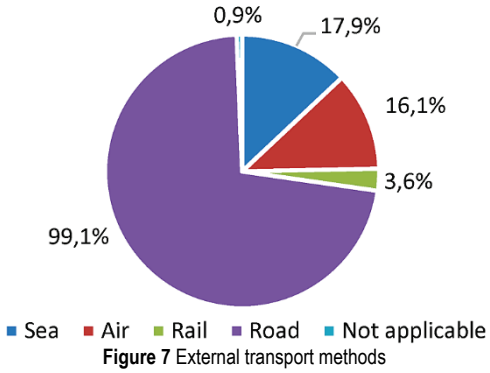


Figure 7 External transport methods

When it comes to external transport, most of the companies in Croatia use road transport, as could be expected since this kind of transport is the most common with the highest proportion among the other transportation possibilities. This is why 92% of the companies use diesel fuel for the drive of external transport vehicles, only 5.4% use electrical driving while 2.7% use gas. The need for autonomous vehicles is low, rated with 1.95 with a significant difference noticed between those familiar and unfamiliar with Industry 4.0 or 5.0. 60.7% of the companies do not track the driver's condition in real-time, while 39.3%

claim that they do. The predicted savings of the implementation of the green technologies in outer transport is rated with 2.49 where 1 represents the prediction of the minimal savings and 5 very high savings.

5 DISCUSSION

Industry 4.0 was presented back in 2011, so for more than ten years it is present and known on the market. The evidence of its complete implementation in the manufacturing system on a global level remains low, although the implementation of certain elements becomes more common over time. Yet, as the research results have shown, 60% of participants are not even familiar with this term. Compared to the previously conducted research, from 2020 [33], the number of those unfamiliar with Industry 4.0 has increased, which answers RQ1. The same time, previous research was focused on the metal machining industry in which only 46% of the participants were not familiar with the term. On the other hand, in the research conducted in 2016, the average maturity level of the Croatian manufacturing industry is calculated as 2.15 [34], which means that, in average, Croatian manufacturing companies haven't reached the 3rd industrial revolution yet, which implies the use of computer technology. The current perception of the digitalization level of Croatian manufacturing companies is 2.82, which could not be described as a very high perception of digitalization, which means that there are many opportunities for improvement. In this research, no correlation was found between the level of digitalization or familiarity with Industry 4.0/5.0 and the level of interest in implementing renewable energy sources in the company. Interestingly, 70% of the participants have green technologies implemented in the corporate brand or strategy. The higher awareness of digital and green elements and the interest for its implementation was noticed within the companies, which have a green corporate strategy, and in many cases, the significant differences were found compared to those without green corporate strategy, which answers RQ2. The role of the human in the system remains high, with a large percentage of manual transportation (71.4%), while also the perceived level of safety of the worker is high (4.08), along with their awareness of safety measures and regulations (4.02), which can be a good starting point in transition towards the human-centred Industry 5.0 concept. The use of the renewable energy resources are more common in the internal than in the external transport, mainly due to the better availability of the technology (i.e. electric forklifts). The interest for the use of renewable energy resources in the external transport, or even advanced autonomous vehicles remains low, as well as the prediction of the possible savings by its implementation, which answers RQ3.

Interestingly, when observing the results, the medium and large companies in Croatia have shown a bigger interest and awareness towards the digital and green technologies. They were observed as a single group compared towards the micro and small companies. The interest in digitalization of the processes is approximated with 3.42, while those from the micro companies have an approximation of only 2.64.

The financial barrier as in very high investments when implementing digital and green elements does not seem to be a most common barrier within the Croatian companies. The lack of time to develop and implement new technologies and, unavailability of the technologies on the current market and workers' resistance to change have been found as the greatest barriers in implementation of Industry 4.0 or 5.0. The average influence of rate of energy price towards the final product is 3.40, the highest in medium and large companies (4.0) and the lowest again in companies without green strategy (2.88). They perceive customers as not ready to pay for the implementation of green technologies, with an average grade of 1.79, with a difference noticed between the groups in green strategy (1.5/1.92; $p = 0.026$).

