TEHNIČKI GLASNIK

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 17 – BROJ / ISSUE 2

LIPANJ 2023 / JUNE 2023 – STRANICA / PAGES 153-303

/

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 2
Varaždin, June 2023 Pages 153-303

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):
May 5, 2023	May 10, 2023	June 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

ECCOMAS
European Community on Computational
Methods in Applied Sciences
www.eccomas.org

ECCOMAS Thematic Conference

ICCCM 2023

7th International Conference on Computational Contact Mechanics

July 5-7, 2023, Torino - Italy

ORGANIZERS

G. Zavarise Politecnico di Torino, Italy
A. Popp Universität der Bundeswehr München, Germany

CONFERENCE SECRETARIAT

e-mail: icccm2023@polito.it
phone: +39 0832 297414
web: ICCCM2023.polito.it

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20211119133806

Original scientific paper
Received: 2021-11-19, Accepted: 2022-08-26

Development of Early Stage Diabetes Prediction Model Based on Stacking Approach

Ilkay Cinar, Yavuz Selim Taspinar, Murat Koklu*

Abstract: Diabetes is a disease that may pose direct or indirect risks in terms of human health. Early diagnosis can minimize the potential harm of this disease to the body and reduce the probability of death. For this reason, laboratory tests are performed on diabetic patients. The analysis of these tests enables the diagnosis of diabetes. The aim of this study is so quickly diagnose diabetes by using data obtained from patients with machine learning methods. In order to diagnose the disease, k-nearest neighbor (k-NN), logistic regression (LR), random forest (RF) models and the stacking meta model which is created by combining these three models were used. The dataset used in the research includes test samples taken from 520 people. The dataset has 17 features, including 16 input features and 1 output feature. As a result of the classification through this dataset, different classification results were obtained from the models. The classification success of the models LR, k-NN, RF and stacking were found to be 91.3%, 91.7%, 97.9% and 99.6%, respectively. F-score, precision and recall performance metrics were utilized for a detailed analysis of the models' classification results. The obtained results revealed that the stacking model has a sufficient level to be used as a decision support system in the early diagnosis of diabetes.

Keywords: decision support system; diabetes; early stage; machine learning; stacking

1 INTRODUCTION

Diabetes is a disease in which insulin cannot be produced by the pancreas, or insulin sufficiently-produced in the pancreas cannot be used by the human body. Insulin, a type of hormone produced by the pancreas, plays a key role in transferring glucose from consumed nutrients to blood cells in the body and then converting it into the energy. When the body is unable to produce insulin, the level of glucose in the blood increase. High levels of glucose in the blood, on the other hand, can be detrimental to the viscera and lead to dysfunction in the tissue.

Diabetes is generally treated in three subheadings as type 1, type 2, and gestational diabetes. Gestational diabetes is a type of diabetes that occurs during pregnancy only because of the hormonal changes. Common symptoms of diabetes mellitus are polyuria, polydipsia, polyphagia, sudden weight loss, being underweight, obesity, pruritus, delayed recovery, blurred vision, genital thrush, nervousness, muscle stiffness, and etc. [3-5]. Early diagnosis of diabetes is essential for taking preventive measures. Besides, effective treatment at the first stage of the disease will always have additional benefits for patients [6].

Diagnosing diabetes through medical testing may not provide confident results due to the clinical complexity, time-consuming process, and high expenses. On the other hand, thanks to the machine learning algorithms, a disease such as diabetes can be predicted in a short time with lower costs [7]. Machine learning, a sub-branch of artificial intelligence (AI), relates to the development of algorithms and techniques that enable computers to learn based on the past experiences. In other words, the system can define and understand the input data and accordingly make decisions, predictions, and classifications [5, 8].

The contributions of this article can be summarized as follows:

- Studies have been carried out for the early diagnosis of diabetes by using the Early Stage Diabetes Risk Prediction dataset within The University of California, Irvine (UCI) repository of machine learning databases.
- For the early diagnosis of diabetes, transfer learning has been applied by utilizing deep learning architectures VGG16 and VGG19.
- The results obtained through transfer learning have been shared.
- The results obtained based on the literature studies carried out using the Early Stage Diabetes Risk Prediction dataset have been compared.

The remaining parts of the article have been organized as follows: The second chapter includes previous studies that focus on diabetes disease prediction via using machine learning algorithms and have significance in terms of the literature. The third chapter covers the description of the dataset, the research methods, and the explanations on performance metrics. The fourth chapter includes experimental results. And lastly, in the fifth chapter, the results and discussion are presented.

2 RELATED WORKS

Kandhasamy and Balamurali [3] compared the performances J48 decision tree (DT), k-nearest neighbors (K-NN), random forest (RF) and support vector machines (SVM) algorithms in order to classify diabetic patients. The results of the study indicated that the random forest algorithm has the higher classification accuracy compared to the other algorithms.

Perveen et al. [9] conducted a study on the prediction of the disease by using diabetes risk factors. The experimental results of the study in which J48 decision tree, Adaboost and Bagging algorithms were utilized to perform classification operations showed that Adaboost algorithm provides better results than J48 decision tree and Bagging algorithms.

In the study conducted by Husain and Khan [10], the distinctive performances of the ensemble learning model were investigated, for prediction of diabetes at an early stage. An ensemble model has been developed by combining these algorithms to improve the overall prediction accuracy by using different machine learning algorithms. 0.75 AUC and

an accuracy of 96% has been ensured by the model developed after the classification process.

Sisodia and Sisodia [2] aimed at developing a model which can predict the probability of diabetes in patients, with the maximum accuracy. Within the scope of their study, three machine learning classification algorithms were used: DT, SVM and Naive Bayes (NB). As a conclusion, it was figured out that the Naive Bayes has a better performance compared to other algorithms with an accuracy rate of 76.30%.

Alehegn et al. [11] proposed an ensemble method for predicting diabetes. The results obtained with the proposed method were compared with the results of the most common machine learning methods. The proposed method was found to have the highest classification accuracy with a rate of 90.36%.

In their studies, Choudhury and Gupta [12] aimed to detect diabetes by using machine learning techniques such as RF, NB, K-NN, SVM, LR, DT and ANN. The results of the study revealed that the highest classification accuracy (77.61%) among all algorithms belongs to LR.

Alam et al. [13] conducted studies to distinguish the most important qualities for the predictions of diabetes disease and make classification of the disease. In order to determine important qualities, principal component analysis (PCA) method was used in the study. Moreover, artificial neural network (ANN), RF, and K-means clustering methods were utilized for the classification processes. As a result of the classification, it is determined that ANN has the highest accuracy value with a rate of 75.7%.

For the diagnosis of diabetes, Challa and Chinnaiyan [14] used the classification algorithms of DT, SVM, K-NN and RF within the scope of their studies. At a rate of 78.25%, the highest classification accuracy was obtained with the DT.

Rajni and Amandeep [4], by using RB-Bayes, proposed a model to determine whether the person has diabetes or not. Furthermore, they performed classification through SVM, NB and K-NN algorithms and compared with the model they proposed. The results of the classification processes showed that the highest classification accuracy belongs to the proposed RB-Bayes model with the rate of 72.9%.

Kowsher et al. [5], in order to detect early diabetic patients, utilized deep neural networks (ANN) together with seven machine learning algorithms and compared their results. Deep ANN has achieved the highest classification accuracy in detecting diabetic patients. As a result of the classification, an accuracy of 95.14% was obtained.

In their studies, Ayon and Islam [15] used deep ANN in order to effectively detect diabetic patients. They compared the results using 5-fold and 10-fold cross-validation during the training of the neural network. The classification accuracy obtained after 5-fold cross-validation was 98.35%, while the accuracy was found to be 97.11% after 10-fold cross-validation. Following the experimental results, it is figured out that the proposed system provides promising results with 5-fold cross-validation.

Le Minh et al. [16] proposed a model to predict the early onset of diabetes disease. They used Multi-Layer Perceptron (MLP) to reduce the number of input features, while they use Gary Wolf Optimizer (GWO) and Adaptive Particle Swarm Optimization (APSO) to optimize the number of input features. The proposed method provided 97% accuracy for APGWO-MLP.

Harz et al. [17] used artificial neural networks to predict whether a person has diabetes or not. They achieved a prediction accuracy of 98.73% as a result of the study.

Yucelbas and Yucelbas [18] aimed to determine which features effective in the early diagnosis of diabetes, according to gender. It is indicated that the weakness feature was not effective on 108 current research results and that the classification accuracy obtained before the selection of male subjects was found to be 97.86%, with 13 features.

In the study conducted by Ridwan [1], an accuracy of 90.20% and an AUC value of 0.95 were obtained by using the Naive Bayes (NB) classification method.

Hana [19], used neural network and linear discriminant analysis (LDA) algorithm to analyze diabetic patients. While an accuracy of 90.38% was achieved with the LDA algorithm, a classification accuracy of 95.19% was achieved with the neural network.

Kaur and Kumari [8], in the study they conducted, used the R data manipulation tool to develop trends and identify risk factors and patterns. SVM, Radial-Basis Function (RBF) Kernel Support Vector Machine, k-NN, ANN and Multi-factor Dimensionality Reduction (MDR) algorithms was used in order to classify patients as diabetic and non-diabetic. As a result, the highest classification accuracy was achieved in Linear Kernel SVM with 89%.

Abd Rahman et al. [20] aimed to develop a prediction model using three different machine learning algorithms to classify Type 2 diabetes mellitus (T2DM) of the Malaysian population. DT, SVM and NB were used as classification algorithms and as a result, in terms of accuracy (0.87), sensitivity (0.9), specificity (0.8), sensitivity (0.9), F1 score (0.9), and AUC value (0.93), the best overall prediction performance was achieved with the random forest algorithm.

Tripathi and Kumar [21] carried out a study to predict diabetes at an early stage, by utilizing LDA, K-NN, SVM and RF machine learning algorithms. It was observed that the RF algorithm, which achieved a maximum accuracy of 87.66% after the classification processes, performed better than the other algorithms used.

Naz and Ahuja [22] presented a methodology aimed at diabetes prediction using machine learning algorithms for the early diagnosis of diabetes. In the study, the classification processes was performed by using the NB, DT, ANN and Deep Learning (DT) algorithms. As a result, DL achieved the highest classification accuracy of 98.07%.

Hana [23] performed the classification process by the C 4.5 decision tree algorithm to detect diabetes. As a result, a classification accuracy of 93.02% was achieved.

3 MATERIAL AND METHODS

3.1 Database

Within the scope of this study, early-stage diabetes risk estimation dataset was used. This data set consists of a total of 520 people, 320 of whom are diabetic and 200 of whom are non-diabetic [24]. The dataset has a total of 17 features, including 16 input features and 1 output feature. In order for the data to be processed more easily, changes have been made in the values belonging to the features in a way that not

affect the classification result. The features and values the dataset includes are given in Tab. 1.

Table 1 Features and values of dataset

Features	Values	Conversion
Age	16 - 90	
Gender	Male, Female	Male=1, Female=0
Polyuria	Yes, No	Yes=1, No=0
Polydipsia	Yes, No	Yes=1, No=0
Sudden Weight Loss	Yes, No	Yes=1, No=0
Weakness	Yes, No	Yes=1, No=0
Polyphagia	Yes, No	Yes=1, No=0
Genital Thrush	Yes, No	Yes=1, No=0
Visual Blurring	Yes, No	Yes=1, No=0
Itching	Yes, No	Yes=1, No=0
Irritability	Yes, No	Yes=1, No=0
Delayed Healing	Yes, No	Yes=1, No=0
Partial Paresis	Yes, No	Yes=1, No=0
Muscle Stiffness	Yes, No	Yes=1, No=0
Alopecia	Yes, No	Yes=1, No=0
Obesity	Yes, No	Yes=1, No=0
Class	Positive, Negative	Positive=1, Negative=0

While the trainings were carried out, 16 features were given as input and 1 output feature was given as output to the models. Data pre-processing was not performed since there is no missing data in the dataset that would affect the classification success of the models.

3.2 k-Nearest Neighbor

k-NN is a machine learning method based on calculating the distances between data in the dataset [25]. The distance of an object to its neighbors is calculated according to a specified parameter which is called "*k*" and indicates the number of neighbors. Objects are divided into classes according to the specified number of neighbors. Different distance determination methods are used in determining the distance between objects [26]. The k value was determined as 5 in this study, while Euclidean distance was used to determine the distance between objects.

3.3 Logistic Regression

LR algorithm is a machine learning method that can be used in classification problems [27]. There is a linear relationship between the dependent and independent variables in linear regression, hence it is used in the solution of single input and single output problems. Due to this limitation of linear regression, logistic regression algorithm is used. In logistic regression, many independent variables are used to predict the dependent variable, which is the output variable. It is not necessary for the independent variables, i.e. input variables, to be evenly distributed [28].

3.4 Random Forest

RF algorithm is an ensemble machine learning method that contains many decision trees. Each decision tree it contains performs a query on objects randomly taken from the dataset, and the object is placed on a node. Results from each decision tree are voted. The most suitable class for the object is determined as a result of voting the estimates from the trees [29].

3.5 Stacking

Stacking is an ensemble machine learning method that can classify data with results from different classifiers and a training result within itself. The model emerged by the results from different models and the result of training within itself is called the meta-model. The meta-model is expected to provide more successful results than the classifiers that comprise it. However, sometimes, it may give lower results since the classification ability is impacted by many parameters [30]. In the study, the Stacking meta-model was created through from the results of k-NN, LR and RF methods and the result of the training within the model.

3.6 Confusion Matrix

Confusion matrix is a table which is used to see the classification numbers of data samples in the solution of classification problems with machine learning methods. The correctly and incorrectly classified data belong the classes can be reached by using the data in this table [31]. In the table, four data exist as true positive (*TP*), true negative (*TN*), false positive (*FP*) and false negative (*FN*). Accordingly,

- True Positive (*TP*) represents the true classified positive samples,
- True Negative (*TN*) represents the true classified negative samples,
- False Positive (*FP*) represents the false classified positive samples,
- False Negative (*FN*); represents the false negative samples [32].

Tab. 2 shows the placement of these values on the matrix.

Table 2 Confusion matrix

		PREDICTED	
		Negative	Positive
ACTUAL	Negative	*TN*	*FP*
	Positive	*FN*	*TP*

3.7 Performance Evaluation Metrics

For the detailed performance evaluation of the models trained with the data in the dataset, there are also different metrics other than the classification success [33]. F-score, precision, and recall metrics are the other metrics utilized for evaluation of the success of the model [34]. The F-score is a measurement metric which will include all error cost, not just

misclassified samples. The class variable is used to evaluate the performance of models in datasets that are not evenly distributed. It is calculated by taking the harmonic average of precision and recall values [35]. Precision is a metric used to see how many samples classified as true and false positives are actually positive. Recall, on the other hand, is the metric showing how many of the samples that should be predicted positively were classified as positive [36]. The four performance metrics used in the study are calculated by the formulas in Tab. 3.

Table 3 Performance metrics equations

Metrics	Equation
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
F-score	$2\times\frac{Precision\times Recall}{Precision+Recall}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$

These four metric values represent the success of the model. The higher the value, the higher the success of the model [37].

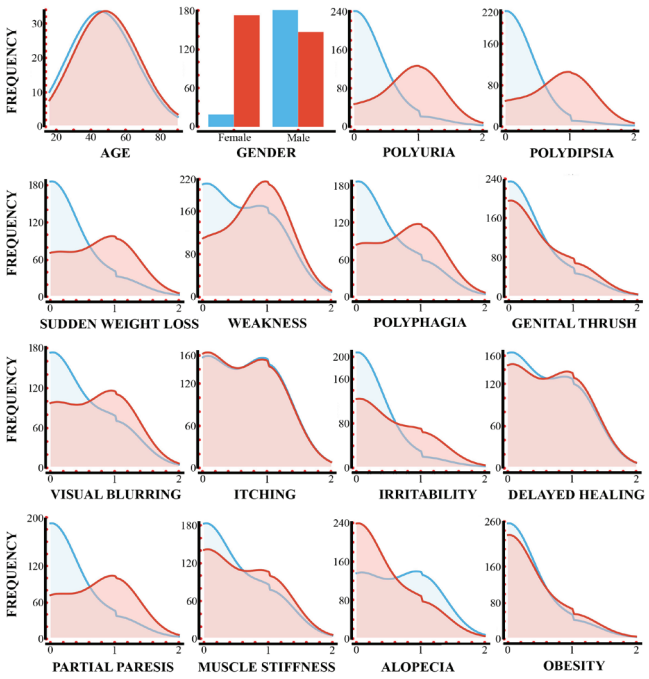


Figure 1 The data distribution charts

4 EXPERIMENTAL RESULTS

Thanks to the data distribution charts, preliminary information can be obtained about their classification success before the models are trained. The distribution ratio of class values also enables to make decision about which kind of testing procedures to be performed on models. The repetition status of each feature in the dataset, that is, the data distribution graphs created according to its frequency are shown in Fig. 1. The result value indicated in blue color indicates negative (0), and the result value indicated in red color indicates positive (1). The output demonstrated in blue color indicates that the value is negative (0), and the output demonstrated in red color indicates that the value positive (1).

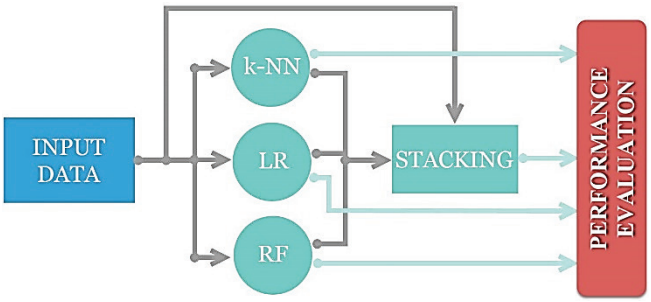
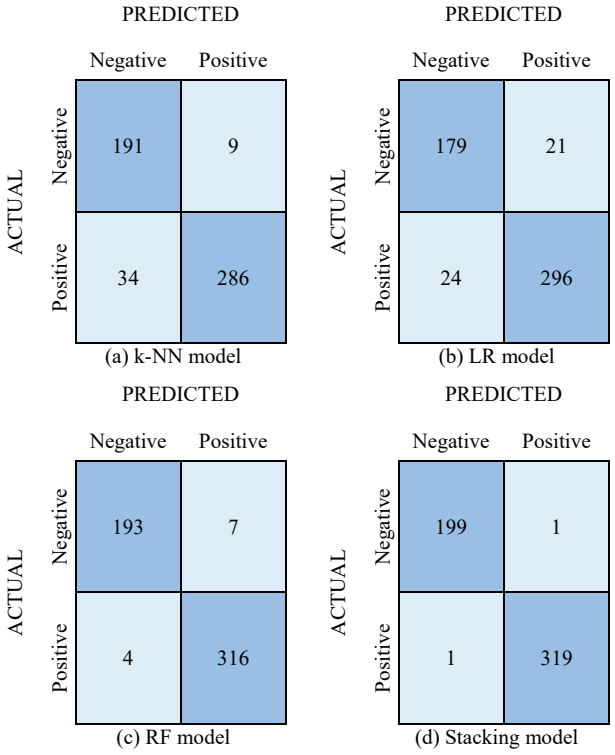


Figure 2 Flow chart of the performed processes

Output values are 200 negative (0), 320 positive (1). After analyzing the data distributions, the training of the models was carried out. The classification processes were performed with the k-NN, LR, RF models, and the Stacking model created by combining these models. The flow chart of the processes performed with the models is given in Fig. 2.

Table 4 Confusion matrix of all models



In order to make a comparison between the accuracies of different classification models, in the study, performance measurements were carried out on the dataset used. The cross-validation method was utilized to obtain a standard in classification and to get rid of the subjectivity of the classification methods performed with the train-test distinction. In cross-validation method, the dataset is divided

into *k* parts. Each part is utilized as a validation set. The remaining *k*–1 part is used for training the algorithm. The average of the success rates obtained as a result of these processes performed k times gives the classification success of the algorithm. In this way, classification success can be measured objectively. The value of k was determined as 10 in the training with a dataset containing 520 lines of data. Confusion Matrices of all methods used are given in order. The performances of the classifiers were compared by using confusion matrix values. Confusion Matrices obtained as a result of the classification with all models are shown in Tab. 4.

In Tab. 4 (a), with the k-NN model, 43 data in total were classified incorrectly. In Tab. 4(b), with the LR model, 45 data in total were classified incorrectly, while 475 data were classified correctly. In Tab. 4(c), 11 data were classified incorrectly and 509 data were classified correctly with the RF model. In Tab. 4(d), with the Stacking meta model based on k-NN, LR and RF models, 2 data were classified incorrectly, while 518 data were correctly classified. The results obtained as a result of the statistical calculations by the data of Confusion Matrix are included in Tab. 5.

Table 5 Performance metrics of models

	F-Score	Precision	Recall
k-NN	0.918	0.923	0.917
LR	0.914	0.914	0.913
RF	0.982	0.973	0.987
Stacking	**0.996**	**0.996**	**0.996**

Classification success rates of LR, k-NN, RF models and the Stacking meta model which is a combination of these models are shown in Tab. 6.

Table 6 Classification success rates of models

	LR	k-NN	RF	Stacking
Accuracy	91.3%	91.7%	97.9%	**99.6%**

According to Tab. 6, LR model has the lowest classification success with a rate of 91.3% while the highest classification success belongs to the Stacking Meta Model with 99.6%. The Stacking Model has the highest F-score value with 0.966, precision 0.996 and recall 0.996 values. The lowest F-score value, on the other hand, belongs to the LR Model with 0.914, precision 0.914, recall 0.913 values. Fig. 3 gives the success rates of the models.

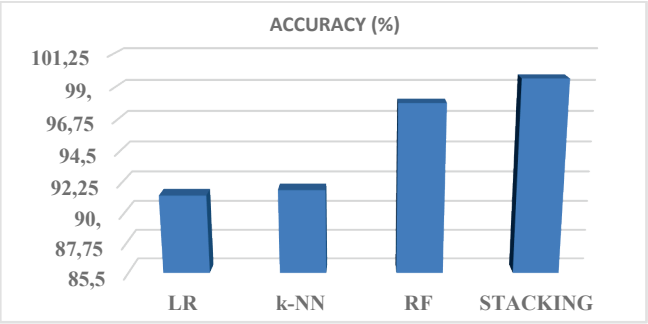


Figure 3 Accuracy of classification models

5 CONCLUSION

The changes in people's diet and lifestyle, may bring about an increase also in the diseases caused by diabetes besides the increase in diabetes disease in the society. Early diagnosis ensures to start treatment of diseases earlier and to halt the diseases' progression. With the classification models formed via using the early diagnosis diabetes dataset created for this purpose, it can be possible to detect diabetes at an early stage. Within the scope of this study, classification processes have been completed for the early diagnosis of diabetes. k-NN, LR, RF and the Stacking meta model created by combining these 3 models were used in classification. The classification successes obtained with these models are 91.7%, 91.3%, 97.9% and 99.4%, respectively. When the success rates are examined, it is understood that the Stacking Model has the highest classification success. The fact that the Stacking Model has a higher success compared to other methods is due to the models that make up this model classify the data correctly and incorrectly. In addition, models become able to classify the data more accurately when they are combined. It was also observed that the classification success of the Stacking Meta-Model, which was created using the models with different *FN* and *FP* classification numbers, is higher. A higher classification success was obtained compared to the results obtained in literature studies using the same data set. The comparison of the proposed model with the models in other studies is given in Tab. 7.