The results of the research could lead to the development of a future strategy for the transformation of the Croatian manufacturing industry towards a green, sustainable, and resilient system by the Industry 5.0 standards. The first step s raising awareness about the concept since familiarity among the companies remains low. Especially it is important to direct the companies towards the human-centred system to overcome the barriers and challenges of Industry 4.0. The ratio of human manual work remains high, which could be a positive aspect to continue the development of new skills and knowledge of human workers. Special caution should be aimed toward the ergonomic design of a workplace and human well-being for motivational level to remain high which directly impacts the productivity. The awareness of a green digital transition is on a relatively satisfactory level, while in many parts the significant difference in the groups of those with or without green corporate strategy is noticed, both regarding the implementation green elements of the elements of Industry 4.0 or 5.0. Green and digital can lead to the positive future development of the industry with a high competitiveness not only on the local, but also on the global market.

6 CONCLUSION

Industry 5.0 is not only a trending concept but is also a needed goal for companies to achieve to remain sustainable and resilient, to keep the human as the most valuable resource in the centre of the manufacturing system, and to remain competitive in the market. Industry 5.0 elements should therefore be implemented in all the segments of the manufacturing companies, such as the logistics with its processes and activities. The digital concept demands nowadays the implementation of green elements whose combination leads to enabling the sustainable development of the company and the local and global industry. The results have shown that the awareness within the Croatian manufacturing companies about Industry 5.0 remains relatively low, while the companies are more open towards the green elements and their implementation rather than the digital. Renewable energy resources are rarely used, as well as energy efficiency tracking, but reusable and recyclable pallets, containers, and other material handling tools are often

used. In internal transport, the electric drive of vehicles is being used, but the energy is not directly coming from renewable sources. The other transport depends on road and diesel fuels while the majority of those with green corporate strategy recognizes the positive impact of future renewable energy resources implementation. As for future work, according to the results, the model for the strategic transformation of a single company towards the green Industry 5.0 should be developed.

7 REFERENCES

[1] Industry 5.0 Available online: https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/industry-50_en (accessed on 20 March 2023).

[2] Nahavandi, S. (2019). Industry 5.0—A Human-Centric Solution. *Sustainability, 11,* 4371. https://doi.org/10.3390/SU11164371

[3] Stentoft, J., Adsbøll Wickstrøm, K., Philipsen, K., & Haug, A. (2021). Drivers and Barriers for Industry 4.0 Readiness and Practice: Empirical Evidence from Small and Medium-Sized Manufacturers. *Production Planning and Control, 32*, 811-828. https://doi.org/10.1080/09537287.2020.1768318

[4] Ghobakhloo, M. (2020). Industry 4.0, Digitization, and Opportunities for Sustainability. *J Clean Prod, 252*, 119869. https://doi.org/10.1016/J.JCLEPRO.2019.119869

[5] Akundi, A., Euresti, D., Luna, S., Ankobiah, W., Lopes, A., & Edinbarough, I. (2022). State of Industry 5.0—Analysis and Identification of Current Research Trends. *Applied System Innovation, 5*. https://doi.org/10.3390/ASI5010027

[6] Potočan, V., Mulej, M., & Nedelko, Z. (2021). Society 5.0: Balancing of Industry 4.0, Economic Advancement and Social Problems. *Kybernetes, 50*, 794-811. https://doi.org/10.1108/K-12-2019-0858/FULL/PDF

[7] Ali, S. S., Kaur, R., & Khan, S. (2023). Evaluating Sustainability Initiatives in Warehouse for Measuring Sustainability Performance: An Emerging Economy Perspective. *Ann Oper Res, 324*, 461-500. https://doi.org/10.1007/S10479-021-04454-W

[8] Chen, J., Liao, W., & Yu, C. (2021). Route Optimization for Cold Chain Logistics of Front Warehouses Based on Traffic Congestion and Carbon Emission. *Comput Ind Eng*, *161*, 107663. https://doi.org/10.1016/J.CIE.2021.107663

[9] Berling, P. & Eng-Larsson, F. (2017). Environmental Implications of Transport Contract Choice - Capacity Investment and Pricing under Volume and Capacity Contracts. *Eur J Oper Res*, *261*, 129-142. https://doi.org/10.1016/J.EJOR.2017.01.056

[10] Bartolini, M., Bottani, E., & Grosse, E. H. (2019). Green Warehousing: Systematic Literature Review and Bibliometric Analysis. *J Clean Prod, 226*, 242-258. https://doi.org/10.1016/J.JCLEPRO.2019.04.055