Table 7 Studies conducted by using the Early Stage Diabetes Risk Prediction dataset

Method	Accuracy (%)	References
NB	90.20	[1]
C4.5 DT	93.02	[23]
ANN	95.19	[19]
APGWO-MLP	97.00	[16]
k-NN	97.86	[18]
ANN	98.73	[17]
Proposed Stacking Model	**99.6%**	

The Stacking Model, which has the highest classification success, can be used as a decision support system in the early diagnosis of diabetes. Achieving 100% success in the field of health is always a desirable conclusion. It is thought that the success of classification can be increased via different machine learning approaches.

6 REFERENCES

[1] Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan), 4*(1), 15-21. https://doi.org/10.47970/siskom-kb.v4i1.169

[2] Sisodia, D. & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, 132, 1578-1585. https://doi.org/10.1016/j.procs.2018.05.122

[3] Kandhasamy, J. P. & Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, 47, 45-51. https://doi.org/10.1016/j.procs.2015.03.182

[4] Rajni, R. & Amandeep, A. (2019). RB-Bayes algorithm for the prediction of diabetic in Pima Indian dataset. *International Journal of Electrical Computer Engineering, 9*(6), 4866. https://doi.org/10.11591/ijece.v9i6.pp4866-4872

[5] Kowsher, M., Turaba, M. Y., Sajed, T., & Rahman, M. M. (2019, December). Prognosis and treatment prediction of type-2 diabetes using deep neural network and machine learning classifiers. *The 22nd International Conference on Computer and Information Technology (ICCIT),* 1-6, IEEE. https://doi.org/10.1109/ICCIT48885.2019.9038574

[6] Jain, D. & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal, 19*(3), 179-189. https://doi.org/10.1016/j.eij.2018.03.002

[7] Kahramanli, H. & Allahverdi, N. (2008). Design of a hybrid system for the diabetes and heart diseases. *Expert systems with applications, 35*(1-2), 82-89. https://doi.org/10.1016/j.eswa.2007.06.004

[8] Kaur, H. & Kumari, V. (2020). Predictive modelling and analytics for diabetes using a machine learning approach. *Applied computing and informatics.* https://doi.org/10.1016/j.aci.2018.12.004

[9] Perveen, S., Shahbaz, M., Guergachi, A., & Keshavjee, K. (2016). Performance analysis of data mining classification techniques to predict diabetes. *Procedia Computer Science*, 82, 115-121. https://doi.org/10.1016/j.procs.2016.04.016

[10] Husain, A. & Khan, M. H. (2018). Early diabetes prediction using voting based ensemble learning. In *International conference on advances in computing and data sciences*. 95-103, Springer. https://doi.org/10.1007/978-981-13-1810-8_10

[11] Alehegn, M., Joshi, R., & Mulay, P. (2018). Analysis and prediction of diabetes mellitus using machine learning algorithm. *International Journal of Pure Applied Mathematics, 118*(9), 871-878.

[12] Choudhury, A. & Gupta, D. (2019). A survey on medical diagnosis of diabetes using machine learning techniques, in *Recent developments in machine learning and data analytic,* 67-78, Springer. https://doi.org/10.1007/978-981-13-1280-9_6

[13] Alam, T. M., Iqbal, M. A., Ali, Y., Wahab, A., Ijaz, S., Baig, T. I. ... & Abbas, Z. (2019). A model for early prediction of diabetes. *Informatics in Medicine Unlocked*, 16, 100204. https://doi.org/10.1016/j.imu.2019.100204

[14] Challa, M. & Chinnaiyan, R. (2019, September). Optimized Machine Learning Approach for the Prediction of Diabetes-Mellitus. In *International Conference on Computational Vision and Bio Inspired Computing,* 321-328, Springer. https://doi.org/10.1007/978-3-030-37218-7_37

[15] Ayon, S. I. & Islam, M. (2019). Diabetes Prediction: A Deep Learning Approach. *International Journal of Information Engineering Electronic Business, 11*(2). https://doi.org/10.5815/ijieeb.2019.02.03

[16] Le, T. M., Vo, T. M., Pham, T. N., & Dao, S. V. T. (2020). A Novel Wrapper–Based Feature Selection for Early Diabetes Prediction Enhanced with a Metaheuristic. *IEEE Access*, 9, 7869-7884. https://doi.org/10.1109/ACCESS.2020.3047942

[17] Harz, H. H., Rafi, A. O., Hijazi, M. O., & Abu-Naser, S. S. (2020). Artificial Neural Network for Predicting Diabetes Using JNN. *International Journal of Academic Engineering Research (IJAER), 4*(10).

[18] Yucelbas, C. & Yucelbas, S. (2020). Determination of Effective Attributes in the Early Diagnosis of Gender-Based Diabetes. *Current Researches in Architecture Engineering Sciences,* 97.

[19] Hana, F. M. (2020). Perbandingan Algoritma Neural Network dengan linier discriminant analysis (LDA) pada klasifikasi penyakit diabetes. *Jurnal Bisnis Digital Dan Sistem Informasi*, *1*(1), 21-29.

[20] Abd Rahman, M. H. F., Salim, W. W. A. W., & Abd Wahab, M. F. (2020). Risk prediction analysis for classifying type 2 diabetes occurrence using local dataset. *Biological and Natural Resources Engineering Journal, 3*(1), 48-61.

[21] Tripathi, G. & Kumar, R. (2020, June). Early prediction of diabetes mellitus using machine learning. *In 2020 8th International Conference on Reliability, INFOCOM Technologies and Optimization (Trends and Future Directions) (ICRITO),* 1009-1014, IEEE. https://doi.org/10.1109/ICRITO48877.2020.9197832

[22] Naz, H. & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders, 19*(1), 391-403. https://doi.org/10.1007/s40200-020-00520-5

[23] Hana, F. M. (2020). Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4. 5. *Jurnal SISKOM-KB (Sistem Komputer dan Kecerdasan Buatan), 4*(1), 32-39. https://doi.org/10.47970/siskom-kb.v4i1.173

[24] Islam, M. F., Ferdousi, R., Rahman, S., & Bushra, H. Y. (2020). Likelihood prediction of diabetes at early stage using data mining techniques. *In Computer Vision and Machine Intelligence in Medical Image Analysis,* 113-125. Springer, Singapore. https://doi.org/10.1007/978-981-13-8798-2_12

[25] Tasdemir, S., Saritas, I., Ciniviz, M., & Allahverdi, N. (2011). Artificial neural network and fuzzy expert system comparison for prediction of performance and emission parameters on a gasoline engine. *Expert Systems with Applications, 38*(11), 13912-13923. https://doi.org/10.1016/j.eswa.2011.04.198

[26] Aslan, M. F., Durdu, A., & Sabanci, K. (2020). Human action recognition with bag of visual words using different machine learning methods and hyperparameter optimization. *Neural Computing and Applications, 32*(12), 8585-8597. https://doi.org/10.1007/s00521-019-04365-9

[27] Cruyff, M. J., Böckenholt, U., Van Der Heijden, P. G., & Frank, L. E. (2016). A review of regression procedures for randomized response data, including univariate and multivariate logistic regression, the proportional odds model and item response model, and self-protective responses. *Handbook of Statistics,* 34, 287-315. https://doi.org/10.1016/bs.host.2016.01.016

[28] Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A., & Mansor, S. (2018). Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN). *Geomatics, Natural Hazards and Risk, 9*(1), 49-69. https://doi.org/10.1080/19475705.2017.1407368

[29] Speiser, J. L., Miller, M. E., Tooze, J., & Ip, E. (2019). A comparison of random forest variable selection methods for classification prediction modeling. *Expert systems with applications,* 134, 93-101. https://doi.org/10.1016/j.eswa.2019.05.028

[30] Ghalejoogh, G. S., Kordy, H. M., & Ebrahimi, F. (2020). A hierarchical structure based on stacking approach for skin lesion classification. *Expert Systems with Applications,* 145, 113127. https://doi.org/10.1016/j.eswa.2019.113127

[31] Ozkan, I. A. (2020). A Novel Basketball Result Prediction Model Using a Concurrent Neuro-Fuzzy System. *Applied Artificial Intelligence, 34*(13), 1038-1054. https://doi.org/10.1080/08839514.2020.1804229

[32] 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

[33] Kocer, S. & Tumer, A. E. (2017). Classifying neuromuscular diseases using artificial neural networks with applied Autoregressive and Cepstral analysis. *Neural Computing and Applications, 28*(1), 945-952. https://doi.org/10.1007/s00521-016-2383-8

[34] Saura, J. R. (2021). Using data sciences in digital marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge, 6*(2), 92-102. https://doi.org/10.1016/j.jik.2020.08.001

[35] Chicco, D. & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics, 21*(1), 1-13. https://doi.org/10.1186/s12864-019-6413-7

[36] Cinarer, G., Emiroglu, B. G., & Yurttakal, A. H. (2020). Prediction of Glioma Grades Using Deep Learning with Wavelet Radiomic Features. *Applied Sciences, 10*(18), 6296. https://doi.org/10.3390/app10186296

[37] Saritas, I. (2012). Prediction of breast cancer using artificial neural networks. *Journal of Medical Systems, 36*(5), 2901-2907. https://doi.org/10.1007/s10916-011-9768-0

Authors' contacts:

Ilkay Cinar
Department of Computer Engineering, Selcuk University,
Alaaddin Keykubat Campus, 42075 Konya, Turkey
ilkay.cinar@selcuk.edu.tr
ORCID: 0000-0003-0611-3316

Yavuz Selim Taspinar
Doganhisar Vocational School, Selcuk University
Alaaddin Keykubat Campus, 42075 Konya, Turkey
ytaspinar@selcuk.edu.tr
ORCID: 0000-0002-7278-4241

Murat Koklu
(Corresponding author)
Department of Computer Engineering, Selcuk University
Alaaddin Keykubat Campus, 42075 Konya, Turkey
Tel: +90 332 223 33 48, mkoklu@selcuk.edu.tr
ORCID: 0000-0002-2737-2360

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20211214151555

Original scientific paper
Received: 2021-12-14, Accepted: 2022-09-01

Exergy Analysis of Thermal Power Plant for Three Different Loads

Nurdin Ćehajić

Abstract: This paper presents the energy and exergy analysis of thermal power plant Tuzla in Tuzla, Bosnia and Herzegovina. The main aim of this paper is to analyze the components of a 200 MW steam power plant unit in order to identify and quantify the sites with the highest exergy losses and to calculate exergy efficiency values of all components when operating at nominal load. The influence of the change in ambient temperature and block load on the value of exergy losses and exergy efficiency was taken into analysis. The analysis further includes the impact of steam block operation without high-pressure and low-pressure heaters on the exergy efficiency of the steam block. The goal of the analysis is to determine the functional state of individual steam block components after a long period of exploitation and maintenance in order to take appropriate measures to improve their technical performance. Exergy losses during nominal operation of the steam power plant unit are the largest in boiler and amount to 313.42 MW, followed by a turbine with 205.60 MW, condenser 1 with 6.03 MW, condenser 2 with 5.75 MW, while other components of the steam power plant have exergy losses in the range of 0.03 to 2.15 MW. Operation of the unit at nominal load without HPH results in an exergy efficiency decrease from 5.60 to 9.80 %, while in case of operation without HPH and LPH it results in a decrease in exergy efficiency from 9.86 to 16.40 % depending on the pattern used to calculate. The conclusion after the analysis indicates that the biggest exergy losses are in the boiler and turbine and consequently these components have the lowest exergy efficiency values. The increase in ambient temperature has different effects on individual components of the thermal power plant, increasing exergy losses of the boiler while reducing the turbine exergy losses and condensers.

Keywords: dead state; exergy analysis; exergy efficiency; exergy losses; steam power plant

1 INTRODUCTION

Energy and exergy analysis of power generation systems are essential for the efficient utilization of energy resources. Therefore these analyses became interesting for researches and scientists in recent years. The most commonly - used method for analyzing energy conversion process is the first law of thermodynamics. However, a method combining the first and second law of thermodynamics has been increasingly used recently. This method is used to calculate exergy and exergy losses in order to determine the efficiency of use of available energy. Exergy analysis enables defining the difference between energy losses to the environment and the internal irreversibility of the process [1].

Exergy analysis evaluates the performance of system and process components, as well as the evaluation of exergy at individual points of the energy transformation process. Based on the obtained data, it is possible to assess efficiency and determine the places in the process with the greatest losses. [2]. It is for these reasons that today's approach to process analysis includes exergy analysis, which provides a more realistic view of the process and is a useful tool for engineering evaluation [3]. It enables a better assessment of the efficiency of the complete system, better optimization, designing and improving the performance of energy systems.

A large number of researchers have sought to understand and improve the operation of thermal power plants, steam turbines and advanced cycles, using the method of energy and exergy analysis. Exergy analysis of energy systems in general and thermal power plants was dealt with by Aljundi et al. [4]. Yang, et al. [5] investigated 660 MW ultra-supercritical steam power plant in China who have shown that, heavier exergy destruction is caused by exhaust flue gases with 73.51% of the total boiler subsystem. The exergy analysis of various thermal power plants led to the conclusion that the boiler is the main source of exergy losses [6-12]. Many researchers have linked exergy to the cost analysis of the thermal power plants [13]. Gogoi and Talukdar [14] analyzed how the pressure in the boiler and the fuel flow rate affect the parameters of the boiler, and found a significant influence of these two parameters on the performance of the energy cycle. Kanoglu, et al. [15] have analyzed and evaluated different efficiencies of energy conversion and heat transfer taking into account energy systems with constant flow (turbines, compressors, pumps, heat exchangers, etc.), various power plants, cogeneration plants and refrigeration systems. Rashad and Maihy [16] analyzed the exergy and energy of the Shobra El - Khima power plant in Cairo and found that the highest exergy destruction occurred in the turbine (about 28% at different loads), while the highest energy loss was recorded in the condenser (55% at different loads). Sengupata, et al. [17] analyzed the exergy of a supercritical coal-fired steam power plant with a capacity of 210 MW at the design values of the parameters and at different loads. Živić, Galović, Avsec and Holik [18] they analyzed four variables at the inlet to the turbine, namely: the ratio of gas inlet temperature to the turbine, the ratio of compressor outlet and inlet pressure and inlet air temperature to the compressor, and the isentropic efficiency of the compressor and turbine. The air temperature at the entrance to the turbine was kept constant, while the temperature of the flue gases at the entrance to the turbine varied from 900 to 1200 °C.

The aim of this paper is to analyze the 200 MW unit of thermal power plant in Tuzla from the perspective of energy and exergy. The primary task is the exergy analysis of thermal power plant components at nominal operating mode, as well as the impact of exergy losses and thermal power plant operation without high - pressure and low - pressure heaters on exergy efficiency.

For the operating modes at 90 % and 80 % the load, the exergy efficiencies will be calculated and a comparative analysis will be performed. Also, the influence of the outside

temperature on exergy losses of boiler, turbine and both steam condensers will be analyzed.

2 PLANT DESCRIPTION

After the completion of construction, Tuzla thermal power plant 200 MW unit was for the first time synchronized with the grid in 1974 and a test facility has started that day. Prior to modernization, the unit had 153668 operating hours and 24267303 MWh of electricity submitted to the electricity grid. In the period from 2006 to 2008, the unit was revitalized by installing a new DCS control system, replacing electrostatic precipitators, coal mills, slag and ash transport systems, reconstructing boiler, installing electro - hydraulic turbine control and a new generator sealing system.

Tuzla thermal power plant 200 MW unit has a single-axle, three - cylinder, condensing turbine installed with two steam outputs and one intermediate heating. Each steam outlet from the turbine is connected to a special condenser. Inter - heating is performed between high - pressure and medium - pressure parts of the turbine.

The high - pressure section consists of 12 stages, the medium-pressure section of 11 stages, while the low - pressure part which is divided into two parts has 4 stages of rotor blades. The turbine is equipped with 7 uncontrolled extraction points used to preheat feedwater before it enters the boiler. The above mentioned 200 MW unit has 4 low pressure and 3 high pressure regenerative system heaters [19].

Extraction points are located at different turbine stages is as follows:

- I extraction point for HPH 7 - beyond 9th stage
- II extraction point for HPH 6 - beyond 12th stage (which is also the output from the high pressure section to the intermediate heating)
- III extraction point for HPH 5 - beyond 15th grade
- IV extraction point for LPH 4 - beyond 8th grade
- V extraction point for LPH 3 - beyond 21st degree
- VI extraction point for LPH 2 - beyond 23rd grade
- VII extraction point for LPH 1 - beyond 25th grade.

The data used for the thermodynamic analysis of the 200 MW unit are based on normative tests from 2014 at the state of the unit of 202 000 operating hours, with the data that the unit operated 6000 hours after the overhaul. The tests were performed for the operation at 100 % unit load (200 MW power) and steam production 600 t/h, 90 % unit load (180 MW power) and steam production 540 t/h, and 80 % unit load (160 MW power) with steam production 480 t/h. Boiler heating surfaces were cleaned.

Numerical analyses (energy and exergy analyses) performed in this paper do not require knowledge of the steam turbine or any other steam system component's internal structure [20-22]. The diagram of the 200 MW steam unit is shown in Fig. 1.

The operating conditions of the power plant are summarized in Tab. 1.

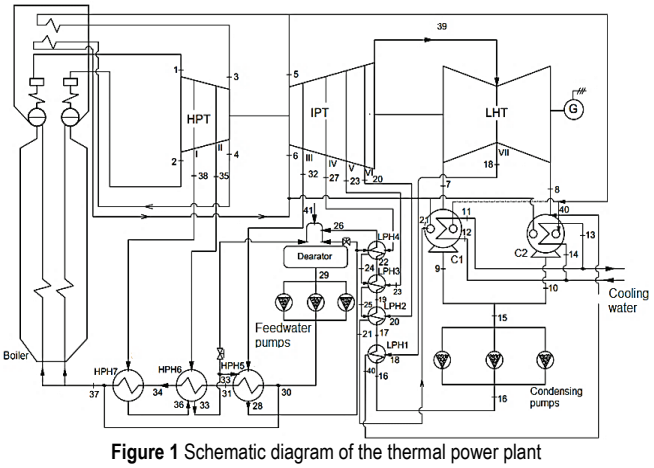


Figure 1 Schematic diagram of the thermal power plant

Table 1 Operating conditions of the thermal power plant

Operating condition	Value
Fuel mass flow rate	219.40 t/h
Lower heating value of fuel	8347.10 kJ/kg
Inlet gas volumetric flow rate to burners	577141 Nm^3/h
Feedwater inlet temperature	241.92 °C
Steam flow rate	608.50 t/h
Steam temperature	534.60 °C
Steam pressure	125.25 bar
Power output	195.99 MW
Number of turbine steam extraction points	7
Cooling water mass flow rate	28000 t/h
Isentropic efficiency of pumps	65 %
Boiler energy efficiency	88.24 %

3 THERMODYNAMIC ANSLYSIS

Exergy is the ability of a system to perform useful work when moving to a final state in equilibrium with the environment. In general, exergy is not conserved as energy, but destroyed in the system. Exergy destruction is a measure of irreversibility and is a source of performance loss. Through exergy analysis, it is possible to estimate the value of exergy losses, as well as the size and source of thermodynamic inefficiency of the heating system.

Mass, energy and exergy balances for any control volume at steady state, with negligible potential and kinetic energy changes, can be expressed, respectively, by

$$\sum m_i = \sum m_e \tag{1}$$

$$Q - W = \sum m_e h_e - \sum m_i h_i \tag{2}$$

$$E_{\text{heat}} - W = \sum m_e e_{xe} - \sum m_i e_{xi} \tag{3}$$

where the net exergy transfer by heat (E_{heat}) at temperature T is given by

$$E_{\text{heat}} = \sum \left(1 - \frac{T_o}{T}\right) \cdot Q \tag{4}$$

and the specific exergy is given by

$$e_x = h - h_0 - T_o(s - s_o) \tag{5}$$

Total exergy was calculated according to the formula

$$E_x = m\left[h - h_0 - T_o(s - s_o)\right] \tag{6}$$

where E_x, T, m, h and s indicate the total exergy rate, temperature, mass flow rate, enthalpy and entropy, respectively. The subscript 0 shows the dead state condition. Plant exergy efficiency can be defined as [23]:

$$\eta_{xe\ \text{plant1}} = \frac{E_{x\ \text{net},\,e}}{E_{xi}} \tag{7}$$

where $E_{x\ \text{net},\,e}$ and E_{xi} are net exergy at the output and exergy at the input, in order and are calculated:

$$E_{x\ \text{net},\,e} = W_T - W_{\text{own consum}} \tag{8}$$

$$E_{xi} = E_{x1} + E_{x2} + E_{x5} + E_{x6} - E_{x3} - E_{x37} - E_{x4} \tag{9}$$

where W_T denotes turbine power and $W_{\text{own consum}}$ refers to the auxiliary devices consuming 10 % of net power generation. E_x represents the exergy rate and subscripts indicate state points in Fig. 1. Exergy losses during coal combustion in the boiler and exergy losses related to exhaust gases were neglected in this analysis. On the other hand, plant exergy efficiency can be defined as:

$$\eta_{xe\ \text{plant2}} = \frac{E_{x\ \text{net},\,e}}{m_{\text{fuel}} \cdot e_{\text{fuel}}} \tag{10}$$

This definition takes into account the irreversibility of the heat transfer from gases to water in boiler pipe systems.

In Eq. (10) m_{fuel} stands for fuel mass flow rate and e_{fuel} is specific fuel exergy that can be expressed as:

$$e_{\text{fuel}} = \varphi \cdot LHV \tag{11}$$

where $\varphi = 1.05$ is exergy factor and LHV is fuel lower heating value [23]. The above forms are used for the analysis of the steam block and the ambient temperature is 293.15 K and the pressure is 101.3 kPa. The thermodynamic properties of the working fluid at the state points from Fig. 1 were calculated REFROP 8 software [24] and summarized in Tab. 2.

Thermodynamic properties of the working fluid and exergy values in the state points from Fig. 1 for operation of thermal power plant with 100%, 90 % and 80 % load were calculated and summarized in Tab. 2. The values of the parameters in the state points next to which the load is not specified are valid for the nominal load.

Values of LHV and mass flows of coal used for thermodynamic analysis are presented in Tab. 3.

For work in stationary mode and by choosing each component from Fig. 1 as control volume, exergy losses and exergy efficiencies can be calculated in the manner shown in Tab. 4.