[11] Carli, R., Digiesi, S., Dotoli, M., & Facchini, F. (2020). A Control Strategy for Smart Energy Charging of Warehouse Material Handling Equipment. *Procedia Manufacturing, 42*, 503-510. https://doi.org/10.1016/j.promfg.2020.02.041

[12] Lewczuk, K., Kłodawski, M., & Gepner, P. (2021). Energy Consumption in a Distributional Warehouse: A Practical Case Study for Different Warehouse Technologies. *Energies*, *14*(9), 2709. https://doi.org/10.3390/EN14092709

[13] Helo, P. & Ala-Harja, H. (2018). Green Logistics in Food Distribution – a Case Study. *International Journal of Logistics Research and Applications*, *21*(4), 464-479. https://doi.org/10.1080/13675567.2017.1421623

[14] Balaji, K. S., Ramasubramanian, B., Vinay, M. S. S., Reddy, D. T., Dheeraj, C., Subash, K.T., & Anbuudayasankar, S. P. (2021). A Demand-Based Relocation of Warehouses and Green Routing. *Mater Today Proc*, *46*(17), 8438-8443. https://doi.org/10.1016/J.MATPR.2021.03.476

[15] Ene, S., Küçükoğlu, İ., Aksoy, A., & Öztürk, N. (2016). A Genetic Algorithm for Minimizing Energy Consumption in Warehouses. *Energy, 114*, 973-980. https://doi.org/10.1016/J.ENERGY.2016.08.045

[16] Carli, R., Dotoli, M., Digiesi, S., Facchini, F., & Mossa, G. (2020). Sustainable Scheduling of Material Handling Activities in Labor-Intensive Warehouses: A Decision and Control Model. *Sustainability, 12*(8), 3111. https://doi.org/10.3390/SU12083111

[17] Chen, X., Wang, X., Kumar, V., & Kumar, N. (2016). Low Carbon Warehouse Management under Cap-and-Trade Policy. *J Clean Prod, 139*, 894-904. https://doi.org/10.1016/J.JCLEPRO.2016.08.089

[18] Sukjit, S. & Vanichchinchai, A. (2020). An Assessment of Motivations on Green Warehousing in Thailand. *The 7th IEEE International Conference on Industrial Engineering and Applications, ICIEA 2020*, 539-542. https://doi.org/10.1109/ICIEA49774.2020.9102035

[19] Larina, I. V., Larin, A. N., Kiriliuk, O., & Ingaldi, M. (2021). Green Logistics - Modern Transportation Process Technology. *Production Engineering Archives, 27*, 184-190. https://doi.org/10.30657/PEA.2021.27.24

[20] Amine, J., Elhassania, M., & El Idrissi Adiba, E. B. (2022). Issues of Transport Problems with Electric Vehicles. *The 14th IEEE International Conference of Logistics and Supply Chain Management, LOGISTIQUA 2022*, 1-6. https://doi.org/10.1109/LOGISTIQUA55056.2022.9938026

[21] Lanza, G., Passacantando, M., & Scutellà, M. G. (2022). The Green Sequencing and Routing Problem. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13557 LNCS*, 231-244. https://doi.org/10.1007/978-3-031-16579-5_16

[22] Liu, Z. & Lan, H. (2013). Analysis of Warehouse Location in Low-Carbon Supply Chain Based on the Cost. In: Chen, F., Liu, Y., & Hua, G. (eds) LTLGB 2012. Springer, Berlin, Heidelberg, 727-733. https://doi.org/10.1007/978-3-642-34651-4_99

[23] Klimecka-Tatar, D., Ingaldi, M., & Obrecht, M. (2021). Sustainable Developement in Logistic-A Strategy for Management in Terms of Green Transport. *Management Systems in Production Engineering, 29*, 91-96. https://doi.org/10.2478/MSPE-2021-0012

[24] Fichtinger, J., Ries, J. M., Grosse, E. H., & Baker, P. (2015). Assessing the Environmental Impact of Integrated Inventory and Warehouse Management. *Int J Prod Econ, 170*, 717-729. https://doi.org/10.1016/J.IJPE.2015.06.025