Table 2 Thermodynamic properties, energy and exergy flow rates of state points in Fig. 1

State point /load		ϑ (°C)	p (bar)	m (t/h)	h (kJ/kg)	s (kJ/kgK)	E_x (MW)
1	100%	534.77	125.36	304.82	3434.64	6.57661	127.46
	90%	534.96	125.35	285.32	3435.20	6.58347	119.18
	80%	535.02	125.23	248.90	3435.73	6.5832	104.02
2	100%	534.58	125.14	303.65	3433.77	6.56920	127.08
	90%	534.20	125,52	279,61	3434.67	6.57324	117.00
	80%	534.83	125.14	253.18	3435.15	6.5714	106.01
3	100%	316.50	22.21	268.29	3056.48	6.77803	79.69
	90%	313.54	21.17	254.96	3052.35	6.79205	75.15
	80%	302.99	18.10	218.49	3036.46	6.8322	62.73
4	100%	314.12	22.19	268.29	3051.00	6.76913	79.48
	90%	310.15	21.26	245.94	3044.28	6.77639	72.26
	80%	300.57	18.20	215.89	3030.67	6.8243	61.77
5	100%	538.61	20.70	268.29	3552.88	7.52483	100.32
	90%	536.32	19.51	254.96	3548.94	7.54695	94.60
	80%	533.09	16.69	218.49	3544.53	7.6127	79.63
6	100%	535.20	20.70	268.29	3545.30	7.51548	99.96
	90%	532.56	19.51	245.94	3540.60	7.53663	90.89
	80%	532.10	16.69	215.89	3542.35	7.61000	78.60
7		35.38	0.0559	226.10	2565.33	8.35741	7.41
8		32.26	0.0537	226.10	2563.00	8.36853	7.10
9		35.41	0.0559	310.15	146.88	0.52143	0.91
10		32.31	0.0537	243.31	135.89	0.46881	0.092
11		27.79	1.30	14000	116.62	0.40618	2.13
12		22.10	1.70	14000	92.86	0.32633	0.76
13		27.35	1.30	14000	114.78	0.40005	2.02
14		22.10	1.70	14000	92.86	0.32633	0.76
15	100%	33.85	0.054	553.46	141.83	0.48954	0.18
	90%	38.05	0.0681	511.26	159.39	0.54632	0.30
	80%	36.80	0.060	454.05	151.24	0.52005	0.21
16	100%	33.91	0.27	553.46	142.06	0.49021	0.19
	90%	38.30	0.22	511.26	160.45	0.54967	0.31
	80%	36.95	16.40	454.05	156.26	0.5309	0.44
17		60.71	6.37	553.46	254.64	0.83979	1.76
18		65.90	0.255	17.21	2619.19	7.82621	1.56
19		95.43	6.27	553.46	400.27	1.25469	5.16
20		173.23	1.22	22.10	2821.29	7.62652	3.60
21		99.87	1.01	84.05	418.54	1.30554	0.90
22		120.99	6.17	553.46	508.28	1.53811	9.24
23		250.40	2.39	21.67	2970.77	7.62746	4.48
24		127.84	3.08	37.11	537.20	1.61167	0.69
25		106.89	2.34	58.78	448.28	1.38413	0.74
26		146.20	5.97	536.70	615.96	1.80298	13.43
27		337.12	3.08	10.75	3145.33	7.81965	5.03
28		163.72	9.44	15.91	691.91	1.97952	0.51
29	100%	162.30	8.17	618.21	685.67	1.96553	19.25
	90%	162.20	8.10	574.86	685.23	1.96454	17.87
	80%	162.15	7.90	510.55	685.00	1.96406	15.86
30	100%	162.20	170.00	610.99	694.75	1.94554	21.45
	90%	162.20	170.00	567.16	694.75	1.94554	20.00
	80%	162.40	172.00	504.65	695.73	1.9472	17.86
31		177.70	169.20	610.99	761.51	2.09642	29.10
32		450.04	9.71	15.89	3371.56	7.63378	5.01
33		199.42	22.86	71.88	850.08	2.32413	3.42
34		218.72	167.41	610.99	942.31	2.48056	36.90
35		370.77	22.86	37.40	3178.34	6.96304	11.80
36		221.22	34.68	26.37	949.57	2.52706	1.55
37	100%	241.92	166.15	610.99	1048.35	2.69148	44.39
	90%	241.10	167.30	567.16	1035.98	2.66714	40.38
	80%	232.14	169.30	504.65	1003.29	2.60242	34.01
38		480.15	34.70	26.37	3406.78	7.10433	9.69
39		173.23	1.117	469.40	2821.66	7.67201	74.83
40		32.80	0.25	17.21	137.468	0.47522	0.005
41		450.00	7.00	5.70	3375.21	7.7884	6.24
0		20.00	1.01	-	84.01	0.29648	0

Table 3 Values used for thermodynamic analysis

Load	Coal *LHV* (kJ/kg)	Coal mass flow (t/h)
100%	8347.10	237.10
90%	8207.20	219.14
80%	8006.70	210.60

Table 4 Expressions of exergy efficiency and exergy destruction rate for each component

Boiler	$I_{\mathrm{B}} = m_{\mathrm{fuel}} e_{\mathrm{fuel}} + (E_{x37} + E_{x3} + E_{x3}) - (E_{x1} + E_{x2} + E_{x5} + E_{x6})$ $\eta_{\mathrm{exB}} = \frac{E_{xi}}{m_{\mathrm{fuel}} e_{\mathrm{fuel}}}$
Turbine	$I_{\mathrm{TUR}} = (E_{x1} + E_{x2} + E_{x5} + E_{x6}) - (E_{x\ \mathrm{extr}} + E_{x7} + E_{x8} + W_{\mathrm{el}})$ $E_{x\ \mathrm{extr}} = E_{x38} + E_{x35} + E_{x32} + E_{x27} + E_{x23} + E_{x20} + E_{x18}$ $\eta_{\mathrm{exTUR}} = \frac{W_{\mathrm{el}}}{(E_{x1} + E_{x2} + E_{x5} + E_{x6}) - (E_{x\ \mathrm{extr}} + E_{x7} + E_{x8})}$
Condenser 1	$I_{\mathrm{C1}} = (E_{x7} + E_{x21} + E_{x12}) - (E_{x11} + E_{x9})$ $\eta_{\mathrm{exC1}} = \frac{E_{x11} - E_{x12}}{E_{x7} + E_{x21} - E_{x9}}$
Condenser 2	$I_{\mathrm{C2}} = (E_{x8} + E_{x14} + E_{x40}) - (E_{x13} + E_{x10})$ $\eta_{\mathrm{exC2}} = \frac{E_{x13} - E_{x14}}{E_{x8} + E_{x40} - E_{x10}}$
Condensing pumps	$I_{\mathrm{CP}} = W_{\mathrm{CP}} + E_{x15} - E_{x16}$ $W_{\mathrm{CP}} = m_{15}(h_{16} - h_{15})$ $\eta_{\mathrm{exCP}} = \frac{E_{x16} - E_{x15}}{W_{\mathrm{CP}}}$
LPH 1	$I_{\mathrm{LHP1}} = (E_{x16} + E_{x18}) - (E_{x17} + E_{x40})$ $\eta_{\mathrm{exLHP1}} = \frac{E_{x17} - E_{x16}}{E_{x18} - E_{x40}}$
LPH 2	$I_{\mathrm{LHP2}} = (E_{x17} + E_{x20} + E_{x25}) - (E_{x19} + E_{x21})$ $\eta_{\mathrm{exLHP2}} = \frac{E_{x19} - E_{x17}}{(E_{x20} + E_{x25}) - E_{x21}}$
LPH 3	$I_{\mathrm{LHP3}} = (E_{x19} + E_{x23} + E_{x24}) - (E_{x25} + E_{x22})$ $\eta_{\mathrm{exLHP3}} = \frac{E_{x22} - E_{x19}}{(E_{x23} + E_{x24}) - E_{x25}}$
LPH 4	$I_{\mathrm{LHP4}} = (E_{x22} + E_{x27} + E_{x28}) - (E_{x24} + E_{x26})$ $\eta_{\mathrm{exLHP4}} = \frac{E_{x26} - E_{x22}}{(E_{x28} + E_{x27}) - E_{x24}}$
Deaerator	$I_{\mathrm{DEA}} = E_{x26} + E_{x41} + E_{x29}$ $\eta_{\mathrm{exDEA}} = \frac{m_{26}(e_{x29} - e_{x26})}{m_{41}(e_{x41} - e_{x29})}$
Feedwater pumps	$I_{\mathrm{FWP}} = W_{\mathrm{FWP}} + E_{x29} - E_{x30}$ $W_{\mathrm{FWP}} = m_{29}(h_{30} - h_{29})/\eta_{\mathrm{p}}$ $\eta_{\mathrm{exFWP}} = \frac{E_{x30} - E_{x29}}{W_{\mathrm{FWP}}}$
HPH 5	$I_{\mathrm{HPH5}} = (E_{x30} + E_{x32} + E_{x33}) - (E_{x31} + E_{x28})$ $\eta_{\mathrm{exHPH5}} = \frac{E_{x31} - E_{x30}}{(E_{x32} + E_{x33}) - E_{x28}}$
HPH 6	$I_{\mathrm{HPH6}} = (E_{x31} + E_{x35} + E_{x36}) - (E_{x34} + E_{x33})$ $\eta_{\mathrm{exHPH6}} = \frac{E_{x34} - E_{x31}}{(E_{x35} + E_{x36}) - E_{x33}}$
HPH 7	$I_{\mathrm{HPH7}} = (E_{x34} + E_{x38}) - (E_{x37} + E_{x36})$ $\eta_{\mathrm{exHPH7}} = \frac{E_{x37} - E_{x34}}{E_{x38} - E_{x36}}$

4 RESULTS AND DISCUSSION

Exergy losses of all components of the thermal power plant are shown in Fig. 2. It was found that the exergy destruction rate of the boiler is dominant over all other irreversibility in the cycle. Boiler exergy losses alone amount to 59 % of losses in the plant, while the exergy destruction rate of the condenser is only 0.84 to 1.07 %. Other components (HPH, LPH, feedwater pumps, condensing pumps and deaerator) have an exergy loss percentage of 0.001 to 0.4 %. Moreover, research shows that 38.50 % of exergy losses occur in turbine.

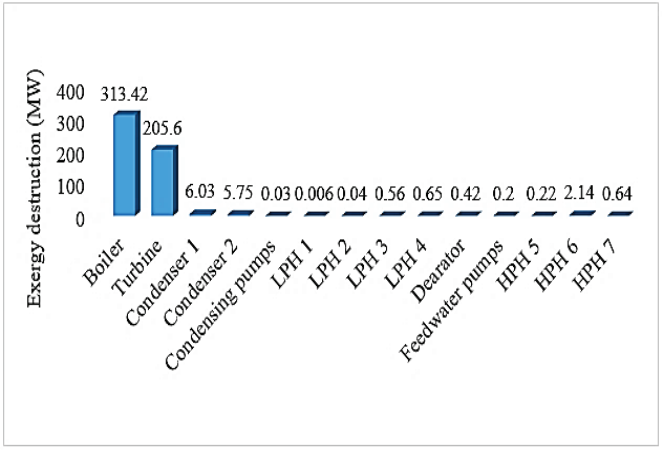


Figure 2 Exergy destruction of the thermal power plant components for nominal operation mode

The values of exergy destruction, the percentage values of exergy destruction and the exergy efficiency of all components for the nominal operation of the block are calculated and given in the Tab. 5.

Table 5 Exergy destruction and exergy efficiency of the thermal power plant components for nominal operation mode

	Exergy destruction (MW)	Percent exergy destruction (%)	Percent exergy efficiency (%)
Boiler	313.42	58.67	44.49
Turbine	205.60	38.49	49.42
Condenser 1	6.03	1.07	18.50
Condenser 2	5.75	0.84	17.95
Condensing pumps	0.03	0.006	28.57
LPH 1	0.006	0.001	99.56
LPH 2	0.04	0.007	98.80
LPH 3	0.56	0.10	87.08
LPH 4	0.65	0.12	86.61
Deaerator	0.42	0.078	54.22
Feedwater pumps	0.20	0.037	91.66
HPH 5	0.22	0.041	96.51
HPH 6	2.15	0.40	78.47
HPH 7	0.64	0.12	91.84

The exergy efficiencies of the thermal power plant components were calculated and shown in the Fig. 3. It is

found that condensing pumps with the exergy efficiencies of 28.60 % are the least efficient devices in the plant and LPH1 with exergy efficiency 99.60 % is the most efficient one. Components with lower exergy efficiency values are condenser 1 (18.50 %), condenser 2 (17.95 %), boiler (44.50 %), turbine (49.42 %) and deaerator (54.22 %).

The influence of the change in ambient temperature on the values of exergy losses of boiler, turbine and both steam condensers during operation of the unit at nominal load are shown in Fig. 4.

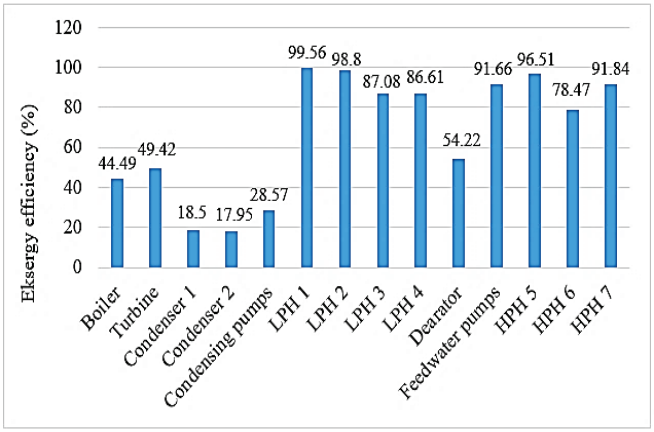


Figure 3 Exergy efficiency of the thermal power plant components for nominal operation mode

Fig. 4 shows that with increasing ambient temperature the boiler exergy losses increase and turbines and both steam condensers decrease. More detailed analysis of the influence of ambient temperature on the exergy efficiency of steam condensers for three different loads of the 200 MW unit was processed in the research of the authors of this paper [25].

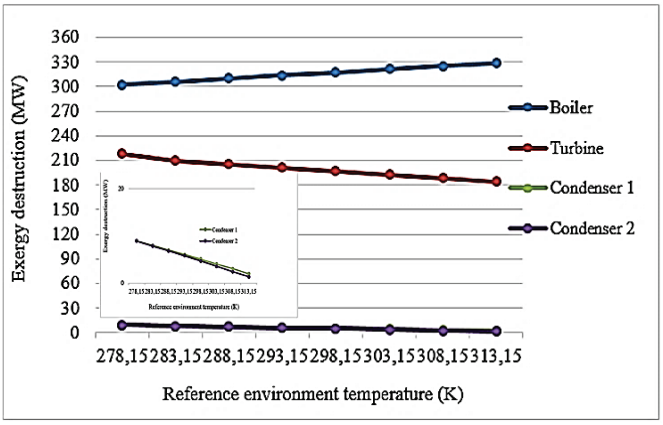


Figure 4 Exergy destruction in function of environment temperature

As previously presented, exergy efficiency power plant can be calculated based on two different methods using two different equations, Eq. (7) and Eq. (10).

Eq. (7) takes into account the energy carried by working fluid, neglecting the irreversibility of combustion process in furnace. Eq. (10) is based on exergy carried by the fuel combusted in furnace where irreversibility of the combustion process and exergy losses in exhaust gases are not neglected.

Thus obtained exergy efficiency values of the thermal power plant unit for operation at 100 %, 90 % and 80 % load are shown in the Fig. 5. For power plant unit operating at 100 % load, the values of exergy efficiencies are 77.26 % (Eq. (7)) and 34.37 % (Eq. (10)). The obtained values of exergy efficiency according to Eq. (10) refer to the coal consumption of 237.10 t/h, coal lower heating value of 8347.10 kJ/kg, the boiler efficiency of 87.88 % and the power at the generator terminals of 195.99 MW. Exergy efficiencies at 90 % load are 76.89 % and 35 %. These values were obtained for coal consumption of 219.20 t/h, coal lower heating value of 8207.20 kJ/kg, boiler efficiency of 88.24 % and electric generator power output of 182.30 MW. Operating at 80 % load, unit exergy efficiency values 75.58 % and 32.96 %. At the same time coal consumption is 210.60 t/h, coal lower heating value 8006.70 kJ/kg, boiler efficiency of 86.50 % and power at generator terminals of 161.50 MW.

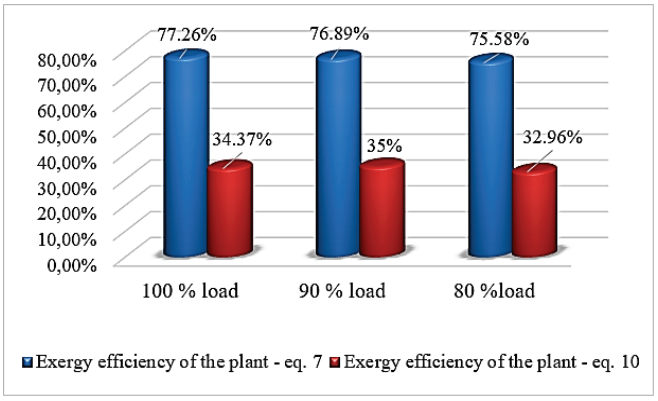


Figure 5 Exergy efficiency of the thermal power plant unit for operation at 100 %, 90 % and 80 % load

The influence of the 200 MW unit operation at the nominal regime without HPH in one and without HPH and LPH in the other case, on its exergy efficiency was calculated and shown in the Fig. 6. In the first case, when operating without HPH, efficiencies according to Eqs. (7) and (10) are 67.74 % and 28.80 %, respectively. These exergy efficiencies values are lower by 5.6 % and 9.8 % compared to operating a thermal power plant with HPH.

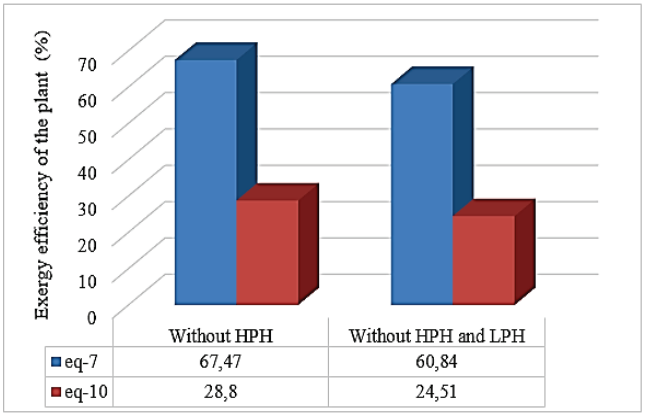


Figure 6 Exergy efficiency of the thermal power plant for operation at 100 % load without HPH and without HPH and LPH

In the other case, when operating thermal power plant unit without HPH and LPH, the exergy efficiencies amount to 60.84 % and 24.51 %, respectively, which are 9.90 % and 16.40 % lower than the values when operating with HPH and LPH.

Fig. 7 shows the exergy efficiencies of the boiler and turbine when the unit is operating at nominal mode without HPH and without HPH and LPH. Operation of the unit at nominal load without HPH and without HPH and LPH results with boiler efficiencies of 42.40 % and 40.30 %, respectively. Compared to the same values when unit is operating with HPH and LPH, these values are lower by 2 % and 4.2 %. Furthermore the exergy efficiencies of the turbine are 44.20 % and 40 % in the case when unit is operating without HPH and operating without HPH and LPH. When both HPH and LPH are operative, the turbine has by 5.2 % and 10 % higher exergy efficiency than in the previously mentioned case.

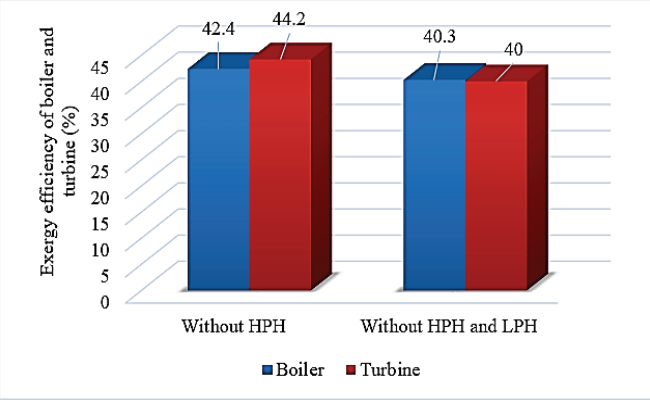


Figure 7 Exergy efficiency of boiler and turbine when unit is operating without HPH and without HPH and LPH

5 CONCLUSIONS

In this article, an analysis of the energy and exergy of a 200 MW steam block was performed, as well as the influence of the change in ambient temperature on the values of exergy losses and exergy efficiency. The analysis of this energy system revealed that the largest exergy loss is in the boiler and is 58.7%, followed by the turbine with an exergy loss of 38.5%. The percent exergy destruction in the condenser 1 and condenser 2 was 1.07 % and 0.84 % respectively, while all heaters, dearator and pumps destroyed less than 1 %.

With the change in ambient temperature, the percentage of exergy losses and exergy efficiency of all components also changed, but the conclusion remained the same, that the boiler and the turbine primarily affect the irreversibility of the analyzed cycle. The exergy efficiency of the block when working at nominal load is the highest, regardless of the method of calculating it. Different ways of calculating exergy efficiency lead to different values of exergy efficiencies and they are significantly less when the irreversibility of heat transfer in the boiler from flue gases to steam is taken into account in the calculation.

The exergy efficiency of the unit when operating at nominal load and without high - pressure heaters is 6 to 10 % lower, depending on the calculation method, while when operating the unit without high - pressure and low pressure heaters, it is lower by 10 to 16.5 %.

The exergy efficiencies of the boiler and turbine during the operation of the unit at nominal load and without high pressure heaters are 2 to 4.2 % less compared to the operation of the unit with high pressure heaters. The exergy efficiencies of the turbine at this load and operation without high pressure and low pressure heaters are further reduced and are 5 to 10 % lower. The operation of the unit at nominal load and without high pressure and low pressure heaters leads to a greater reduction in the exergy efficiency of the turbine than the boiler.

The analysis confirmed previous researches that indicate that the boiler within the steam block has the highest exergy losses. The analysis of exergy losses and exergy efficiency indicates that certain components such as LPH3, LPH4 and HPH6 have significantly lower values of exergy efficiency and that their performance can be improved with certain measures through revitalization or maintenance. Also, there is room for increasing exergy efficiency in both steam condensers. The fact that about 3 MW of cooling water exergy is released into the atmosphere at the cooling tower indicates the possibility of installing commercially available technologies for its use and generation of additional electrical and thermal energy.

NOMENCLATURE

Abbreviations:

HPH	High pressure regenerative system heaters
LPH	Low pressure regenerative system heaters
LHV	Fuel lower heating value (kJ/kg)

Latin Symbols:

m	mass flow rate (kg/s)
Q	heat transfer rate to the system (W)
W	work rate or power done by the system (W)
h	specific enthalpy (J/kg)
e_x	specific exergy (J/kg)
E_x	total exergy rate (W)
T	temperature (K)
s	specific entropy (J/kg K)
ϑ	temperature (°C)
p	pressure (bar)

Greek symbols:

φ	exergy factor
η	exergy efficiency

Subscripts:

net	netto
own consump	own consumption
TUR	turbine
o	dead state conditions
B	boiler
ex	exergy
extr	extraction
el	electrical
C1	condenser 1
C2	condenser 2
CP	condensing pump
FWP	feed water pumps
DEA	deaerator

P pump
i inlet
e exit

6 REFERENCES

[1] Kopac, M. & Hilalci, A. (2007). Effect of ambient temperature on the efficiency of the regenerative and reheat Çatalağzı power plant in Turkey. *Applied Thermal Engineering, 27*(8-9), 1377-1385. https://doi.org/10.1016/j.applthermaleng.2006.10.029

[2] Rosen, M. & Dincer, I. (2004). Effect of varying dead - state properties on energy and exergy analyses thermal systems. *International Journal of Thermal Sciences, 43(2),* 121-133 https://doi.org/10.1016/j.ijthermalsci.2003.05.004

[3] Utlu, Z. & Hepbasli, A. (2007). A review on analyzing and evaluating the energy utilization efficiency of countries. *Renewable and Sustainable Energy Reviews, 11*, 1-29. https://doi.org/10.1016/j.rser.2004.12.005

[4] Kanoglu, M., Dincer, I., & Rosen, A. M. (2007). Understanding Energy and Exergy Efficiencies for Improved Energy Management in Power Plants, *Energy Policy, 35*, 3967-3978. https://doi.org/10.1016%2Fj.enpol.2007.01.015

[5] Yang, Y., Wang, L., Dong, C., Xu, G., Morosuk, T., & Tasatsaronis, G. (2013). Comprehensive Exergy – Based Evaluation and Parametric Study of a Coal – fired Ultra Supercritical Power Plant. *Aplied Energy, 112,* 1087-1099. https://doi.org/10.1016/j.apenergy.2012.12.063

[6] Aljundi, H. (2009). Energy and Exergy Analysis of a Steam Power Plant in Jordan, *Applied Thermal Engineering*, *29*, 324-328. https://doi.org/10.1016/j.applthermaleng.2008.02.029

[7] Ehsan, A. & Yilmazoglu, M. Z. (2011). Design and Exergy Analysis of a Thermal Power Plant Using Different Types of Turkish Lignite, *International Journal of Thermodynamics, 14(3),* 125-133. https://doi.org/10.5541/ijot.288

[8] Erdem, H. H., Akkaya, A. V., Cetin, B., Dagdas, A., Sevilgen, H. K., Sahin, B., Teke, I., Gungor, C., & Atas, C. (2009). Comparative Energetic and Exergetic Performance Analyses for Coal - Fired Thermal Power Plants in Turkey. *International Journal of Thermal Sciences, 48,* 2179-2186 https://doi.org/10.1016%2Fj.ijthermalsci.2009.03.007

[9] Hastia, S., Aroonwilasa, A., & Veawab, A. (2013). Exergy Analysis of Ultra Supercritical Power Plant. *Energy Procedia, 37,* 2544-2551. https://doi.org/10.1016/j.egypro.2013.06.137

[10] Hongbin, Z. & Yuman, C. (2009). Exergy Analysis of a Steam Power Plant with Direct Air - Cooling System in China. *IEEE School of Mechatronics Engineering*, 1-4. https://doi.org/10.1109%2FAPPEEC.2009.4918346

[11] Ghasemzadeh, B. & Sahebi, Y. (2010). Optimization and Exergy Analysis for Advanced Steam Turbine Cycle. The *2nd International Conference on Mechanical and Electronics Engineering,* 2, 190.

[12] Dincer, I. & Muslim, H. A. (2001). Thermodynamic Analysis of Reheat Cycle Steam Power Plants, *International Journal of Energy Research, 25,* 727-739. https://doi.org/10.1002/er.717

[13] Ray, T. K., Datta, A., Gupta, A., & Ganguly, R. (2010). Exergy Based Performance Analysis for Proper O&M Decisions in a Steam Power Plant. Energy *Conversion and Management, 51*, 1333-1344. https://doi.org/10.1016/j.enconman.2010.01.012

[14] Gogoi, T. K. & Talukdar, K. (2014). Thermodynamic Analysis of A Combined Reheat Regenerative Thermal Power Plant and Water – LiBr Vapor Absorption Refrigeration System. *Energy Conversion and Management, 78*, 595-610. https://doi.org/10.1016/j.enconman.2013.11.035

[15] Kanoglu, M., Cengel, Y. A., & Dincer, I. (2012). *Efficiency Evaluation of Energy Systems*, Springer, New York. https://doi.org/10.1007/978-1-4614-2242-6

[16] Rashad, A. & Maihy, A. E. (2009). Energy and Exergy Analysis of a Steam Power Plant in Egypt. *The 13th International Conference on Aerospace Sciences & Aviation Technology*, 1-12.

[17] Sengupta, S., Datta, A., & Duttagupta, S. (2007). Exergy Analysis of a Coal - Based 210 MW Thermal Power Plant. *International Journal of Energy Research*, *31*(1), 14-28 https://doi.org/10.1002/er.1224

[18] Živić, M., Galović, A., Avsec, J., & Holik, M. (2016). Exergy analysis of a Brayton cycle with variable physical properties and variable composition of working substance. *Technical Gazette, 23*(3), 801-808. https://doi.org/10.17559/TV-20160208112755

[19] Ćehajić, N. & Hafizović, S. (2021). Comparative analysis of power distribution, exergy losses and exergy efficiencies of the 200 MW steam turbine for three operating modes with and without turbine steam extraction. *Tehnika, 76*(6), 765-773. https://doi.org/10.5937/tehnika2106765c

[20] Noroozian, A., Mohammadi, A., Bidi, M., & Ahmadi, M. H. (2017). Energy, exergy and economic analyses of a novel system to recover waste heat and water in steam power plants. *Energy conversion and management*, *144*, 351-360. https://doi.org/10.1016/j.enconman.2017.04.067

[21] Kumar, S., Kumar, D., Memon, R. A., Wassan, M. A., & Ali, M. S. (2018). Energy and exergy analysis of a coal fired power plant. *Mehran University Research Journal of Engineering & Technology*, *37* (4), 611-624. https://doi.org/10.22581/muet1982.1804.13

[22] Adibhatla, S. & Kaushik, S. C. (2017). Energy, exergy, economic and environmental (4E) analyses of a conceptual solar aided coal fired 500 MWe thermal power plant with thermal energy storage option. *Sustainable Energy Technologies and Assessments*, *21*, 89-99. https://doi.org/10.1016/j.seta.2017.05.002

[23] Kotas, T. J. (1995). *The Exergy Method in Thermal Plant Analysis,* second ed. Krieger, Malabar

[24] Lemmon, E., Huber, M., & McLinden, M. (2007). *NIST Reference Fluid Thermodynamic and Transport Properties REFPROP 8, NIST Standard Reference Database 23*.

[25] Ćehajić, N. & Žigić, H. (2021). Parametric analysis of steam condenser block of 200 MW, *Scientific Review, 1*(8).

Author's contacts:

Nurdin Ćehajić, PhD
JP Elektroprivreda BiH, Termoelektrana Tuzla,
21. april 4 Bukinje, 75203 Tuzla, Bosnia and Herzegovina
E-mail: nurdin.cehajic@epbih.ba

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220504103756

Original scientific paper
Received: 2022-05-04, Accepted: 2022-11-02

Error Analysis for Designed Test and Numerical Integral by Using UED in Material Research

Maosheng Zheng*, Jie Yu, Haipeng Teng, Yi Wang

Abstract: In this article, the error analysis for designed test and definite integral in employing uniform experimental design is analogically developed on basis of midpoint rule in rectangle method for assessing definite integral through conducting discretized sampling approximately. It concluded that the discrete sampling-point by means of good lattice point in the evaluations of definite integral and maximum value of a function is promised with higher accuracy, and the predicted entire error of this uniform sampling point method for above problems decreases with the number of sampling points significantly.

Keywords: definite integral; error analysis; midpoint rule; rectangle method; uniform experimental design

1 INTRODUCTION

Experiment design is an essential topic to scientific and industrial developments. How do we arrange and design experiments to get effective results with less number of trials? Some approaches are proposed to answer this encountered question frequently, such as *Response Surface Methodology*, *Orthogonal Experimental Design*, and *Uniform Experimental Design*, etc. Better design could lead to results that are more effective.

In 1978, due to the need for missile designs, a special demand for total trial number being no greater than 50 with 18 factor levels of a five-factor experiment was faced to Prof. K. T. Fang [1], He and Prof. Y. Wang thus proposed a novel solution by employing number-theoretical methods [2]. They created a brand new experimental design methodology to conduct the design of the problem in that time, which was named as Uniform Experimental Design (UED). UED is attributed to number - theoretical method or quasi-Monte Carlo method. This novel methodology was employed to the design of missiles successfully, and got series of meaningful achievements in China due to its wide applications [1, 2].

Early in 1950s, Ulam and Von Neumann developed Monte Carlo method, which is a statistical stimulation. The main idea of their method is to convert an analysis problem into a probability problem and then employ a statistical simulation to deal with the problem to get a same solution. This seems an effective solution for some difficult analysis problems, including the approximate estimation of complicated definite integrals. The general idea of Monte Carlo method is the need of a set of stochastic numbers to enable to perform this statistical simulation. The precision of this method strongly depends on the independence and uniformity of stochastic numbers.

Almost in the same period of 1950s, deterministic methods were also proposed to deal with some difficult analysis problems by some mathematicians, which aimed to give solution by using uniformly distributed points in space instead of random numbers like that in Monte Carlo method [3], such as, Korobov put forward the concept of a point set, which is uniformly distributed. Since then Hua et al developed the good lattice point (GLP) method in 1960s for evaluation of definite integral approximately, which is with low-discrepancy on basis of number theory [2, 3]. Therefore, this kind of method is called a number-theoretical method or quasi-Monte Carlo method naturally thereafter. UED can be seen as one of the successful applications of the number-theoretical method [1, 2]. This methodology was subsequently used in evaluations for approximation of multiple-integral successfully.

There are many beneficial features of GLP [4-6], besides the uniformity of distribution of sampling-point over the specific domain and good space-filling characteristic.

Currently, the UED finds its wide applications over the world, it spreads in designs of Chinese medicine, chemical reaction and missile, as well as Ford Motor Co. Ltd for its design of standard exercises and automotive as computer experiments for providing a support of the preliminary design of production [7].

1.1 Essential Characteristics of UED

The essential characteristics of UED involves [1, 2, 7]:

A) **Homogenization**
The distribution of specimen point is homogeneously scattered in the variable space; therefore it gains a surname of "space filling design" occasionally [1, 2, 7]. A number of "Uniform Design Table" (UDT) was specifically developed by Fang to arrange the distribution of the specimen point for UED [8], which is fully deterministic.

B) **Entire Mean Model**
UED intends to get an outcome consequence, which is with minimum deviation of the entire averaged value from the actual total averaged value through uniform distribution of specimen points.

C) **Robustness**
UED is expected to be used in many cases with robustness regardless of variation of model.

1.2 Basic Principle of UED

The fundamental principle of UED is as followings:

1) Entire Mean Model

The fundamental hypothesis is the existence of a deterministic relationship of the response g *vs.* the independent input variables $x_1, x_2, x_3, ..., x_s$, the formula of the response can be expressed as,

$$g = G(x_1, x_2, x_3, ..., x_s), x = \{x_1, x_2, x_3, ..., x_s\} \in C^s. \quad (1)$$

The further hypothesis is that the entire averaged value of the response g on $C^s = [0, 1]^s$ is,

$$\overline{E(g)} = \int_{C^s} G(x_1, x_2, x_3, ..., x_s)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3, ..., \mathrm{d}x_s. \quad (2)$$

Moreover, if one takes m sampling points $q_1, q_2, q_3, \ldots, q_m$ on C^s to conduct an average value of g, then the averaged value of g over these m specimen points is,

$$\overline{g(D_m)} = \frac{1}{m}\sum_{i=1}^{m} G(q_i). \quad (3)$$

In Eq. (3), $D_m = \{q_1, q_2, q_3, \ldots, q_m\}$ represents a design of such m specien points.

Fang et al indicated that the deviation $\varepsilon = \overline{E(g)} - \overline{g(D_m)}$ of the specimen point set on C^s and D_m will be quite small provided the specimen points $q_1, q_2, q_3, \ldots, q_m$ are uniformly distributed in the domain C^s.

2) Uniform Design Table

In order to provide an appropriate application for UED, a number of UDT as well as the "Utility Table" were conducted [8], with which the location of specimen points can be specifically determined with convenience.

3) Regression

In general, under condition of discretization with sampling points, an approximate expression for response $r' = R'(x_1, x_2, x_3, ..., x_m)$ *vs.* the independent input variables can be regressed to reveal the resemble formation.

1.3 Aim of This Study

Until now, the estimation of the accuracy of applying each UDT to conduct actual problem is unclear though the discrepancy of each point set of UDT is provided by Fang in his book [5].

In this paper, the study of entire error of definite integral and maximum value by using sampling points from UDT is preliminarily conducted in the point of view of practical application.

2 ERROR ANALYSIS FOR APPLYING UED TO CONDUCT ACTUAL PROBLEMS

2.1 Distribution Rule of UED in Space

A uniform table $U_n(n^q)$ has n rows and q columns, and q is the Euler function for a positive integer n, $q = \varphi(n) = n\cdot(1 - 1/p_1)\cdot(1 - 1/p_2)\cdot\ldots\cdot(1 - 1/p_v)$ for each n. According to number theory, it assumes that there is a unique prime decomposition $n = p_1^{r_1} \cdot p_2^{r_2} \cdot \ldots \cdot p_v^{r_v}$ for each n [1, 2, 8], in which $r_1, r_2, ..., r_v$ express positive integers and $p_1, p_2, \ldots, p_v$ indicate different primes. Furthermore, the number of independent factors at most is $t = \varphi(n)/2 + 1$, i.e., $s \le t$ for each n, s is the actual number of the independent variables of the studied problem [1, 2, 8].

In the hyper cubic $[0, 1]^s$, the specimen point is homogeneously distributed. Especially, for one independent variable case, the design $X^* = [1/2n, 3/2n, 5/2n, ..., (2n–3)/2n, (2n–1)/2n]^T$ is the unique design on [0, 1] with low - discrepancy or under star discrepancy of $D^*(X^*) = 1/2n$ [1, 2, 8].

Obviously, the homogeneously spreading of the specimen point in the above design for one independent variable case is the same as that of the Midpoint Rule in rectangle method of definite integral [9-11].

Following midpoint rule, the definite integral $E(g) = \int_{x_0-\delta/2}^{x_0+\delta/2} g(x)\mathrm{d}x$ around position x_0 with a subsection δ is approximated by $I = g(x_0)\cdot\delta$, which is with the local error of $\varepsilon_M = \delta^3\cdot g''(x_0)/24$ [9-11], where $g''(x_0)$ indicates the second derivative at location x_0, in general ε_M is negligible as δ is sufficiently small.

According to mean value theorem of integral, one could always find a proper position ζ within $[x_0 - \delta/2, x_0 + \delta/2]$, such that the value of the function $g(\zeta)$ meets the demand of $g(\zeta)\cdot\delta = E(g) = \int_{x_0-\delta/2}^{x_0+\delta/2} g(x)\mathrm{d}x$. However, above discussion indicates that the value of the integral approximates to $I = g(x_0)\cdot\delta$ with the small local error of $\varepsilon_M = \delta^3\cdot g''(x_0)/24$ [9-11], therefore, $g(\zeta)$ approximately equals to $g(x_0)$, i.e., $g(\zeta) \approx g(x_0)$.

Furthermore, the local error of the function $g(x)$ around x_0 within area of $[x_0 - \delta/2, x_0 + \delta/2]$ is about $g'(x_0)\cdot\delta/2$.

Moreover, considering the maximum value of the function $g(x)$ within its range of $[a, b]$, it supposes that if the function $g(x)$ within its range $[a, b]$ derives its maximum value $g(x_l)$ at a discrete point $x_l = a + (2l - 1)(b - a))/2n$, $l \in [1, 2, 3, \ldots, n]$, then the realistic error of the actual maximum of this function $g_{max}(x)$ from the nominal maximum $g(x_l)$ at a discrete point x_l is $E_{actual} = g_{max}(x) - g(x_l)$, and it could be approximately estimated by,

$$E_{est.} = g'(x_l)\cdot\delta/2 \approx \left|g(x_{l+1}) - g(x_l)\right|/2. \quad (4)$$

Table 1 Distribution of 11 sampling points within domain [0, 1.5]

No.	Position	e^x
1	0.0682	1.0706
2	0.2045	1.2270
3	0.3409	1.4062
4	0.4773	1.6117
5	0.6136	1.8471
6	0.7500	2.1170
7	0.8864	2.4263
8	1.0227	2.7808
9	1.1591	3.1871
10	1.2955	3.6527
11	1.4318	4.1863

As an example, lets' conduct the error analysis for e^x within range of [0, 1.5] by using 11 discrete uniform sampling points [12, 13]. As to such problem, following the

procedure of UED [1, 2, 8], the 11 sampling points are uniformly distributed within range of [0, 1.5], see Tab. 1.

The actual value of function e^x is 4.4817 at $x = 1.5$, while the maximum value of discretized function e^x in Tab. 1 is 4.1863. The actual error is $E_{\text{actual}} = 0.2954$, while the estimated value by using Eq. (4) is $E_{\text{est.}} = 0.2668$, which is close to E_{actual}.

2.2 Error Analysis of Applying UED for Maximum Value

In general, as to a s-dimensional problem, it assumes that the discretization is conducted for a function $\vec{x}_p$ within its domain $[0, 1]^s$, and the maximum $g(\vec{x}_p)$ of this function $g(\vec{x})$ is at the discrete point $\vec{x}_p$ due to this discretization, then the error of the maximum $g(\vec{x}_p)$ at the discrete point $\vec{x}_p$ from the actual maximum $g_{\max}(\vec{x})$ of the function $g(\vec{x})$ due to this discretization is $E_{\text{actual}} = g_{\max}(\vec{x}) - g(\vec{x}_p)$, and it can be estimated by following equation,

$$E_{\text{est.}} \approx \frac{1}{2\gamma}\sum_{i=1}^{\gamma}\left|g(\vec{x}_{p+i}) - g(\vec{x}_p)\right|. \tag{5}$$

In Eq. (5), γ is the number of nearest neighbour of $\vec{x}_p$.

Thus, the error of the maximum $g(\vec{x}_p)$ at the discrete point $\vec{x}_p$ from the actual maximum $g_{\max}(\vec{x})$ of the function $g(\vec{x})$ corresponding to the discretization can be estimated by using Eq. (5) in principle.

2.3 Error Analysis of Applying UED for Integral

As to midpoint rule [9-11], the definite integral $\overline{E(g)} = \int_a^b g(x)\mathrm{d}x$ can be approximated by a summation, i.e., $g(D_n) = \frac{(b-a)}{n}\sum_{i=1}^{n} g\left[a + (i-1/2)\cdot(b-a)/n\right]$ in 1-dimension, while the entire error is,

$$E_{\text{M}} = \frac{1}{24}\cdot\frac{(b-a)^3}{n^2}\cdot\frac{1}{n}\sum_{i=1}^{n} g''\left[a + (i-1/2)\cdot(b-a)/n\right] \approx$$
$$\approx \frac{(b-a)^2}{24n^2}\cdot\left\{g'\left[b-(b-a)/2n\right] - g'\left[a+(b-a)/2n\right]\right\}, \tag{6}$$

where $g''(a + (i - 1/2)\cdot(b - a)/n)$ is the 2nd derivative at position "$a + (i - 1/2)\cdot(b - a)/n$", and $g'[a + (b - a)/2n]$ expresses the 1st derivative at position "$a + (b - a)/2n$".

If the integrand $g(x)$ behaves a wavy form, the summarizing $\varepsilon_n = \sum_{i=1}^{n} g''\left[a + (i-1/2)\cdot(b-a)/n\right]$ is quite tiny; or else the summarizing ε_n remains for some monotone function like e^x. However even in latter case good precision could be obtained with 11 specimen points which are homogeneously distributed [12, 13]. Take the integral $\int_0^{1.5} e^x\mathrm{d}x$ as an example, which is with the precise value of $\overline{E(g)} = \int_0^{1.5} e^x\mathrm{d}x = 3.4817$, while the summarizing of the discrete specimen points in manner of GLP is $\overline{g(D_{11})} = \frac{1.5}{11}\sum_{i=1}^{11} e^{[(i-0.5)\cdot 1.5/11]} = 3.4790$, the actual total error is 0.0027, while the predicted value of total error is of 0.0025, which is not far from the actual error.

For higher dimensions, the entire error could be estimated analogically by,

$$\left|E_{\text{M}}\right| \le \frac{1}{24\cdot s\cdot n}\sum_{i=1}^{s}(b_l - a_l)^3\cdot M(l). \tag{7}$$

In Eq. (7), the term $M(l)$ indicates $\max|g''(x)\|$ in l^{th} independent variable, $x_l \in [a_l, b_l]$.

Until now, the error estimation of applying UED for integral is estimated in principle.

3 APPLICATIONS

3.1 Error Analysis of Maximum Value of Functions in Their Domains by Using Discrete Uniform Sampling Points

Lets' study the function $f_1(x, y) = \ln(x + 2y)$ in the domain of $[1.4, 2.0] \times [1.0, 1.5]$ by using discrete uniform sampling points first. Assume that the actual maximum value of the function $f_1(x, y)$ within the domain is $f_{1\max}(x, y)$, and the maximum value of the function $f_1(x, y)$ due to the discretization is at a discrete point $\vec{x}_p$ and denoted by $f_1(\vec{x}_p)$. Fig. 1 shows the variations of the actual error $E_{\text{actual}} = |f_{1\max}(x, y) - f_1(\vec{x}_p)|$ *vs.* the number of sampling points n, together with the estimated error $E_{\text{est.}}$. It can be seen that the tendency of the variation of the actual error $E_{\text{actual.}}$ is the same as the those of the estimated error E_{est}, which decreases significantly with the increase of the number of sampling points n.

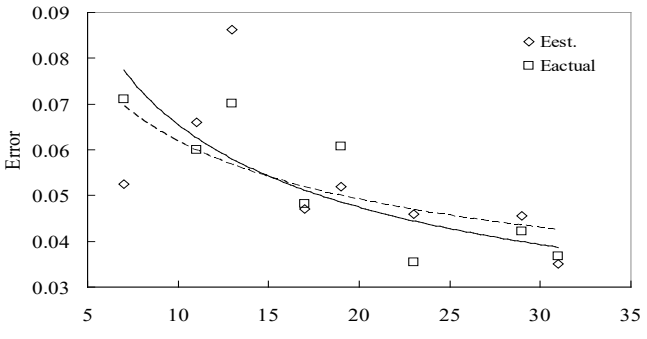


Figure 1 Comparison of the variations of the actual error E_{actual} and the estimated error $E_{\text{est.}}$ *vs.* number of sampling points n for $f_1(x, y) = \ln(x + 2y)$

Next, Lets' study the function $f_2(x, y) = 1 + 2x^2 + 2y^3$ in the domain of $[1.4, 2.0] \times [1.0, 1.5]$ by using discrete uniform sampling points. Fig. 2 shows the comparison of the variations of the actual error E_{actual} and the estimated error $E_{\text{est.}}$ *vs.* the number of sampling points n for $f_2(x, y)$. It can be seen again that the tendency of the variations of the actual error E_{actual} and the estimated error $E_{\text{est.}}$ is the same, which

decreases with the increase of the number of specimen points n obviously. The accuracy of the estimation varies with the exact detail of the function $f(x, y)$.