[25] Hammond, G. P., Howard, H. R., & Tuck, A. (2012). Risk Assessment of UK Biofuel Developments within the Rapidly Evolving Energy and Transport Sectors. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 226*(5), 526-548. https://doi.org/10.1177/1748006X12448147

[26] Grosse, E. H. (2023). Application of Supportive and Substitutive Technologies in Manual Warehouse Order Picking: A Content Analysis. *Int J Prod Res*, 1-20, https://doi.org/10.1080/00207543.2023.2169383

[27] Oran, I. B. & Cezayirlioglu, H. R. (2021). AI - Robotic Applications in Logistics Industry and Savings Calculation. *Journal of Organizational Behavior Research, 6*, 148-165. https://doi.org/10.51847/JUXQMVCVQF

[28] Rega, A., Di Marino, C., Pasquariello, A., Vitolo, F., Patalano, S., Zanella, A., & Lanzotti, A. (2021). Collaborative Workplace Design: A Knowledge-Based Approach to Promote Human–Robot Collaboration and Multi-Objective Layout Optimization. *Applied Sciences, 11*. https://doi.org/10.3390/APP112412147

[29] Hao, J., Shi, H., Shi, V., & Yang, C. (2020). Adoption of Automatic Warehousing Systems in Logistics Firms: A Technology–Organization–Environment Framework. *Sustainability, 12*, 5185. https://doi.org/10.3390/SU12125185

[30] Z-Test Definition: Its Uses in Statistics Simply Explained with Example. https://www.investopedia.com/terms/z/z-test.asp (accessed on 25 March 2023).

[31] An Introduction to t Tests | Definitions, Formula and Examples https://www.scribbr.com/statistics/t-test/ (accessed on 25 March 2023).

[32] Pearson Product-Moment Correlation - When You Should Run This Test, the Range of Values the Coefficient Can Take and How to Measure Strength of Association. https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php (accessed on 25 March 2023).

[33] Trstenjak, M., Opetuk, T., Cajner, H., & Tosanovic, N. (2020). Process Planning in Industry 4.0—Current State, Potential and Management of Transformation. *Sustainability, 12*, 5878, https://doi.org/10.3390/SU12155878

[34] Babić Z., Veža, I., & Pavić, I. (2016). Ranking of Enterprises with Regard to Industrial Maturity Level Using AHP and TOPSIS - CROSBI. In *Proceedings of the Proceedings of ISHAP2016*, Saaty, W. R., Ed., Creative Decisions Foundation: London, 124-125.

Authors' contacts:

Maja Trstenjak, dr. sc.
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lucica 5, 10 000 Zagreb, Croatia
maja.trstenjak@fsb.hr

Miljenko Mustapić, dr. sc.
University North,
Trg dr. Zarka Dolinara 1, 48 000 Koprivnica, Croatia
mmustapic@unin.hr

Petar Gregurić, mag. ing. mech
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lucica 5, 10 000 Zagreb, Croatia
petar.greguric@fsb.hr

Tihomir Opetuk, asist. prof. dr. sc.
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lucica 5, 10 000 Zagreb, Croatia
tihomir.opetuk@fsb.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20.....

Categorization of paper
Received: y-m-d, Accepted: y1-m1-d1

14pt
14pt

Article Title Only in English (Style: Arial Narrow, Bold, 14pt)

14pt

Ivan Horvat, Thomas Johnson, Marko Marić (Style: Arial Narrow, Normal, 10pt)

14pt

Abstract: Article abstract contains maximum of 150 words and is written in the language of the article. The abstract should reflect the content of the article as precisely as possible. TECHNICAL JOURNAL is a trade journal that publishes scientific and professional papers from the domain(s) of mechanical engineering, electrical engineering, civil engineering, multimedia, logistics, etc., and their boundary areas. This document must be used as the template for writing articles so that all the articles have the same layout. (Style: Arial Narraw, 8pt)

8pt

Keywords: keywords in alphabetical order (5-6 key words). Keywords are generally taken from the article title and/or from the abstract. (Style: Arial Narraw, 8pt)

10pt
10pt

1 INTRODUCTION (Article Design)

(Style: Arial Narrow, Bold, 10pt)

10pt

(Tab 6 mm) The article is written in Latin script and Greek symbols can be used for labelling. The length of the article is limited to eight pages of international paper size of Letter (in accordance with the template with all the tables and figures included). When formatting the text the syllabification option is not to be used.