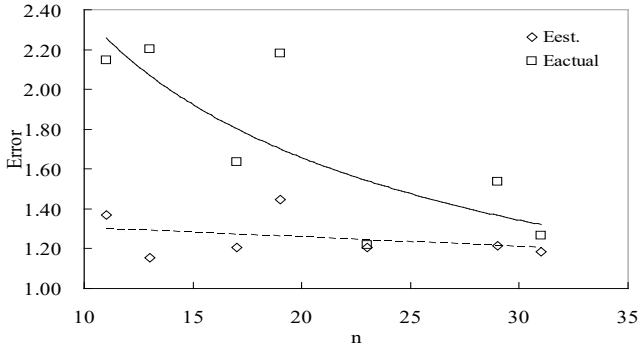


Figure 2 Comparison of the variations of the actual error E_{actual} and the estimated error $E_{est.}$ *vs.* number of sampling points n for $f_2(x, y) = 1 + 2x^2 + 2y^3$

3.2 Error Analysis of Definite Integral in 2-D

Under condition of higher dimensions, Fang set up a number of UDTs and the corresponding utility tables specifically for appropriate utility of UED, which is based on GLP and number-theoretic methods [1, 2, 8].

Now, lets' take a definite integral of a 2-dimensional problem as an example, i.e.,

$$H = \int_{x=1.4}^{2.0} \mathrm{d}x \int_{y=1.0}^{1.5} H(x, y)\mathrm{d}y \equiv \int_{x=1.4}^{2.0} \mathrm{d}x \int_{y=1.0}^{1.5} \ln(x + 2y)\mathrm{d}y. \quad (8)$$

The UDT $U_{17}(17^8)$ is tried to be used to conduct the definite integral, 17 sampling points are included [12, 13].

The accurate data of this integration is 0.429560 [14].

The specimen points of $U_{17}(17^8)$ within the integral range [1.4, 2.0] × [1.0, 1.5] are distributed uniformly and shown in Tab. 2, the values of function $H(x, y)$ at each discrete location are displayed in Tab. 2 as well. The symbols x_{10} and y_{20} in Tab. 2 show the nominal locations of the original table $U_{17}(17^8)$ within the domain [1, 17] × [1, 17] [2, 3].

Table 2 Positions of the specimen points within domain [1.4, 2.0] × [1.0, 1.5] by means of $U_{17}(17^8)$

No. = x_0	y_0	x	y	H
1	11	1.4176	1.3088	1.3951
2	5	1.4529	1.1324	1.3131
3	16	1.4882	1.4559	1.4816
4	10	1.5235	1.2794	1.4067
5	4	1.5588	1.1029	1.3257
6	15	1.5941	1.4265	1.4922
7	9	1.6294	1.25	1.4181
8	3	1.6647	1.0735	1.3381
9	14	1.7	1.3971	1.5028
10	8	1.7353	1.2206	1.4295
11	2	1.7706	1.0441	1.3504
12	13	1.8059	1.3676	1.5132
13	7	1.8412	1.1912	1.4407
14	1	1.8765	1.0147	1.3625
15	12	1.9118	1.3382	1.5235
16	6	1.9471	1.1617	1.4518
17	17	1.9824	1.4853	1.6000

In according with the method of UED [1, 2, 8], the integration H in the domain [1.4, 2.0] × [1.0, 1.5] becomes a summation in the discrete specimen points as

$$H \approx \frac{0.6 \times 0.5}{17} \sum_{i=1}^{17} H(x_i, y_i). \quad (9)$$

Summarizing Eq. (9) acquires a result, says 0.429613, which causes an error of 5.3×10^{-5} as respect to its actual result of 0.429560 [14], while the estimated error from the sampling points of the uniform design is 7.6×10^{-5}.

3.3 Comparative Assessment of Discrete Specimen Points by Means of GLP with Monte Carlo Simulation for Integral

$$F \equiv \int_0^6 f(x)\mathrm{d}x = \int_0^6 (2 - x/3)\mathrm{d}x$$

The accurate value of the definite integral of $F = \int_0^6 (2 - x/3)\mathrm{d}x$ is 6.

Han and Ren once assessed this integral by employing Monte Carlo algorithm [15].

Now, lets' try to re-study this integral by means of GLP with 11 discrete specimen points for comparison [15]. The locations of the specimen points and the discrete values of the function $f(x)$ are given in Tab. 3.

Table 3 Locations of the specimen points and the discrete data of the function

No.	x	$f(x)$
1	0.2727	1.9091
2	0.8182	1.7273
3	1.3636	1.5455
4	1.9091	1.3636
5	2.4545	1.1818
6	3	1
7	3.5455	0.8182
8	4.0909	0.6365
9	4.6364	0.4545
10	5.1818	0.2727
11	5.7273	0.0909

Following the procedure of UED, the definite integral F now becomes a summation of the 11 discrete specimen points within the integral domain [0, 6]. The summation gives a value of $\frac{6}{11}\sum_{i=1}^{11} f(x_i) = 6$, which is exactly and luckily same as that of the accurate value of this integral. However, the use of Monte Carlo simulation results in a varying error, which changes even up to 1000 random specimen points [15]. As was shown in [15], the Monte Carlo simulation gives an error of 0.1214 at 1000 specimen points, while the error of the simulated result reaches to 0.2908 with 100 specimen points [15]! This result reflects the merit of the assessment with discrete specimen points by means of GLP in evaluating definite integral and maximum value of a function once more.

4 CONCLUSION

Error analysis of designed test and definite integral by employing UED to conduct discrete specimen is analyzed in

this paper, the estimation of error is performed with the aid of midpoint rule in rectangle method for predicting definite integral analogically. The study indicates the decreasing tendency of the entire error of specimen point in predictions of definite integral and maximum value of a function *vs.* the number of specimen points, and the proper number of specimen points can be determined by the promised requirement of accuracy conversely.

Conflict Statement

There is no conflict of interest.

5 REFERENCES

[1] Fang, K. T., Liu, M. Q., Qin, H., & Zhou, Y. D. (2018). *Theory and Application of Uniform Experimental Design*. Science Press, Beijing, China, Springer Nature, Singapore. https://doi.org/10.1007/978-981-13-2041-5

[2] 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

[3] Hua, L-K. & Wang, Y. (1981). *Applications of Number Theory to Numerical Analysis*. Berlin, New York: Springer-Verlag & Beijing: Science Press.

[4] Elsawah, A. M., Fang, K.-T., & Deng, Y. H. (2021). Some interesting behaviors of good lattice point sets. *Communications in Statistics - Simulation and Computation, 50*(11), 3650-3668. https://doi.org/10.1080/03610918.2019.1628988

[5] Yang, L., Zhou, Y., & Liu, M.-Q. (2021). Maximin distance designs based on densest packings. *Metrika, 84*, 615-634. https://doi.org/10.1007/s00184-020-00788-w

[6] Huang, Y. & Zhou, Y. (2022). Convergence of uniformity criteria and the application in numerical integration. *Mathematics, 10*, 3717. https://doi.org/10.3390/math10193717

[7] Liu, M. Q., Lin, D. K. J., & Zhou, Y. (2020). The Contribution to Experimental Designs by Kai-Tai Fang. In *Contemporary Experimental Design, Multivariate Analysis and Data Mining, Festschrift in Honour of Professor Kai-Tai Fang*, eds. by J. Fan, J. Pan, Springer, 21-35, Cham, Switzerland. https://doi.org/10.1007/978-3-030-46161-4

[8] Fang, K. T. (1994). *Uniform Design and Uniform Design Table*. Science Press, Beijing, China.

[9] Holmes, M. H. (2016). *Analysis Introduction to Scientific Computing and Data Analysis*. Springer Nature Switzerland AG, Cham, Switzerland. https://doi.org/10.1007/978-3-319-30256-0

[10] Izaac, J. & Wang, J. (2018). *Computational Quantum Mechanics*. Springer Nature Switzerland AG, Cham, Switzerland. https://doi.org/10.1007/978-3-319-99930-2

[11] Stickler, B. A. & Schachinger, E. (2016). *Basic Concept in Computational Physics*, 2nd Ed. Springer Nature Switzerland AG, Cham, Switzerland. https://doi.org/10.1007/978-3-319-27265-8

[12] Yu, J., Zheng, M., Teng, H., & Wang, Y. (2022). An efficient approach for calculating a definite integral with about a dozen of sampling points. *Vojnotehnicki Glasnik*, 70(2), 340-356. https://doi.org/10.5937/vojtehg70-36029.

[13] Zheng, M., Teng, H., Yu, J., Cui, Y., & Wang, Y. (2023). *Probability-Based Multi-objective Optimization for Material Selection.* Springer Science and Business Media LLC, Singapore. https://doi.org/10.1007/978-981-19-3351-6

[14] Song, S. & Chen, S. (2004). Two Effective Quadrature Schemes for Calculating Double Integration. *Journal of Zhengzhou University, 36*, 16-19. https://en.cnki.com.cn/Article_en/CJFDTOTAL-ZZDZ200401003.htm

[15] Han, J. & Ren, W. (2007). Monte Carlo Integration and Quasi-Monte Carlo Integration. *Journal of Shanxi Normal University (Natural Science Ed.), 121*(1), 13-17.

Authors' contacts:

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

Jie Yu
School of Life Science, Northwest University,
Xi'an, 710069, China

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-20220604112718

Original scientific paper
Received: 2020-06-04, Accepted: 2022-08-10

Quick Review: Uncertainty of Optimization Techniques in Petroleum Reservoir Management

Amir Naser Akhavan*, Seyed Emad Hosseini, Mohsen Bahrami

Abstract: The notable increase in petroleum demand, together with a decline in discovery rates, has highlighted the desire for efficient production of existing oil wells worldwide. Mainly, the productivity of the existing large oil fields makes us consider the principles of managing reservoirs to make the most of extraction. At the same time, many different uncertainties in the course of the developing oil field, including geological, operational, and economic uncertainties, have a detrimental impact on the reservoir's effective production, which is why dealing with uncertainty is crucial for maximizing output. There is a broad variety of studies on managing oil reservoirs under uncertainty information in the literature. In this study a short review of earlier works has been done on optimization strategies and management of uncertainty in reservoir production.

Keywords: management of reservoir; management of uncertainty; optimization approaches

1 INTRODUCTION

For many years, oil has defined the world economy, this trend will most likely continue in the coming years. Demand for oil has increased with the population growth, and this demand needs to be met with proper optimization. Optimization management between the two factors of oil field development and oil production methods should be done that overall demand does not face problems. Proper planning for optimization requires ground and underground, data these data always have uncertainties that should be considered.

Uncertainty is due to incomplete and imprecise knowledge as a result of limited sampling of the subsurface heterogeneities. Well data and seismic data have incomplete coverage and finite resolution. Sub ground and oil Reservoirs are heterogeneous and difficult to predict away from wells or seismic data. Ignoring uncertainty and locking in important model parameters and choices amounts to an assumption of perfect knowledge and is generally an unacceptable approach. One way to reduce the effects of uncertainty on optimizations is to model them. Understanding the (1) sources of uncertainty, (2) methods to represent uncertainty, (3) the formalisms of uncertainty, and (4) uncertainty modeling methods and workflows were essential for the integration of all reservoir information sources and providing good models for decision making in the presence of uncertainty and improvement of optimization.

Geophysical prospecting consists of making a quantitative inference about subsurface properties from geophysical measurements. Due to many ineluctable difficulties, observed data are almost always insufficient to uniquely specify the rock properties of interest. Hence, inevitable uncertainty remains after the estimation. The sources of the uncertainty arise from many factors: inconsistency in data acquisition conditions, insufficient available data as compared to the subsurface complexities, limited resolution, imperfect dependence between observed data and target rock properties, and our limited physical knowledge. While the uncertainty has been identified for a long time, quantitative framework to discuss the uncertainty has not been well established.

In this study we examine several sources of uncertainty in the development of the oil field make the estimation of the future productivity of a reservoir inaccurate. In general, uncertainty in information about the management of reservoirs falls into four categories.

2 LITERATURE REVIEW

2.1 Conceptual Model

Based on field studies the following conceptual model has been proposed to explain the subject.

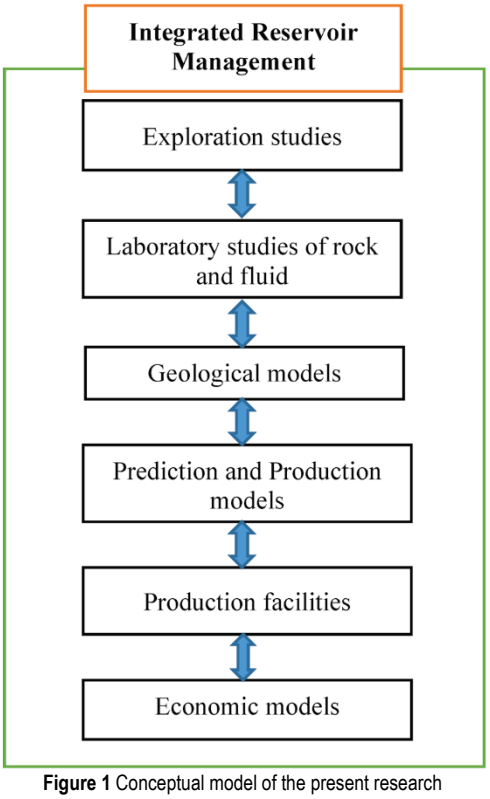


Figure 1 Conceptual model of the present research

2.2 Uncertainty

2.2.1 Engineering Data Uncertainty

Methods such as gas-injection, intermittent water and gas injection (WAG), smart water, and thermal and polymeric methods are commonly used to increase oil recovery. Injecting water into reservoirs is one of the oldest oil recovery methods that has been used for many years. Over time, this method has undergone many alterations and improvements, which has led to its use as the most popular oil recovery method in sandstone and carbonate reservoirs.

Establish a relationship between different types of oil production techniques that consider all factors on oil production for proper optimization is necessary. Uncertainties of Geological, fluid mobility, laboratory, and field heterogeneity are the engineering uncertainties that we will examine in the following.

2.2.2 Uncertainty of Geophysical and Geological Information

Two main types of uncertainty affect our confidence in the results from numerical models: parametric uncertainty and structural uncertainty. Parametric uncertainty arises because of incomplete knowledge of model parameters such as empirical quantities, defined constants, initial conditions, and boundary conditions. Structural uncertainty in models arises because of inaccurate treatment of dynamical, physical, and chemical processes, inexact numerical schemes, and inadequate resolutions. Uncertainty from geology is usually related to seismic data, which is classified as Structural uncertainty.

Seismic data used in the construction of a reservoir system are unclear. These uncertainties relate to data collection, analysis, and statistical explanation. Typically, uncertainties occur as a result of errors in data collection, conflicting explanations, error in converting depth data, error in preliminary explanation, and the wavelength map error that has to do with the crest of the reservoir.

Probably uncertainties are mostly geological. In geological information, uncertainties arise due to sedimentation, rock nature (lithology), rock extension region, and rock physical properties, which leads to the following uncertainties:

- of the reservoir's gross volume
- of the size and direction of sedimentation
- of difference in extension of rock type
- of porosity data
- of the net/gross ratio
- of contact between fluids.

These uncertainties have an impact on the assessment of on-site hydrocarbons and the movement of fluid across the reservoir.

2.2.3 Uncertainty of Dynamic Information

Petroleum reservoirs are very heterogeneous and it is difficult to predict the movement of fluid in them and always cause uncertainties that prevent proper and practical optimization. To reduce these uncertainties, various models and simulation methods should be used to perform the correct optimization to compensate for global oil demand. For example one of this method is grey bootstrap is proposed to resolve some problems about evaluation of the uncertainty in the process of dynamic. This method can evaluate the uncertainty without any prior information about probability distribution of random variables, separating trends with known and unknown law.

At this point, uncertainties of all variables that have an impact on the flow of a fluid through the reservoir are addressed. These variables include absolute, vertical-horizontal, and relative permeability (the measurement of a rock's ability, to transmit fluids), fault transmissibility, rate of injection, productivity index, pores, and skin (around a wellbore), capillary pressure curve, aquifers (water-bearing portions). Such uncertainties influence both the calculation of the reserve and the change of flow rates with time.

2.2.4 Uncertainty of PVT Information

PVT data are known to be the least uncertain data. Lack of certainty of PVT data affects the capacity of process units, hydrocarbon transportation, and marketing. Among the uncertainties in this category are as follows:

- Uncertainty of fluid tests
- Uncertainty of fluid structure
- Uncertainty of the calculation of PVT properties
- Uncertainty of interfacial tension.

2.2.5 Uncertainty of Field Performance Information

The well drilled in Pennsylvania was the first example of a well that showed the earth's layers, but with the drilling of other wells, a variety of different layers of the earth emerged, and engineers concluded that the earth's layer was different in each basin. Therefore, information of different layers and surface data should be used to minimize uncertainties

Furthermore, data on field performance might as well be swayed by the following uncertainties:

- The cost of oil production is usually calculated systematically and accurately; nevertheless, calculations of the water-oil ratio (WOR) and the gas-oil ratio (GOR) are occasionally performed;
- The rate of production fluctuation is normally evened out as it can occur at short durations; the rate of gas is not calculated correctly, especially if parts of it is burned;
- Injection information is less accurate than production information as a result of errors in the calculation stage, loss of fluid at a different time because of leakage in the skin or flow behind the piping system; and
- Pressures gauged at a certain phase of the flow analysis are generally less reliable than those acquired during shut-in.

2.2.6 Uncertainty of Economic Information

In all different parts of human life, the economy is the most effective factor. In the oil industry, drilling a well or

using an oil production method or using new technology is only used when it has an economic benefit. The main purpose of optimization is to reduce the risks that may impede the economic benefits of an operation. By accurately recognizing the uncertainties, the most optimal model for an economic operation can be used. For example, the use of nanoparticles to wettability Alteration and reduce sand production is a technique that has received much attention from researchers in the past few years, but since its economic uncertainty is very high, it has not been widely used in industry.

Production management is challenged by some uncertain risks of going up or down in oil prices. Conventional production enhancement approaches concentrate on net present value over time. The lack of reliability of long-term predictions is the key problem of many strategies. Given the time-dependent nature and the instability of oil prices, more often than not, it makes oil risky production [1, 2].

2.2.7 Uncertainty of Political Information

The uncertainty of the political system is a key feature affecting the local investment climate, which firms and entrepreneurs must consider when deciding to start, expand, or contract their businesses. Given the impact of oil on the economy and relations between countries and its key role in determining world powers, it is governed by very complex policies. Investors and entrepreneurs engaged in oil trade in different countries must act in accordance with that country's policy and also consider the impact of factors such as sanctions.

In order to predict the future of their business, these investors must pay close attention to the behavior of countries and world powers in order to avoid destructive global policies. Many countries have a major source of oil revenue, which, given the lifespan of their reservoirs, which are in the semi-finals, is forced to use techniques that reduce uncertainties and are able to meet demand. Therefore, their policy should be formulated in such a way as to give the leading companies investment security so that they do not face any problems.

2.2.8 Uncertainty of Environmental Information

Coping with the above uncertainty would be a serious factor in the production of the oil field. Being less sensitive to uncertainty along with the implementation of calculations are two methods that have been found to be exceedingly contradictory in some research. In this study, however, we present a small report of certain optimization techniques in the development of the petroleum field, based on the data on uncertainty to be able to both reduce uncertainty and sensitivity to its data.

In this study, we present an up-to-date analysis of optimization methods used to solve these problems by first analyzing the cause of uncertainty and then reviewing some practical optimization techniques against technical problems.

3 RESEARCH METHOD

The principles of robust optimization have been introduced primarily in the research papers of engineering design.

The first approach to deal with uncertainties is probably stochastic (linear) programming in the form of a risk factor to manage robustness, whereas robust optimization is known to be more useful in engineering fields, according to the study by Mulvey and Bai [3, 4].

The person who discussed the great implications regarding Robust Optimization in Engineering [5] was Taguchi, who is well recognized for developing a leading design strategy and has gained a lot of interest in the last few years.

Risk management should be addressed, too, when we deal with oil project investments. The recovery of oil is severely affected by geological, financial, and technical risks of exploration & production operations. The major elements of risk reduction [6] are the collection of relevant data and flexibility.

As a result of risk quality dynamics and the amount of risk exposure (RE) that threaten risk management's efficiency, it has been recommended that dynamic risk assessment in oil production and robust optimization programs be examined [7].

Van Essen et al. (2009) introduced the Robust Optimization (RO) approach to minimize the risk of geological uncertainties inherent in the development stage of the oil field, by implementing a series of discoveries that explain a range of possible geological systems to address geological uncertainty data [8].

NPV included a single objective with fixed oil pricing, and it was the related objective function. They also used a standard gradient-based optimization strategy wherein they access the gradients through an adjoin formulation.

Alhuthali et al. (2010) evened up the time of arrival of waterfronts across all producers with the aid of several geological realizations and implemented two optimization parameters of expected value and standard deviation, in linear form with a risk aversion coefficient; actually, they have employed the gradient and the analytical form of Hessian calculation of the objective function [9].

Almeida et al. (2010) attempted, under technological and geological uncertainties, to generate a pro-active approach and specify a project using a genetic algorithm to optimize the single objective NPV [10].

Chen and Hoo (2012) present a link between the Markov chain Monte Carlo (MCMC) and the Kalman filter ensemble (EnKF) in trying to collect changes of certain variables to monitor the amount of water pumped to a reservoir entitled the water-flooding program. They accomplished this by employing an efficient model-based system involving the uncertain parameter changes and a specific low-order model developed from a first principle model [11].

Oil production increased (by 9.0 percent and 8.2 percent with EnKF and MCMC adjustments, respectively), and water production decreased in the final total net present value in the parameter update model. The findings also indicated that

maximizing the reservoir's oil production had an effect on the amount of water added to drive the contained oil out, so it is of utmost importance to change the uncertain geological variables (porosity and permeability) to optimize the reservoir's oil production.

In water flooding modeling, Capolei, et al. (2013) also included an open-loop modeling scenario with no input and a closed-loop modeling scenario with geological uncertainty. To bolster the RO technique, they developed an updated robust proposed methodology (modified RO) with larger profits and less risk. The gains were calculated according to the predicted NPV, while the risk was calculated according to the normal NPV deviation [12]. Yasari, et al. (2013) put forth an interesting theory to minimize uncertainty sensitivity while no measurement data were expected to be available [13]. As such, by using a derivative-free Evolutionary Multi-objective Optimization (EMO) technique in the context of an updated Non-dominated Sorting Genetic Algorithm (NSGA), known as NSGA-Il, they established the robust optimization technique to generate several Pareto-optimal alternatives without theoretical deduction of the dynamic reservoir systems. And in 2015, in an effort to obtain the optimal - yet robust - waterflood strategies, they offered multi-objective optimization formulations. Two multi-objective, Pareto-based robust optimization models have been tested to overcome the permeability uncertainties.