10pt

1.1 Subtitle 1 (Writing Instructions)

(Style: Arial Narrow, 10pt, Bold, Align Left)

10pt

The document format is Letter with margins in accordance with the template. A two column layout is used with the column spacing of 10 mm. The running text is written in Times New Roman with single line spacing, font size 10 pt, alignment justified.

Article title must clearly reflect the issues covered by the article (it should not contain more than 15 words).

Body of the text is divided into chapters and the chapters are divided into subchapters, if needed. Chapters are numbered with Arabic numerals (followed by a period). Subchapters, as a part of a chapter, are marked with two Arabic numerals i.e. 1.1, 1.2, 1.3, etc. Subchapters can be divided into even smaller units that are marked with three Arabic numerals i.e. 1.1.1, 1.1.2, etc. Further divisions are not to be made.

Titles of chapters are written in capital letters (uppercase) and are aligned in the centre. The titles of subchapters (and smaller units) are written in small letters (lowercase) and are aligned left. If the text in the title of the subchapter is longer than one line, no hanging indents.

10pt

Typographical symbols (bullets), which are being used for marking an item in a list or for enumeration, are placed at a beginning of a line. There is a spacing of 10pt following the last item:

- Item 1
- Item 2
- Item 3

10pt

The same rule is valid when items are numbered in a list:

1) Item 1
2) Item 2
3) Item 3

10pt

1.2 Formatting of Pictures, Tables and Equations

(Style: Arial Narrow, 10pt, Bold, Align Left)

10pt

Figures (drawings, diagrams, photographs) that are part of the content are embedded into the article and aligned in the centre. In order for the figure to always be in the same position in relation to the text, the following settings should be defined when importing it: text wrapping / in line with text.

Pictures must be formatted for graphic reproduction with minimal resolution of 300 dpi. Pictures downloaded from the internet in ratio 1:1 are not suitable for print reproduction because of unsatisfying quality.

10pt

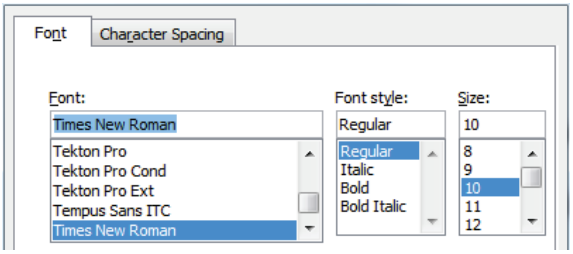


Figure 1 Text under the figure [1]
(Style: Arial Narraw, 8pt, Align Centre)

10pt

The journal is printed in black ink and the figures have to be prepared accordingly so that bright tones are printed in a satisfactory manner and are readable. Figures are to be in colour for the purpose of digital format publishing. Figures in the article are numbered with Arabic numerals (followed by a period).

Text and other data in tables are formatted - Times New Roman, 8pt, Normal, Align Center.

When describing figures and tables, physical units and their factors are written in italics with Latin or Greek letters,

while the measuring values and numbers are written upright.

10pt

Table 1 Table title aligned centre

(Style: Arial Narraw, 8pt, Align Centre)

	1	2	3	4	5	6
ABC	ab	ab	ab	ab	ab	ab
DEF	cd	cd	cd	cd	cd	cd
GHI	ef	ef	ef	ef	ef	ef

10 pt

Equations in the text are numbered with Arabic numerals inside the round brackets on the right side of the text. Inside the text they are referred to with equation number inside the round brackets i.e. ".... from Eq. (5) follows ...." (Create equations with MathType Equation Editor - some examples are given below).

10pt

$$F_{\text{avg}}(t, t_0) = \frac{1}{t}\int_{t_0}^{t_0+t} F\left[q(\tau), p(\tau)\right] \mathrm{d}\tau, \tag{1}$$

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cdot\cos\frac{\alpha-\beta}{2}, \tag{2}$$

$$(\boldsymbol{AB})^{\mathrm{T}} = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}. \tag{3}$$

10pt

Variables that are used in equations and also in the text or tables of the article are formatted as *italics* in the same font size as the text.