The test studies showed that the proposed approach delivered better performance in providing a robust optimal alternative(s based on Pareto (injection policies) against permeability uncertainties that were accurate for the original group of realizations [14].

In contrast with the alternative equivalence of certainty and robust optimization techniques, the mean-variance parameter's potential to help minimize the significant inherent geological uncertainties has been suggested for production optimization.

Through their study, it became clear that maximizing certainty equivalence and robust optimization remain to be risky solutions. Still, the efficiency of mean-variance optimization in risk management and reducing the degree of uncertainty in optimizing efficiency is quite remarkable.

Siraj, et al. (2015) suggested a multi-objective optimization question that takes into account financial and model uncertainties to subsidize the adverse effects, that is, the risk of these uncertainties on production output [15].

Without significantly sacrificing the main priority of economic life-cycle efficiency, they established improved robustness. In order to describe the financial and geological uncertainty domain, they also provided a set of different oil price possibilities and geological system realizations. An average NPV among these groups is the main priority.

Their second goal was to optimize the pace of oil production to minimize risk since the risk of uncertainty grows with time. The multi-objective model was applied separately in a dynamic or lexicographic fashion for both types of uncertainties.

Geological uncertainty greatly affects the optimum well placement strategy and has to be considered in the question of optimization of well placement. A geological realization control mechanism for well placement against geological uncertainty was established by Rahim and Li (2015) [16].

Hanssen and Foss (2015) framed the question of optimization as a two-level stochastic programming question, and the outcome was a technique to run the wells, rather than a single set point acquired by the deterministic problem. The principles of risk theory are super beneficial as a result of high degrees of uncertainty in model-based financial modeling of the water-flooding mechanism in oil reserves. They proposed an inverted risk management system in another study to optimize the lower tail (worst instances) of the distribution of the economic objective function but without seriously sacrificing the upper tail (best instances). Within geological uncertainty, they found the worst robust optimization scenario and Conditional Value-at-Risk (CVaR) measure to optimize the worst problem(s). Also, a deviation method of semi-variance was included in geological and economical uncertainty defined by a set of geological system realizations and a set of different oil price scenarios to optimize the worst cases [17, 19].

Foroud, et al., and Siraj, et al. (2016, 2017) stressed the geological system as a primary cause of uncertainty in the simulation of petroleum reservoirs that can lower the reliability of optimization process outcomes of simulation. The clustering algorithms such as the Kernel K-means Method (KKM) were suggested to pick a generic subset of geological systems and reduce the overall calculation cost during the process of simulation.

The strategy of some researchers is to control uncertainty in the field design. They proposed new methodologies to quantify and minimize risks based on an efficient production plan and decision-making procedure. It is called risk management. In the construction of elaborate petroleum fields, Santos, et al. (2017), for instance, considered the robust risk assessment strategy by introducing resilience to the production mechanism and developing a robust production plan.

The proposed method is based on the performance evaluation of all possibilities of an algorithmically optimized production plan, which aims to further evolve the optimization procedure and minimize risk. Multi-Attribute Utility Theory (MAUT) through multiple objectives (technological and financial indicators) is the essence of this system [22]. They recommended systematic, objective methods in subsequent work to quantify the expected value of flexibility (EVF). In the production process, this approach applies to complex reservoirs with several uncertainties shaping the selection of the production strategy [23].

Silva, et al. proposed (2017) a five-stage approach to estimate the importance of flexibility under exogenous and endogenous uncertainties in oil production operations, and each stage is split into certain secondary stages to identify the specifics of problem analysis and strategy development [24].

Measuring the uncertainties of reservoir aquifer response by conducting a complete simulation of fluids flow on a wide range of models also assumes prohibitive intractable calculation of costs and time. Some methods suggest an estimated solution (flow proxies) to address this challenge [25] or organize the realizations inside a multidimensional sphere depending on the flow results received through an estimated (computationally cheaper) design [26].

In order to measure the uncertainties for a broader class of variables, Bardy 2019 employed both methods and

combined the complex performance of the entire group of models [27].

Olalotiti-Lawal provided (2018) a novel technique for calibration of the subsurface system and quantification of uncertainty through Markov chain Monte Carlo (MCMC), wherein proper mixing is improved by contact among parallel Markov chains. This approach substantially increases the convergence of the sampling performance without loss [2].

In 2018, Zanbouri and Salahshoor proposed a novel robust modeling approach to specify a group of robust surrogate systems with unorganized uncertainty for economical performance estimation of an uncertain petroleum reservoir during the water flooding phase, based on geological uncertainty as a serious hurdle in the development of petroleum fields. In this process, the MIMO surrogate system combined with the desired nonlinear NPV objective function was recognized to produce a new updated, robust surrogate system in a configuration form of multi-input single-output (MISO) and allow direct economic quality evaluation calculation [5].

The robust optimization method was developed by Mudhafar et al. in 2018 to evaluate the optimal intervals of gas injection, soaking, and oil processing in diverse reservoirs within geological uncertainties [9].

In his analysis, the robust optimization method within geological uncertainties showed higher recovery of oil and NPV than nominal realization optimization, providing the decision-maker with a degree of freedom to substantially reduce the plan's risk.

The redevelopment of Brownfield is a highly valued answer to manage the drop in production and to optimally position the infill well to optimize recovery and reduce operating costs given the unstable climate of oil price [11].

A new procedure for robust and efficient well placement optimization within geological uncertainty was suggested by Hutahaean 2019. Multi-objective aided background matching, Bayesian posterior estimation, and well positioning optimization were incorporated into the multi-objective environment through various geological systems in their conceptual workflow. The proposed workflow provides robust and reliable optimal decisions in placing the infill well over multiple history match models [12].

In oilfield production and reservoir operations, well placement efficiency is a serious hurdle because reservoir asymmetries generate deeply non-smooth, discontinuity, non-convex cost functions comprising several local optimums. It is also important to run a massive number of simulations on the reservoir.

Several optimizing strategies can be categorized into two classes of approaches to assess the location of the well: gradient-based ex, and derivative-free ex approaches [23, 24]. Optimization techniques have recently been implemented to design and reduce the computational problem of well-placement optimization within uncertainty [27, 3].

In order to find an infill drilling scheme for vertical or horizontal well positioning optimization, Jesmani 2020 used the Simultaneous Perturbation Stochastic Approximation (SPSA) method, a local optimization technique, which would reduce computation significantly [13].

3.1 Propose an Optimal Model of Integrated Reservoir Management

Basically, there are five main factors in integrated reservoir management: well design and management, reservoir properties, reservoir modeling, surface facility design and economy. The first three cases are surveyed below. The integration of these three makes it possible to propose and design efficient and economical ways to enhance production from petroleum reservoirs (as schematically shown in Fig. 2).

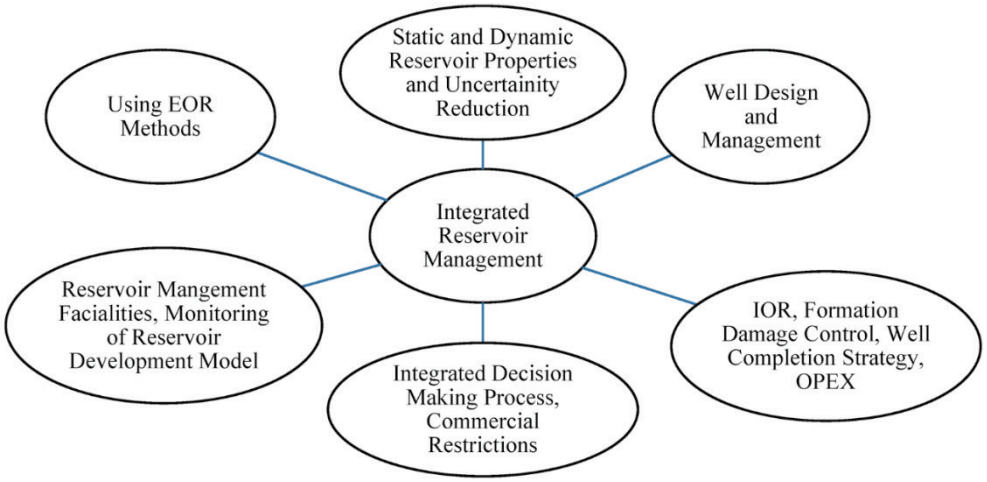


Figure 2 Integrated Reservoir Management

Reservoir properties use two main methods: structural modeling and stratigraphy modeling. In relation to structural modeling, interactive modeling software now helps to examine the compatibility between horizons and seismic faults and observations made in wells. In stratigraphic modeling, core information is used in an integrated manner to extract rock types based on geological and petrophysical criteria. In practice, a multivariate statistical experiment and analysis of well logs is performed, and the resulting cross-

plots are analyzed jointly with petrophysical core data to identify rock types.

The advantages of this approach are twofold: 1. Using the information available in all wells; 2. Calibration of geological facies in terms of information flow characteristics in several wells, Therefore, the identified rock types remain significant for both the sedimentologist and the reservoir engineer.

The purpose of describing the reservoir is to improve the geological modeling of the reservoir, thereby reducing subsequent uncertainties in the reservoir model and assigning dynamic properties to network blocks on a good scale to reduce uncertainties in production forecasting. These studies include performing the required laboratory tests (mainly measuring relative permeability and capillary pressure) in real reservoir conditions to observe fluid properties, wettability conditions, and saturation endpoints. Although much more complex and time consuming than conventional laboratory studies, SCAL results are much more reliable for calibrating the reservoir model. Special methods and equipment have been developed for this purpose. The role of flow units defined in the scale of the reservoir model can be easily related to the distribution of rock types in the exact geological model [27].

In order to adapt to different development plans, the repository simulator must be implemented with a number of options. Thus, the reservoir simulator must consider the mechanical effects of the rock. Optimization process to simulate hybrid effects such as gas injection and management Due to the complex well adaptation and modeling of heterogeneities, the simulator should be run with unstructured networking facilities. Finally, in order to be able to perform heavy calculations and use it to make quick decisions, the simulator must be able to run on parallel machines.

In addition, the integration of dynamic data greatly contributes to the reliability of the geological model for subsequent reservoir applications. This integration can be done in the early stages of field development using well test resources. Later, new dynamic information from the wells will allow the geological model to be updated. In practice, advanced mathematical methods are now available to model the geological model with good experimental results [19]. They include: Inversion techniques of simulated well experiments, such as the gradient method, to adjust the petrophysical properties. And a gradual deformation technique to adjust the geological model itself, the distribution of facies, reservoir boundaries and fault position. In the field of reservoir properties, there are complete software lines. This is the case for IFP with Reservoir Modeling Line (LMR).

4 CONCLUSION

Uncertainty is due to incomplete and imprecise knowledge as a result of limited sampling of the subsurface heterogeneities. In this study, optimization strategies in petroleum reservoir planning are introduced. It has been shown the output of each method independently. It can be seen that for reservoir management priority, many optimization studies have so far focused on production optimization. Mainly, the efficiency of the existing large petroleum fields causes us to consider the principles of reservoir management in order to increase EOR. There are several sources of uncertainty in the oil field development that misrepresent the future reservoir productivity. In general, uncertainty in information about reservoir management is divided into four categories. At the same time, many different ambiguities in the course of the developing oil field, including geological, operational and economic uncertainties, have a devastating effect on the effective production of this reservoir.

5 REFERENCES

[1] Schulze-Riegert, R. & Ghedan, S. (2007). Modern Techniques for History Matching.

[2] Christiansen, L. H., Capolei, A., & Jorgensen, J. B. (2016). Time-explicit methods for joint economical and geological risk mitigation in production optimization. *Journal of Petroleum Science and Engineering, 146*, 158-169. https://doi.org/10.1016/j.petrol.2016.04.018

[3] Christiansen, L. H., Hørsholt, S., & Jørgensen, J. B. (2018). A Least Squares Method for Ensemble-based Multi-objective Oil Production Optimization. *IFAC, 51*(8), 7-12. https://doi.org/10.1016/j.ifacol.2018.06.347

[4] Bai, D., Carpenter, T., & Mulvey, J. (1997). Making a Case for Robust Optimization Models. *Management Science, 43*, 895-907. https://doi.org/10.1287/mnsc.43.7.895

[5] Mulvey, J., Vanderbei, R., & Zenios, S. (1995). Robust optimization of large-scale systems. *Operation. Research, 43*, 264-281. https://doi.org/10.1287/opre.43.2.264

[6] Beyer, H.-G. & Send-off, B. (2007). Robust optimization – A comprehensive survey. *Computer Methods in Applied Mechanics and Engineering, 196*(33), 3190-3218. https://doi.org/10.1016/j.cma.2007.03.003

[7] Hayashi, S. H. D., Ligero, E. L., & Schiozer, D. J. (2010). Risk mitigation in petroleum field development by modular implantation. *Journal of Petroleum Science and Engineering, 75*(1), 105-113. https://doi.org/10.1016/j.petrol.2010.10.013

[8] Xie, G. (2010). Dynamic risk management in petroleum project investment based on a variable precision rough set model. *Technological Forecasting and Social Change, 77*(6), 891-901. https://doi.org/10.1016/j.techfore.2010.01.013

[9] Van Essen, G. M. (2009). Robust water flooding optimization of multiple geological scenarios. *SPE Journal, 14*(1), 202-210. https://doi.org/10.2118/102913-PA

[10] Authalic, A. H. (2010). Optimizing smart well controls under geologic uncertainty. *Journal of Petroleum Science and Engineering, 73*(1), 107-121. https://doi.org/10.1016/j.petrol.2010.05.012

[11] Almeida, L. F., Vellasco, M. M. B. R., & Pacheco, M. A. C. (2010). Optimization system for valve control in intelligent wells under uncertainties. *Journal of Petroleum Science and Engineering, 73*(1), 129-140. https://doi.org/10.1016/j.petrol.2010.05.013

[12] Chen, Y. & Hoo, K. A. (2020). Model parameter uncertainty updates to achieve optimal management of a reservoir. *Control Engineering Practice, 20*(10), 1042-1057. https://doi.org/10.1016/j.conengprac.2012.01.007

[13] Capolei, A. (2013). Waterflooding optimization in uncertain geological scenarios. *Computational Geosciences, 17*(6), 991-1013. https://doi.org/10.1007/s10596-013-9371-1

[14] Yasari, E. (2021). Application of multi-criterion robust optimization in water-flooding of oil reservoir. *Journal of Petroleum Science and Engineering, 109*, 1-11. https://doi.org/10.1016/j.petrol.2013.07.008
[15] Yasari, E. & Pishvaie, M. R. (2015). Pareto-based robust optimization of water-flooding using multiple realizations. *Journal of Petroleum Science and Engineering, 132*, 18-27. https://doi.org/10.1016/j.petrol.2015.04.038
[16] Siraj, M. M., Van den Hof, P. M. J., & Jansen, J. D. (2015). Handling risk of uncertainty in model-based production optimization: a robust hierarchical approach. *IFAC-PapersOnLine, 48*(6), 248-253. https://doi.org/10.1016/j.ifacol.2015.08.039
[17] Rahim, S. & Li, Z. (2015). Well Placement Optimization with Geological Uncertainty Reduction. *IFAC-PapersOnLine, 48*(8), 57-62. https://doi.org/10.1016/j.ifacol.2015.08.157
[18] Hanssen, K. G. & Foss, B. (2015). Production Optimization under Uncertainty - Applied to Petroleum Production. *IFAC-PapersOnLine, 48*(8), 217-222. https://doi.org/10.1016/j.ifacol.2015.08.184
[19] Siraj, M. M., Van den Hof, P. M. J., & Jansen, J. D. (2016). Robust optimization of water-flooding in oil reservoirs using risk management tools. *IFAC-PapersOnLine, 49*(7), 133-138. https://doi.org/10.1016/j.ifacol.2016.07.229
[20] Siraj, M. M. (2018). Scenario-based robust optimization of water flooding in oil reservoirs enjoys probabilistic guarantees. *IFAC-PapersOnLine, 51*(8), 102-107. https://doi.org/10.1016/j.ifacol.2018.06.362
[21] Foroud, T., Seifi, A., & AminShahidy, B. (2016). An efficient optimization process for hydrocarbon production in presence of geological uncertainty using a clustering method: A case study on Brugge field. *Journal of Natural Gas Science and Engineering, 32*, 476-490. https://doi.org/10.1016/j.jngse.2016.04.059
[22] Siraj, M. M., Van den Hof, P. M. J., & Jansen, J. D. (2017). An adaptive robust optimization scheme for water-flooding optimization in oil reservoirs using residual analysis. *IFAC-PapersOnLine, 50*(1), 11275-11280. https://doi.org/10.1016/j.ifacol.2017.08.1632
[23] Santos, S. M. G., Gaspar, A. T. F. S., & Schiozer, D. J. (2017). Risk management in petroleum development projects: Technical and economic indicators to define a robust production strategy. *Journal of Petroleum Science and Engineering, 151*, 116-127. https://doi.org/10.1016/j.petrol.2017.01.035
[24] Santos, S. M. G., Gaspar, A. T. F. S., & Schiozer, D. J. (2018). Managing reservoir uncertainty in petroleum field development: Defining a flexible production strategy from a set of rigid candidate strategies. *Journal of Petroleum Science and Engineering, 171*, 516-528. https://doi.org/10.1016/j.petrol.2018.07.048
[25] de Oliveira Silva, M. I., de Santos, A. A., Schiozer, D. J., & de Neufville, R. (2017). Methodology to estimate the value of flexibility under endogenous and exogenous uncertainties. Journal of Petroleum Science and Engineering, 151, 235-247. https://doi.org/10.1016/j.petrol.2016.12.026
[26] Bardy, G., Biver, P. G. Caumon, P., Renard, V., Corpel, V., & King, P. R. (2014). Proxy comparison for sorting models and assessing uncertainty on oil recovery profiles. *ECMOR XIV - 14th European Conference on the Mathematics of Oil Recovery*. EAGE. https://doi.org/10.3997/2214-4609.20141901
[27] Josset, L. & Lunati, I. (2013). Local and Global Error Models to Improve Uncertainty Quantification. *Math Geosci, 45*, 601-620. https://doi.org/10.1007/s11004-013-9471-4

Authors' contacts:

Amir Naser Akhavan, Assistant Professor
(Corresponding author)
Management, Technology, and Science Departments,
AmirKabir Technology University Tehran,
No. 350, Hafez Ave, Valiasr Square,
1591634311 Tehran, Iran
E-mail: akhavan@aut.ac.ir

Seyed Emad Hosseini
Management, Technology, and Science Departments,
AmirKabir Technology University Tehran,
No. 350, Hafez Ave, Valiasr Square,
1591634311 Tehran, Iran
E-mail: seyedemadh@yahoo.com

Mohsen Bahrami
Mechanical Engineering Solid Design Department,
AmirKabir University of Technology Tehran,
No. 350, Hafez Ave, Valiasr Square,
1591634311 Tehran, Iran
E-mail: mbahrami@aut.ac.ir

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220714131724

Original scientific paper
Received: 2022-07-14, Accepted: 2022-12-12

Prototyping and Integration of Educational Low-Cost Mobile Robot Platform

Petar Piljek*, Denis Kotarski, Alen Šćuric, Tomislav Petanjek

Abstract: This paper describes the process of designing and prototyping a low-cost robotic platform based on existing equipment and projects that enable extracurricular STEM activities in Croatia and beyond. A robotic platform with a differential drive configuration was chosen for education from an early age due to its simplicity and a wide range of cheap and compatible components from which it can be made. From the aspect of integration into extracurricular or curricular activities, the BBC micro:bit ecosystem was considered, enabling block-based visual programming. Components with printable parts make up the assembly of the educational robot. The main steps in designing and creating a robot prototype are presented, which consist of the modelling, 3D printing of robot parts, and assembly into a functional system. After several stages of testing, an interactive workshop was held with 7th-grade primary school pupils. Further work is planned to create educational material for extracurricular STEM workshops.

Keywords: 3D printing; BBC micro:bit; differential drive; Scratch; STEM education

1 INTRODUCTION

Education in the STEM field is quite diverse, which is made possible by the availability of a variety of components, technologies, and ultimately curricula that include practical work. 3D printers are very popular and are found in a wide range of applications, which is possible by creating a wide range of parts that can be further used in assemblies. The title of the paper [1] asks the question: What to build next, given in mind that the era of affordable 3D printing is underway. An interesting area is robotics which can be integrated into different levels of education with different outcomes. In paper [2], a study on learning and the problem-solving process among junior high school pupils through participation in robotics projects using Lego Mindstorm is presented.

The use of robotics education in Croatia in primary school is carried out mainly through extracurricular activities. The IRIM Association has launched various programs through the Croatian Makers League and other projects [3], which have further popularized the STEM area. One of the projects involved equipping libraries with BBC micro:bit sets of educational boards and 3D printers. This allows the integration of 3D printing with the BBC micro:bit unit, which is especially interesting from the aspect of robotics. In addition to the existing infrastructure, it is possible to further expand it with smaller investments, given that a low-cost system is being considered. Compared to existing robots available on the market such as mBot, keeping in mind a set of robots, for the same price range, it is possible to purchase a 3D printer and create your own robotic platforms. BBC micro:bit boards are versatile and can be used for various purposes such as mobile robots or robotic arms. Paper [4] presents a study that explores pupils' experience when designing programmed technological solutions using a BBC micro:bit board and identifies the technological knowledge in order to successfully solve a real-world task. Furthermore, in [5] study investigates ways of experiencing the process of solving a real-world task with programming material for pupils aged 10 and 14.

Differential drive robot configuration is very interesting from the aspect of education at primary school levels. The principle of operation is intuitive because the movement depends on the angular velocities of the drive motors. An example of a task can be to define the functions that allow for different movements, such as forward, backwards or turning in place. The Croatian Makers League used mBot robots, and the effectiveness of their usage to increase the basic knowledge in programming and robotics for pupils of age 13 is shown in [6]. Such robots can be made as low-cost platforms and can be integrated with existing BBC micro:bit sets. In addition, this type of robot is used at higher levels of education such as fuzzy control of a mobile robot in case of obstacle avoidance [7], using the Arduino eco-system. Arduino-based control is widespread and is particularly interesting in the field of education given its versatility [8], low component cost, and compatibility with various software packages such as MATLAB [9]. As mentioned, interesting platforms are also robotic arms, which in low-cost versions mainly consist of drive components in the form of servo motors [10].

Educational robots can usually be purchased in kits that need to be assembled, or they can be designed and developed mainly by rapid prototyping technologies, including 3D printing. The process of 3D printing (additive manufacturing) is preceded by the process of designing parts. Various tools can be used for 3D modelling, depending on the level of education, from primary school level to engineering education, some of which are presented in [11]. 3D tools used in teaching and extracurricular activities in the field of STEM education are mostly free. Preparation for manufacturing is carried out in so-called slicers and further production is realized by open-source printers [12]. The paper [13] presents a study on approaches to design and 3D printing in teacher education.

In this paper, potentials in the field of STEM education based on the existing infrastructure in Croatia are discussed. The concept of the educational platform has the following goals: do it yourself, take off the shelve components, low cost, and easy to repair and maintain. The components, their possibilities within such a system, and their advantages in the

wider integration of education in the field of robotics are presented. In the robot design phase, it is necessary to construct the parts of the system that, together with the components, form a functional system. For constructed parts, the process of prototyping parts with fused deposition modelling (FDM) technology is presented. Then, the parts and components of the educational mobile robot with differential drive were assembled. A workshop was conducted with pupils up to 7th grade, using a presented platform. Through the workshop, the basic movements of robots were tested, and pupils were introduced to the principles of operation. Also, the reactions of pupils gave the impression that, as expected, there is an increased interest in the STEM area.

2 EDUCATIONAL ROBOTICS IN CROATIA – CURRENT STATE AND POTENTIALS

The education process in the Republic of Croatia consists of an eight-year elementary school level, then secondary school level whose length varies by qualifications, and higher education which is divided into university level and professional studies. Regarding engineering education in the field of mechatronics and robotics, there are a large number of institutions in Croatia that provide such education, such as universities with related technical faculties and universities of applied sciences. At such institutions, some studies consist of theoretical and practical knowledge in the field of robotics through a series of courses, many of which are interdisciplinary. This includes knowledge of mechanical engineering, electrical engineering, computer science, and other branches. It should be noted that engineering education has been established in Croatia. Robotics has been studied more intensively in secondary school level education since the early 2000s, when a new profession, mechatronics technicians were introduced. In gymnasium programs, programming is more common and can be applied to robots. As for extracurricular activities, they are provided through centres for technical culture and various workshops. Commonly, robots are available as kits or didactic sets, which have become more and more advanced and diverse over the years. The emergence of robotics at the elementary school level has so far been mainly carried out through extracurricular activities. There were various workshops, summer schools, and leagues such as the "Liga kumpanija" which took place from the beginning of 2010 and the most widespread Croatian Makers League, which continues today.

In general, as far as STEM in Croatia is concerned, it can be freely said that the IRIM Association launched the STEM revolution in Croatia, but also in the surrounding countries. Various projects made extracurricular activities for various levels of education and lifelong learning possible. The Junior Engineer Academy project distributed equipment such as 3D printers and electronics kits and organized mentor training in the involved secondary schools. The goal is to enable the development of project ideas and to establish and maintain close contacts with companies and universities that provide students with an early insight into the world of engineers and qualified professionals in various fields of technology and science. In addition to secondary schools, the libraries involved in projects "Digital libraries for local development - DL4LD" and "STEM Revolution continues - Libraries" were equipped with 3D printers, and training for librarians was also organized. This project was preceded by the crowdfunding campaign "STEM revolution", which enabled the financing of the mass introduction of BBC micro:bit technology in Croatian schools, libraries, associations, and other institutions that work with children. From the aspect of robotics, since 2014, IRIM has launched the Croatian Makers League as the most massive extracurricular activity in elementary schools in the field of robotics and programming.

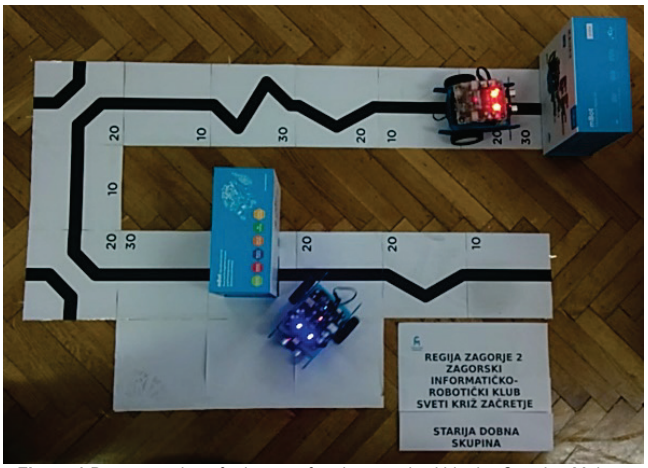


Figure 1 Demonstration of robots performing a task within the Croatian Makers League

The Croatian Makers League is part of the Croatian Makers project, the main goal of which is to enable the broad inclusion of robotics, automation, and programming in education at elementary school age. It is the largest competition of its kind in the EU with more than 12 000 children included per school year in more than 600 schools. The IRIM association donated more than 3000 robots, therefore Croatian Makers League represents IRIM's flagship project in robotics. The competition is divided into two categories: 1st-5th and 6th-8th grade of elementary school and takes place approximately four times a year. Institutions can be involved in the project by purchasing their equipment or by applying for a donation tender. The league was using the mBot platform until 2021 when they switched to another platform based on BBC micro:bit which offers an even easier entry into the world of robotics but also allows for more advanced usage. Fig. 1 shows the task to be performed by two mBot robots within the Croatian Makers League for the 2016/2017 school year. A common mission consists of a line-tracking task which is a typical problem that needs to be solved at that level of education.

From the aspect of understanding the principle of operation of the differential drive and understanding the operation of individual sensors, it is common to program robots to perform missions, such as those involved in education through the Croatian Makers League. Basic mission tasks are line tracking and obstacle avoidance. At the level of elementary school, in the first phase of education, the

basic functions required for the movement of robots are defined, namely forward, backward, turn right and turn left. After the initial phase of education which includes testing the robot's differential drive, various types of sensors are implemented in the system. For the basic functions of the robot required for education at the elementary school level, infrared line-tracking sensors and an ultrasonic distance sensor were considered. Such sensors come in low-cost variants and are used in many educational kits. In the preliminary design phase of a low-cost system, drive and sensor components and their compatibility with control ecosystems were considered. Fig. 2 shows the testing of components that include line tracking sensors and distance sensors.

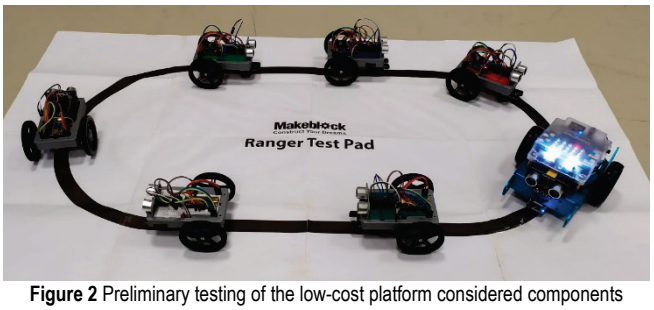


Figure 2 Preliminary testing of the low-cost platform considered components

3 LOW-COST MOBILE ROBOT PLATFORM

The mobile robot with a differential drive configuration is applied in a wide range of missions, which is especially interesting from the aspect of preparing pupils in the STEM field for future technicians, engineers, and scientists. The differential drive configuration of an unmanned ground vehicle (UGV) consists of two actuators (electric motor drives). Such a mobile robot has a simple principle of operation, therefore it can be applied from the earliest age to teach technical skills.

3.1 The Principle of Operation of the Differential Drive

Differential drive allows rotation in place (without translation), and the angle of rotation of the robot (heading angle) is determined by the differences in rotational speeds of the left and right wheels. This type of robot exists in two-dimensional space and has three degrees of freedom (DOF). The operating principle is defined with respect to the assumed two Cartesian coordinate systems (Fig. 3), the base coordinate system ($\mathcal{F}^{\mathrm{E}}$) and the mobile robot coordinate system ($\mathcal{F}^{\mathrm{B}}$). This type of drive configuration represents a non-holonomic mobile robot [15], given that the number of control variables is smaller than the number of robot DOFs, which are represented by the position (x, y) and orientation (ψ) of the robot. The translational v and rotational ω velocities of the robot are defined in $\mathcal{F}^{\mathrm{B}}$ and represent the motion of the robot in 2D space.

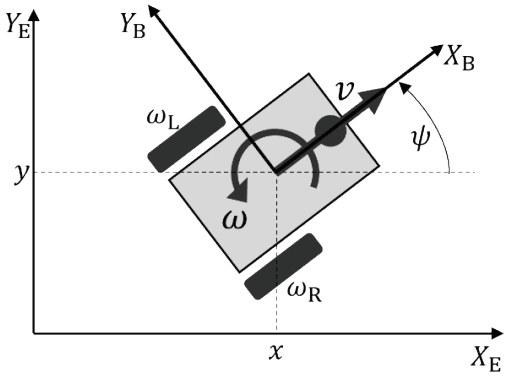


Figure 3 Schematic representation of a robot with a differential drive configuration

From the aspect of engineering education, robot kinematics is defined, where the robot speeds with respect to the base coordinate system are defined by the following expressions

$$\dot{x} = v\,\cos\psi,\quad \dot{y} = v\,\sin\psi,\quad \dot{\psi} = \omega. \tag{1}$$

The translational and rotational speeds of the robot depend on the angular velocities of the left and right wheels, i.e., the drive motor velocity $\boldsymbol{\Omega} = [\omega_{\mathrm{L}} \quad \omega_{\mathrm{R}}]^{\mathrm{T}}$, and the configuration parameters defined by the wheel diameter d and the distance between the drive wheels l. The robot velocity vector $\boldsymbol{v}^{\mathrm{B}} = [v \quad \omega]^{\mathrm{T}}$ is defined by the robot drive allocation matrix $\boldsymbol{\Gamma}_{\mathrm{R}}$ which maps the angular velocities of the motor to the robot velocity vector.

$$\boldsymbol{v}^{\mathrm{B}} = \boldsymbol{\Gamma}_{\mathrm{R}}\boldsymbol{\Omega} = \begin{bmatrix} \frac{d}{4} & \frac{d}{4} \\ -\frac{d}{2l} & \frac{d}{2l} \end{bmatrix} \begin{bmatrix} \omega_{\mathrm{L}} \\ \omega_{\mathrm{R}} \end{bmatrix} \tag{2}$$

The above Eq. (2) is a problem of direct kinematics which determines the translational and rotational speeds of robots based on the angular velocities of the wheels that represent the input variables of the model. The characteristic of a differential drive is that it has a smaller number of actuators than the number of DOFs, so a robot with this type of drive is a non-holonomic robot. The problem of inverse kinematics is the determination of the angular velocities of the left and right wheels based on the desired state of the robot, or the desired translational and rotational speed of the robot. From the aspect of motor control, there are different types of electric motor drives. Servo motors with continuous rotation were considered for the low-cost prototype.

3.2 System Components

From the aspect of education, the basic component of the educational robot is the control unit that is in charge of the system functioning. Practically speaking, it receives data from the sensor, interprets it and, based on the program, sends commands to the drive module. Since part of the infrastructure in Croatia has been secured through IRIM projects, and considering that since 2021 the most massive robot competition has been used by robots based on BBC

micro:bit boards, it is logical to consider this board as a key robot component. The great advantage of the BBC micro:bit board is the versatility from the aspect of programming where it can be used from elementary school levels to engineering education. It can be programmed in a block-based visual programming language (Scratch) and using Python or JavaScript programming language. One of the advantages is that it is cheap, it is very widespread, and there is a very large community. Furthermore, no software is required, it can be programmed in a browser. Besides, no large computing resources are required, just a PC or tablet with an installed compatible browser. On the hardware side, there are a lot of input/output pins, also, there are numerous additions and compatibility with other components. Fig. 4 shows a BBC micro:bit board with marked parts.

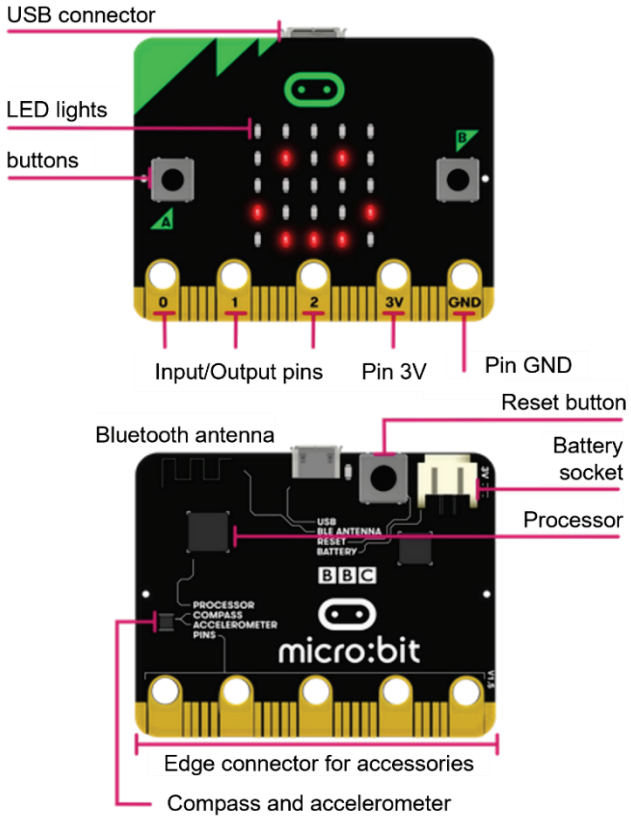


Figure 4 BBC micro:bit control unit

The success of the integration of BBC micro:bit boards as a robotic system has been presented to several commercial robots used primarily in education. An example of such a robot is cyber:bot [15]. If there is a need to involve a large number of pupils, for this price it is possible to provide a device for 3D printing and the necessary components for making robots, including BBC micro:bit boards. With a little time and imagination, robots can be made relatively easily. The advantages of such an approach are multiple. Through different phases, pupils get acquainted with electronics, 3D modelling, and 3D printing. In addition to being easy and cheap to build and repair, it is also easy to do system upgrades with new options. Such robots can also be combined with other control components like an Arduino board. The choice of components depends primarily on the level at which it is intended to be programmed, i.e., the programming language.

Two servo motors were used to drive the mobile robot. The advantage of this type of motor is compactness since the servo motor consists of a gearbox and feedback electronics. Such motors are controlled by a PWM signal, which is interesting because this signal modulation is used in a broader sense of robotics, from UGV robots to unmanned aerial vehicles (UAVs). Also, servo motors are often used in educational robotic arms. The servo motors with a price range of up to 10 $ were used in the different phases of system testing and they are shown in Tab. 1.

Table 1 Considered low-cost servo motors

Continuous servo	Size (mm)	Weight (grams)	Operating voltage (V)	RPM @ 6V
FEETECH FS90R	22,5×12,1×23,4	9	3-6	130
FEETECH FT90MR	23,2×12×25,5	12,5	3-6	100
FEETECH FS5103R	40,8×20,1×38	36	4,8-6	62

A basic sensor package was tested that includes IR sensors used in tasks involving line tracking and an ultrasonic distance sensor for obstacle avoidance tasks. The sensor package can be further expanded. The power supply of the mobile robot depends on the selected control, drive, and sensor components. These components do not necessarily operate at the same voltage levels. Different voltage levels can be achieved with buck/boost converters or with multiple batteries with required voltage levels. Fig. 5 shows the wiring diagram of a possible mobile robot, which was later tested.

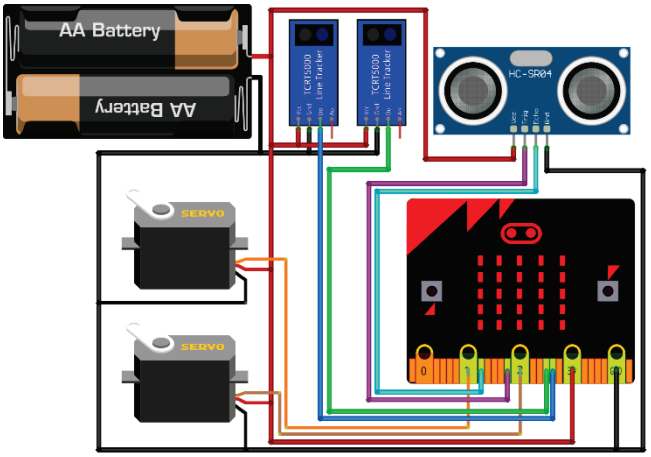


Figure 5 Schematic representation of the mobile robot electronic components connection

4 ROBOT PROTOTYPING AND INTEGRATION

4.1 Prototyping Procedure

The process of prototyping parts consists of the design and production phases, as shown in [16]. Although the same procedure is used at different levels of education, there are differences in approach, tools and learning outcomes. Additive manufacturing technologies enable an ever-widening range of research and applications in education, as

shown in the case of designing a modular educational multirotor UAV that can be used in engineering education [17]. The process of prototyping a low-cost educational mobile robot that can be used at different levels of education, primarily at the primary school, but also at the secondary school level, is considered and presented. The Prusa i3 MK3S 3D printer, which enables FDM printing, was used for parts manufacturing.

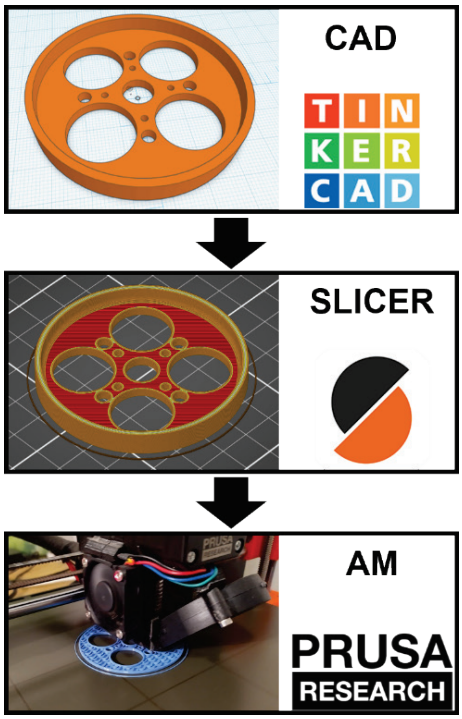


Figure 6 Schematic representation of the prototyping process

From idea to realization, it is necessary to take several steps that include 3D modelling, preparation of models for printing in the so-called slicer software, and part printing, as shown in Fig. 6. TinkerCAD software package for parts 3D modelling and assemblies was considered. Since the Prusa 3D printer was used to make the parts, the parts were prepared in the compatible Prusa Slicer software. Different print parameters can be set in the slicer, depending on the requirements of a particular part. Fig. 7 shows the robot wheel with different print parameters. The advantage of the FDM process is the low cost of materials compared to other technologies. Furthermore, in addition to the considered 3D printer, there are many different printer kits on the market, so the implementation of education can be even less expensive.

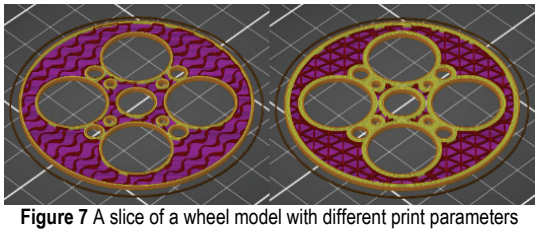


Figure 7 A slice of a wheel model with different print parameters

4.2 Experimental Testing and Integration of Robot Platform

After obtaining all the necessary components and making the final versions of the robot parts, the assembly of the robot can be started. The line tracking task at the primary school level of education can be achieved using several approaches, but in essence, it comes down to reading the left and right IR sensors which can have two output states, logic unit (white) or logic zero (black). By combining these two inputs, the movement of the robot is controlled. The basic set of robot movements consists of forward, backward, left turn, and right turn functions. Fig. 8 shows one of the possible implementations of the line tracking program.

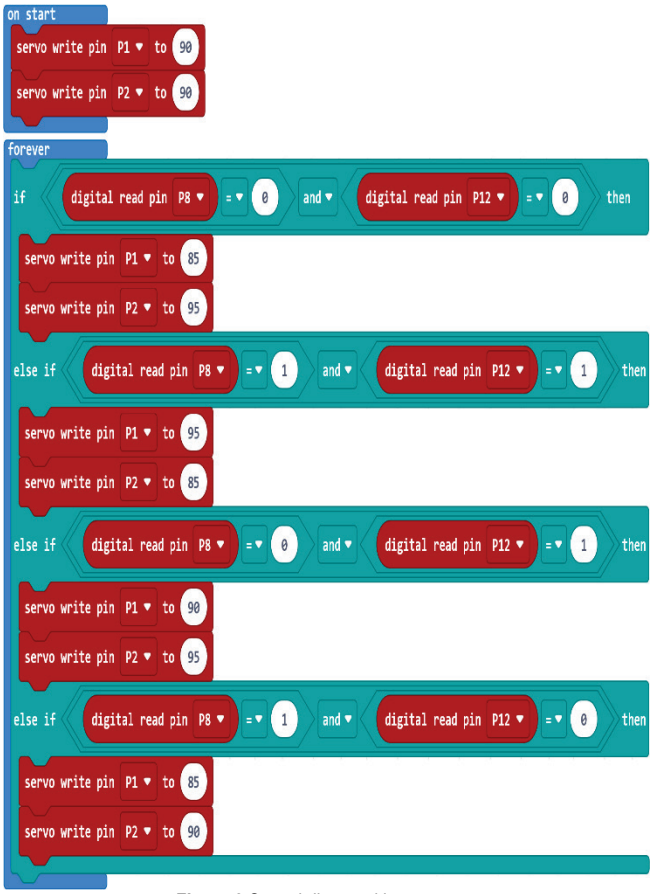


Figure 8 Scratch line tracking program

Figure 9 Experimental testing during a workshop

After successfully testing the robotic platform, a workshop was held as part of extracurricular activities at Mače Elementary School in collaboration with the Trivium STEM Edu association (Fig. 9). The participants of the workshop were 7th-grade pupils. Through the workshop, pupils were shown various tasks that can be implemented on the proposed low-cost platform. The integration of Bluetooth wireless control with another BBC micro:bit or with a phone was also tested.

5 CONCLUSION

In this paper procedure of educational low-cost robot platform design is presented. A mobile robot with a differential drive configuration was described in order to show the working principle and to define system components. In addition to the low cost of such a robot, it is important to emphasize several advantages of this design approach. Such a system can be used at different levels of education because it is easily upgradeable. Low-cost components can be used, but also more professional components can significantly expand the range of education. Furthermore, the considered drive components are also applied in a wide range, from mobile robots, and robotic arms to unmanned aerial vehicles. By conducting a series of experiments and workshops, it has been shown that such a system can be easily integrated into existing institutions and schools that have a 3D printer or plan to purchase it.

6 REFERENCES

[1] Eisenberg, M. (2013). 3D printing for children: What to build next, regarding the 3D printing?. *International Journal of Child-Computer Interaction, 1*(1), 7-13. https://doi.org/10.1016/j.ijcci.2012.08.004

[2] Barak, M. & Zadok, Y. (2009). Robotics projects and learning concepts in science, technology and problem solving. *International Journal of Technology and Design Education, 19,* 289-307. https://doi.org/10.1007/s10798-007-9043-3

[3] See https://croatianmakers.hr/hr/naslovnica/

[4] Cederqvist, AM. (2022). An exploratory study of technological knowledge when pupils are designing a programmed technological solution using BBC Micro:bit. *International Journal of Technology and Design Education, 32*, 355-381. https://doi.org/10.1007/s10798-020-09618-6

[5] Cederqvist, AM. (2021). Designing and coding with BBC micro:bit to solve a real-world task – a challenging movement between contexts. *Education and Information Technologies*. https://doi.org/10.1007/s10639-021-10865-w

[6] Lee, B. Y., Liew, L. H., Khan, M. Y. B. M. A., & Narawi, A. (2020). The Effectiveness of Using mBot to Increase the Interest and Basic Knowledge in Programming and Robotic among Children of Age 13. *Proceedings of the 2020 The 6th International Conference on E-Business and Applications*, 105-110. https://doi.org/10.1145/3387263.3387275

[7] Piljek, P. & Kotarski, D. (2020). Neizrazito upravljanje mobilnim robotom za slučaj izbjegavanja prepreka. *7th International Conference "Vallis Aurea", Focus on: Research & Innovation*, 509-518.