10pt

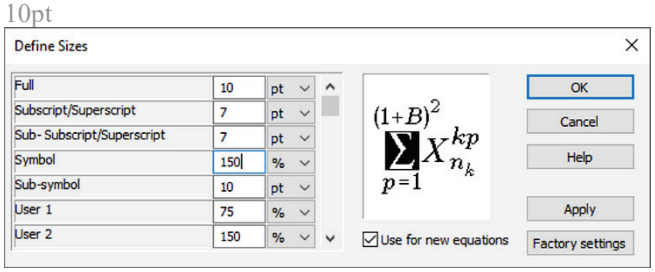


Figure 2 The texts under figures

(Style: Arial Narraw, 8pt, Align Centre)

10pt

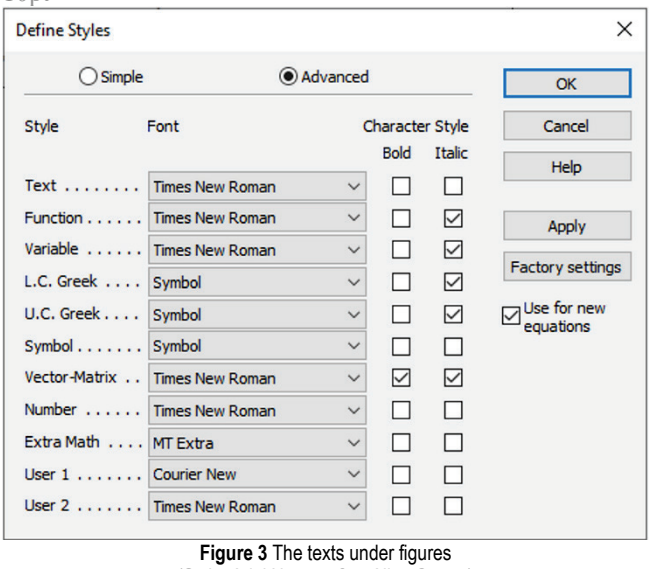


Figure 3 The texts under figures

(Style: Arial Narraw, 8pt, Align Centre)

Figures and tables that are a part of the article have to be mentioned inside the text and thus connected to the content i.e. " … as shown in Fig. 1…" or "data from Tab. 1…" and similar.

10pt

2 PRELIMINARY ANNOTATION

10pt

Article that is offered for publication cannot be published beforehand, be it in the same or similar form, and it cannot be offered at the same time to a different journal. Author or authors are solely responsible for the content of the article and the authenticity of information and statements written in the article.

Articles that are accepted for publishing are classified into four categories: original scientific papers, preliminary communications, subject reviews and professional papers.

Original scientific papers are articles that according to the reviewer and the editorial board contain original theoretical or practical results of research. These articles need to be written in such a way that based on the information given, the experiment can be repeated and the results described can be achieved together with the author's observations, theoretical statements or measurements.

Preliminary communication contains one or more pieces of new scientific information, but without details that allow recollection as in original scientific papers. Preliminary communication can give results of an experimental research, results of a shorter research or research in progress that is deemed useful for publishing.

Subject review contains a complete depiction of conditions and tendencies of a specific domain of theory, technology or application. Articles in this category have an overview character with a critical review and evaluation. Cited literature must be complete enough to allow a good insight and comprehension of the depicted domain.

Professional paper can contain a description of an original solution to a device, assembly or instrument, depiction of important practical solutions, and similar. The article need not be related to the original research, but it should contains a contribution to an application of known scientific results and their adaptation to practical needs, so it presents a contribution to spreading knowledge, etc.

Outside the mentioned categorization, the Editorial board of the journal will publish articles of interesting content in a special column. These articles provide descriptions of practical implementation and solutions from the area of production, experiences from device application, and similar.

10pt

3 WRITING AN ARTICLE

10pt

Article is written in the English language and the terminology and the measurement system should be adjusted to legal regulations, standards and the International System of Units (SI) (Quantities and Units: ISO 80 000 - from Part 1 to Part 14). The article should be written in third person.

Introduction contains the depiction of the problem and an account of important results that come from the articles that are listed in the cited literature.

Main section of the article can be divided into several parts or chapters. Mathematical statements that obstruct the reading of the article should be avoided. Mathematical statements that cannot be avoided can be written as one or more addendums, when needed. It is recommended to use an example when an experiment procedure, the use of the work in a concrete situation or an algorithm of the suggested method must be illustrated. In general, an analysis should be experimentally confirmed.

Conclusion is a part of the article where the results are being given and efficiency of the procedure used is emphasized. Possible procedure and domain constraints where the obtained results can be applied should be emphasized.

10pt

4 RECAPITULATION ANNOTATION

10pt

In order for the articles to be formatted in the same manner as in this template, this document is recommended for use when writing the article. Finished articles written in MS Word for Windows and formatted according to this template must be submitted using our The Paper Submission Tool (PST) (https://tehnickiglasnik.unin.hr/authors.php) or eventually sent to the Editorial board of the Technical Journal to the following e-mail address: tehnickiglasnik@unin.hr

The editorial board reserves the right to minor redaction corrections of the article within the framework of prepress procedures. Articles that in any way do not follow these authors' instructions will be returned to the author by the editorial board. Should any questions arise, the editorial board contacts only the first author and accepts only the reflections given by the first author.

10pt

5 REFERENCES (According to APA)

10pt

The literature is cited in the order it is used in the article. No more than 35 references are recommended. Individual references from the listed literature inside the text are addressed with the corresponding number inside square brackets i.e. "… in [7] is shown …". If the literature references are web links, the hyperlink is to be removed as shown with the reference number 8. Also, the hyperlinks from the e-mail addresses of the authors are to be removed. In the literature list, each unit is marked with a number and listed according to the following examples (omit the subtitles over the references – they are here only to show possible types of references):

9pt

[1] See http://www.bibme.org/citation-guide/apa/

[2] See http://sites.umuc.edu/library/libhow/apa_examples.cfm

[3] (Style: Times New Roman, 9pt, according to APA)

[4] Amidzic, O., Riehle, H. J., & Elbert, T. (2006). Toward a psychophysiology of expertise: Focal magnetic gamma bursts as a signature of memory chunks and the aptitude of chess players. *Journal of Psychophysiology, 20*(4), 253-258. https://doi.org/10.1027/0269-8803.20.4.253

[5] Reitzes, D. C. & Mutran, E. J. (2004). The transition to retirement: Stages and factors that influence retirement adjustment. *International Journal of Aging and Human Development, 59*(1), 63-84. Retrieved from http://www.baywood.com/journals/PreviewJournals.asp?Id=0091-4150

[6] Jans, N. (1993). *The last light breaking: Life among Alaska's Inupiat Eskimos*. Anchorage, AK: Alaska Northwest Books.

[7] Miller, J. & Smith, T. (Eds.). (1996). *Cape Cod stories: Tales from Cape Cod, Nantucket, and Martha's Vineyard*. San Francisco, CA: Chronicle Books.

[8] Chaffe-Stengel, P. & Stengel, D. (2012). *Working with sample data: Exploration and inference*. https://doi.org/10.4128/9781606492147

[9] Freitas, N. (2015, January 6). People around the world are voluntarily submitting to China's Great Firewall. Why? Retrieved from http://www.slate.com/blogs/future_tense/2015/01/06/tencent_s_wechat_worldwide_internet_users_are_voluntarily_submitting_to.html (Style: Times New Roman, 9pt, according to APA)

10pt

10pt

Authors' contacts:

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

8pt

Full Name, title
Institution, company
Address
Tel./Fax, e-mail

Note: Gray text should be removed in the final version of the article because it is for guidance only.


BEM/MRM
46
WIT CONFERENCES
Call for Papers
46th International Conference on Boundary Elements and other Mesh Reduction Methods
3-5 October 2023 | Lisbon, Portugal
Organised by
Wessex Institute, UK
Embry-Riddle Aeronautical University, USA
University of Central Florida, USA
University of Mississippi, USA
Sponsored by
WIT Transactions on Engineering Sciences
www.witconferences.com/bem46

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 17 – BROJ / ISSUE 3

RUJAN 2023 / SEPTEMBER 2023 – STRANICA / PAGES 305-477

Sveučilište
Sjever

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

ISSN 1846-6168 (PRINT) / ISSN 1848-5588 (ONLINE)

TEHNICKIGLASNIK@UNIN.HR – HTTP://TEHNICKIGLASNIK.UNIN.HR