[8] Oltean, S. E. (2019). Mobile Robot Platform with Arduino Uno and Raspberry Pi for Autonomous Navigation. *Procedia Manufacturing, 32*, 572-577. https://doi.org/10.1016/j.promfg.2019.02.254

[9] Tejado, I., Serrano, J., Pérez, E., Torres, D., & Vinagre, B. M. (2016). Low-cost Hardware-in-the-loop Testbed of a Mobile Robot to Support Learning in Automatic Control and Robotics. *IFAC-PapersOnLine, 49*(6), 242-247. https://doi.org/10.1016/j.ifacol.2016.07.184

[10] Bhargava, A. & Kumar, A. (2017). Arduino controlled robotic arm. *International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 376-380, https://doi.org/10.1109/ICECA.2017.8212837

[11] Nemec, R. (2017). Using and Citation of 3D Modeling Software for 3D Printers. *International Journal of Education and Information Technologies, 11*, 160-170.

[12] Kostakis, V., Niaros, V., & Giotitsas, C. (2015). Open source 3D printing as a means of learning: An educational experiment in two high schools in Greece. *Telematics and Informatics, 32*(1), 118-128. https://doi.org/10.1016/j.tele.2014.05.001

[13] Verner, I. & Merksamer, A. (2015). Digital design and 3D printing in technology teacher education. *Procedia CIRP, 36*, 182-186. https://doi.org/10.1016/j.procir.2015.08.041

[14] See https://stemfinity.com/collections/micro-bit/products/cyber-bot-robot-kit-with-micro-bit

[15] Hassan Zarabadipour, H. & Yaghoubi, Z. (2019). Control of a non-holonomic mobile robot system with parametric uncertainty. *Technical Journal, 13*(1), 43-50. https://doi.org/10.31803/tg-20190116100550

[16] Piljek, P., Krznar, N., Krznar, M., & Kotarski, D. (2022). Framework for Design and Additive Manufacturing of Specialised Multirotor UAV Parts. In Răzvan Păcurar (Ed.), Trends and Opportunities of Rapid Prototyping Technologies [Working Title]. IntechOpen. https://doi.org/10.5772/intechopen.102781

[17] Kotarski, D., Piljek, P., Pranjic, M., Grlj, C. G., & Kasac, J. (2021). A Modular Multirotor Unmanned Aerial Vehicle Design Approach for Development of an Engineering Education Platform, *Sensors, 21*(8). https://doi.org/10.3390/s21082737

Authors' contacts:

Petar Piljek, PhD, Assistant Professor
(Corresponding author)
University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture,
Trivium STEM Edu association,
Ivana Lučića 5, 10002 Zagreb, Croatia
petar.piljek@fsb.hr

Denis Kotarski, PhD
Trivium STEM Edu association,
Ciglenica Zagorska 10, 49223 Sveti Križ Začretje, Croatia
denis.kotarski@gmail.com

Alen Šćuric
Trivium STEM Edu association,
Ciglenica Zagorska 10, 49223 Sveti Križ Začretje, Croatia
alenscuric9@gmail.com

Tomislav Petanjek
Trivium STEM Edu association,
Ciglenica Zagorska 10, 49223 Sveti Križ Začretje, Croatia
trivium.stem.edu@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221221115431

Original scientific paper
Received: 2022-12-21, Accepted: 2022-12-22

Dynamic Context Awareness of Universal Middleware based for IoT SNMP Service Platform

Hae-Jun Lee

Abstract: This study focused on the Universal Middleware design for the IoT (Internet of Things) service gateway for the implementation module of the convergence platform. Recently, IoT service gateway including convergence platform could be supported on dynamic module system that is required mounting and recognized intelligent status with the remote network protocol. These awareness concepts support the dynamic environment of the cross-platform distributed computing technology is supported by these idea as a Universal Middleware for network substitution. Distribution system commonly used in recent embedded systems include CORBA (Common Object Request Broker Architecture), RMI (Remote Method Invocation), DCE (Distributed Computing Environment) for dynamic service interface, and suggested implementations of a device object context. However, the aforementioned technologies do not support each standardization of application services, communication protocols, and data, but are also limited in supporting inter-system scalability. In particular, in order to configure an IoT service module, the system can be simplified, and an independent service module can be configured as long as it can support the standardization of modules based on hardware and software components. This paper proposed a design method for Universal Middleware that, by providing IoT modules and service gateways with scalability for configuring operating system configuration, may be utilized as an alternative. This design could be a standardized interface provisioning way for hardware and software components as convergence services, and providing a framework for system construction. Universal Middleware Framework could be presented and dynamic environment standardization module of network protocols, various application service modules such as JINI (Apache River), UPnP (Universal Plug & Play), SLP (Service Location Protocol) bundles that provide communication facilities, and persistence data module. In this IoT gateway, management for based Universal Middleware framework support and available for each management operation, application service component could be cross-executed over SNMP (Simple Network Management Protocol) version 1, version 2, and version 3. The way of SNMP extension service modules are conducted cross-support each module and independent system meta-information that could be built life cycle management component through the MIB (Management Information Base) information unit analysis. Therefore, the MIB role of relation with the Dispatcher applied to support multiple concurrent SNMP messages by receiving incoming messages and managing the transfer of PDU (Protocol Data Unit) between the RFC 1906 network in this study. Results of the study revealed utilizing Universal Middleware that dynamic situations of context objects with mechanisms and tools to publish information could be consisted of IoT to standardize module interfaces to external service clients as a convergence between hardware and software platforms.

Keywords: context awareness; dynamic java; dynamic network; IoT gateway; universal middleware

1 INTRODUCTION

The IoT industry's ongoing transition from service-centered to deliverable network requires the use of auxiliary devices with interface for application integration, execution, and management. The complexity of these costs, decentralized and distributed computing systems presents significant import for external modules. Network protocols are increasingly requiring communication modules to offer context awareness systems that are dynamically, residential, remote interface way, service-oriented, decentralized, energy-efficient, and customized. The dynamic context awareness system with such network management properties must be capable of adjusting its behavior at run-time in response to its perception of its environment and its own state in the unit of effective adaptation. Looking at an emerging trend in network Forensic, internet, and network services grows in the IoT industry of infrastructure, more and more inefficient networks are exploited for temporary purposes. The dynamic context awareness of network forensics has provided researchers insufficient enforcement systems with such as dynamic service platform.

Dynamic Context Awareness should be designed in service deliverable component architecture that is a distributed residential system from the local area network to the broadband network. In the dynamic network system of service awareness, considering distributed system that proposed an implementation model for dynamic run-time context object [1]. Universal Middleware models of CORBA (Common Object Request Broker Architecture), RMI (Remote Method Invocation), and DCE (Distributed Computing Environment), which are representative distributed computing technologies, are implemented by network infrastructure management conventionally based on interface modularization for standardized application service module [2].

While the benefits of convergence solutions have been partially acknowledged in prior research in relation to SAAS (software as a platform) computing system, and edge, more about context object awareness have also converged to another applicable service of an additional component by internal and external network environment [3]. IoT (Internet of Things) and network resources have been transferred to deliver from the service platform as a management protocol. However, not all the device interfaces initialized by such information have been tracking handled [4].

IoT research and standardization have been done for activities and technologies including UPnP (Universal Plug and Play) / DLNA (Digital Living Network Alliance), and PUCC (Peer-to-peer Universal Computing Consortium) that uses Universal Middleware. There has not been concern about networking in a distributed BCN, peer-to-peer/overlay networking management module to important things and add-on valued for a suitable platform for internal network among facility devices over heterogeneous service. Recently, peer-to-peer studies on the peer-to-peer metadata module, and a general middleware platform were researched and described [5]. These devices include desktop, home gateway, mobile phones, tablet computers, cloud servers, and others [5, 28, 29].

The implementable methods that support this model are representative of OSGi (Open Service Gateway initiative),

Android platform, and ASP.NET Core, and there is a difference in the method of constructing a cross-platform centered on the operating system [6]. The majority of cross-platform support for the real-time operating systems can be easy to establish a correlative networks. Universal Middleware provides a service platform for integration way of the deployment of modules and services. This service module can be dynamically loaded and consume services applications that take advantage of the dynamic updating components without the required add-on process. The activation services make the application context object of network module available in the activation services.

In particular, RMI provides a standard interface suitable for the convergence of IoT devices and reliability applications for remote diagnostic management between application services in context object classes supported by Java environments [7]. An application service module for the Facility Device was configured based on Universal Middleware for porting to the residential gateway [8]. Remote diagnostic and device application context specific situations were possible with some MIB (Management Information-Based) customization for SNMP. Explain the dynamic service model and MIB for SNMP customization scenarios that can be adopted. Commonly, The number of nodes that this platform can support or the capabilities of SNMP managers are frequently constrained, making it difficult to use advanced features for more complex networks with tens of thousands network nodes. This MIB database based on Universal Middleware design is a text MIB file that itemizes and describes all contexts on a particular device that can be queried or controlled using SNMP. Globally, each MIB item could be assigned an object identifier domain for SNMP.

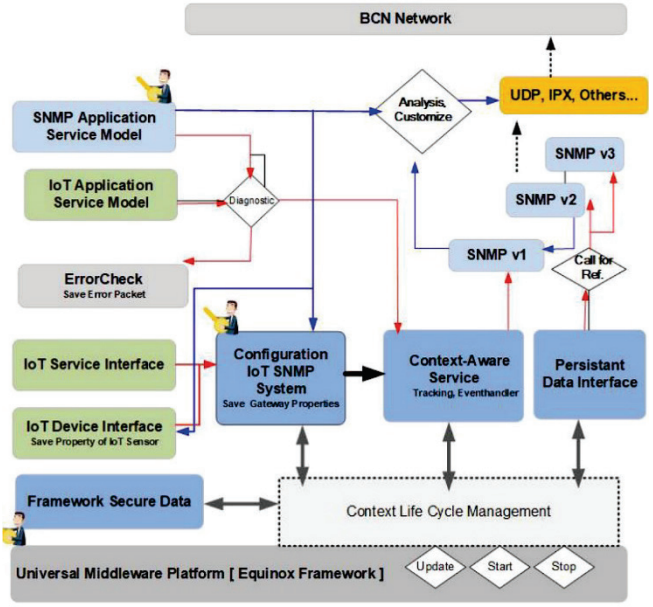


Figure 1 IoT SNMP service platform based on universal middleware

Practically, Eclipse's Equinox framework could be supported deliverables through the bundle context object that deployed standardize modules for transfer network, application components, and various data schema as a provider-centered system design [9]. It is possible to construct a mesh-scenario service between the communication module and the data module for configuring such an IoT environment, which is effective in systems whether it is a dynamic situation for devices or a dynamic situation on a platform.

Fig. 1 describe for IoT service platform depending on the device interface driver with standardized remote way diagnostics use case model. A common IoT application service model based on context bundles that depend on the framework of Universal Middleware. This diagnostic means to catch the SNMP Application Service model on UDP (User Datagram Protocol), IPX (Inter-network Packet Exchange), and other BCN (Broad Cast Network) areas. In addition, version 3 of the SNMP context also depends on the Security service context, as it offers enhanced security. As the SNMP context is used on another IoT-SNMP service platform, the Universal Middleware should be installed for the SNMP management process [10]. The above mentioned properties of the SNMP agent were recognized by this process. The generator is used dynamically for SNMP service implementations built upon this MIB.

In order to configure a scenario-based IoT module that is dynamically configured, concurrency must be provided on the Universal Middleware platform [11]. Until now, there have been studies on models and terminal system modules for implementing SNMP-based network management functions, but there are limitations in applying them to different versions of the operating environment. In this paper, existing methods with these limitations support multiple distributed structures and support them in the form of deliverable bundles, so it is possible to configure a scenario between the internal and the external BCN network of the IPX [12]. By implementing dynamic content based on a life cycle linked with SNMP, service attributes of a terminal may be stored and analysed in real-time. Chapter 2 shows the design of a Universal Middleware platform with dynamic context, and Chapter 3 shows that MIB information containing service attributes is tested through virtualization consoles of multiple devices to implement the proposed method and exchange it normally in a distributed environment.

2 DYNAMIC CONTEXT-AWARENESS OF UNIVERSAL MIDDLEWARE

2.1 Universal Middleware Platform for Context-Awareness

The name and modeling techniques for the contextual recognition procedure are configured differently for each application system. The structure of the contextual media constituting the system is generally composed of a modeling technique for the problem area and an inference technique suitable for the constructed model. In the existing situation recognition system, the basic configuration method consists of the development of middleware or intelligent situations, such as a module that interprets input information, a module that models the interpreted information and determines the situation, and a module that processes the defined situation

according to the application system. By applying Universal Middleware on a network basis suitable for IoT environments, this study designed existing internal services as a dynamic context module that can organically configure devices and service interfaces while maintaining independence from the infrastructure associated with the external network, BCN area. Therefore, even when the existing built system is maintained, the system burden is reduced for managers, and modules can be effectively added for users to dynamically use the service. [13]

Universal Middleware could be provided a system model that extends to the external network interface with tracking and event handler. The scalability of the communication network module, application service module, and data module for each standard module is controlled by these external systems, which also offer convergence services for device interfaces. The system implementation specifications of this model were applied based on the Java-based Equinox framework supported by the Eclipse development environment [14].

An IoT device or the network administrator provides an interface to the management device through an SNMP agent. Fig. 2 shows that virtual information is shared by MIB, a database of performance statistics, and parameters directly related to the current operation of the device, including dynamic management object context. In addition, SNMP provides notify, command, and response, a mechanism by which administrators can communicate with agents, and consists of a dictionary store. IoT service-type information for an object context could be used based on MIB. These MIB resources get the information of device interface context object between SNMP Protocol Service and SNMP Agent.

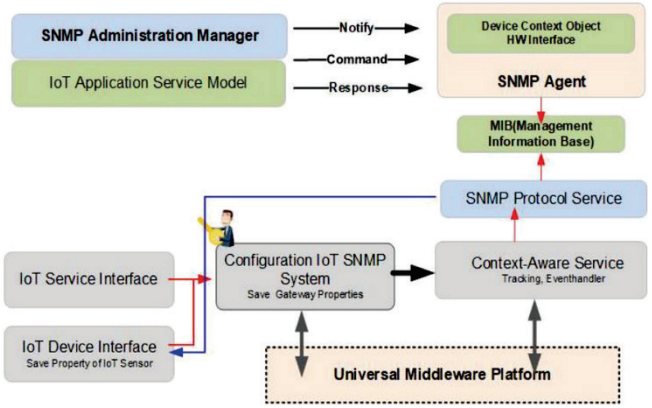


Figure 2 SNMP MIB information with service agent on universal middleware

Dynamic modules of objects for dynamic context awareness in the framework provide a device vendor-independent standards-based approach to Java Virtual Machine for CDC (Connected Device Configuration)-based applications and infrastructure [15]. Universal Middleware platform model communicates local area network and is distributed object system over the communication to provide a residential end-to-end architecture, and supports stable and evolving technical specifications for the open-source projects [16].

IoT service-type information on the object may be used based on the MIB information. Various manufacturing devices can be supported for RFC 1212, RFC 1213, RFC 1513, RFC 1757, and RFC 2021 remote monitoring as defined in the Internet RFC standard [17]. Thus, the private MIB for network components is available on the delivered context-object on the framework. The MIB resources can also be loaded and describes the entirety of all SNMP objects that are available in the network on context bundles, including addresses within this structure are set meta-data of the manufacturer [18]. It could be transferred to allow and distribution service consumers by the network manager, only the manufacturer number has to be registered.

2.2 IoT Convergence Service for SNMP

SNMP application of IoT service on the platform executes the commands issued by the agent that is to communicate messages and to store the MIB database. In this case, an SNMP service model designed according to the RFC 2571 method was suggested to process message contents differently for each version. Firstly, IoT Convergence Service for SNMP Service should be exchanged with handle messages to the network domain [19]. Secondly, the clustering module's convergence method offers security services that authenticate and encrypt messages. Thirdly, from the module to the others controls the access to distributed object context.

IoT SNMP Service declares separate convergence model for version 1 and version 2. Version 3 of SNMP implements that SNMP version 2 application bundles are designed according to RFC 2571. These platform models offer services that the network gateway's SNMP entity's convergence module is connected to [20].

Table 1 Management object context of MIB ID tree structure

Item	Format	Comment
Format	serviceNumber.identfierName.identfierName	Unique
Network Access Property	iso.org.dod.internet.private, console.1001.mySNMP.myUID.aaa.bbb	Last No. Service Instance Object
Configuration Property	Vector String <address>:<port>:<type>:where:	Trap Listener
	SNMP port(Primary/Secondary)	Platform Port
	Physical Name of SNMP Agent	System Location
	Domain Node Full-Name	Domain Name
	Contact Operator Name	System Operator
	Access Version 1, 2 or other future	Community Name

The IoT service domain should be to have a private identity that unambiguously identifies the entity the platform is running on the real-time execution environment. The converged module of platform is constructed dispatch including message processing, security, and access management. Thus, Tab. 1 describes IoT convergence service for SNMP contains a MIB tree control system with items that could be building scenario based. MIB ID should be assigned to context for identifying the object of the following format in Tab. 1.

IoT MIB service format is used by network managers to access the properties of the device type that the service represents that hold properties. To access the properties of a particular service uses the notifications to inform the manager that a predefined event has occurred and the generation of the agent assigns time-stamp fields, object ID, and variable [21].

2.3 Evaluation of Building for Dynamic Context-Awareness

Evaluation of Building for Dynamic Context Object is based on the installation and test environment of a framework that implements Equinox and commercial engine of Universal Middleware. Use the equation editor to show each equation. Thus, a Universal Middleware-based SNMP platform was established for application service bundles for dynamic context awareness experiments [22]. An application service bundling operation for object context applied to the module standard in the Equinox framework supported by the Eclipse 2021 IDE (Integrated Development Environment). For multiple supports of IoT service devices, a configuration management service bundle that supports SNMP version 1, version 2, and version 3 were configured on the platform [23].

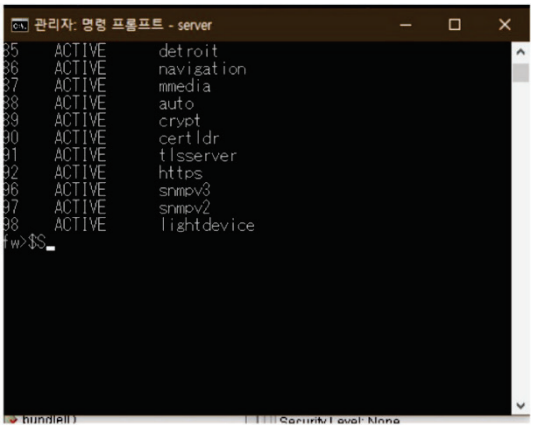


Figure 3 SNMP application service of equinox framework JDK 1.5

Fig. 3 shows the execution monitoring of IoT Service Platform Engine and SNMP services running on a Windows 10 compatible with Equinox framework based on JDK 1.5 JVM (Java Virtual Machine) as an active state [24]. Commonly, SNMP architecture supports a device service with a control module in each version of the architecture by adding DDS (Device Driver Subsystem) to the module for controlling different devices. Context-Aware SNMP apps send messages to the BCN network in XML (Extended Mark-up Language) format which removes the length restriction [25]. The first recognition experiment case is about an SNMP-based IoT convergence service. Fig. 4 shows the experiment and results of setting and changing MIB information after installation of SNMP version 2 and version 3 in a multi-support environment on a service platform that applies Universal Middleware.

Subsequently, Fig. 4 shows a process for setting and driving a service gateway according to the SNMP version on the console after driving for platform evaluation. This evaluation system is used for managing the SNMP MIB framework management network, and configuring service context through the object configuration service through SNMP. The role of the configuration module runs to all awareness of the context object mentioned above itself and should connect to the framework. Also, in order to run with the SNMP management service, there is a need to have connection information about SNMP [31]. In this application test, MIB information constructs a standard interface for lighting devices, an IoT service scenario, to confirm the applicability of multiple preferences for each version 2 of the application service. It is possible to set the authority for the person in charge of the operation, and domain name for reading, writing and notification [26].

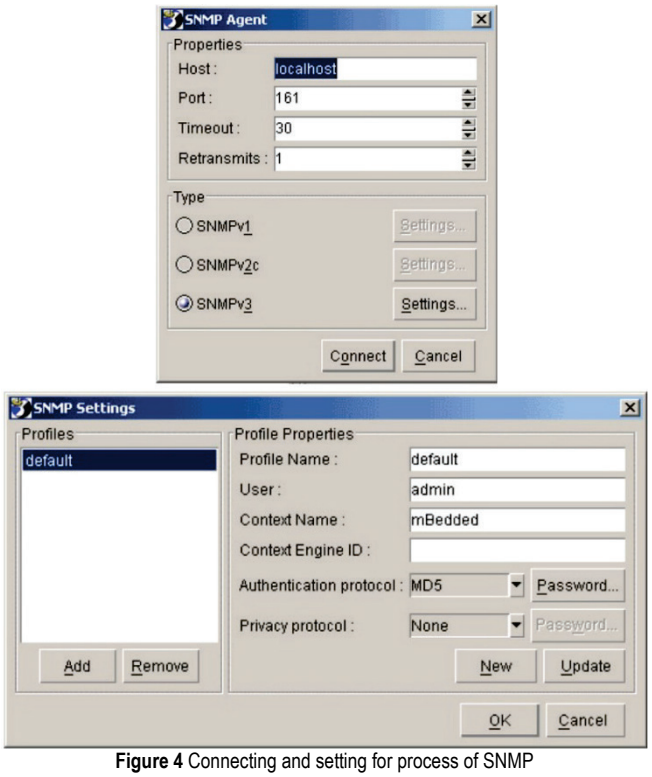


Figure 4 Connecting and setting for process of SNMP

Externally, security services could be extended in version 3 with Trap libraries to connect a multi-platform based on dynamic situations. SNMP agent uses trap notifications to inform the manager that a predefined event has occurred.

The second experiment was Fig. 5 shows MIB information. A number of IoT modules configured on the framework have been configured, and the MIB information of the lighting device for the SNMP service and the configuration levels are configured differently according to versions 2 and 3.

In addition, system Lister information, including access manager information, domain information, and port number, can be additionally set to dynamically expand the interface so that it can be easily configured when linking services between inside and outside the framework.

Fig. 6 shows MIB for Dynamic Query Execution of Context Awareness. The condition of dynamic information located for each function of the device indicated by the MIB is changed, and status information can be checked in real-

time through activity information. In addition, through an internet address and trap nodes information, it is possible to check the complex service status with notification functions including the failure status of the device.

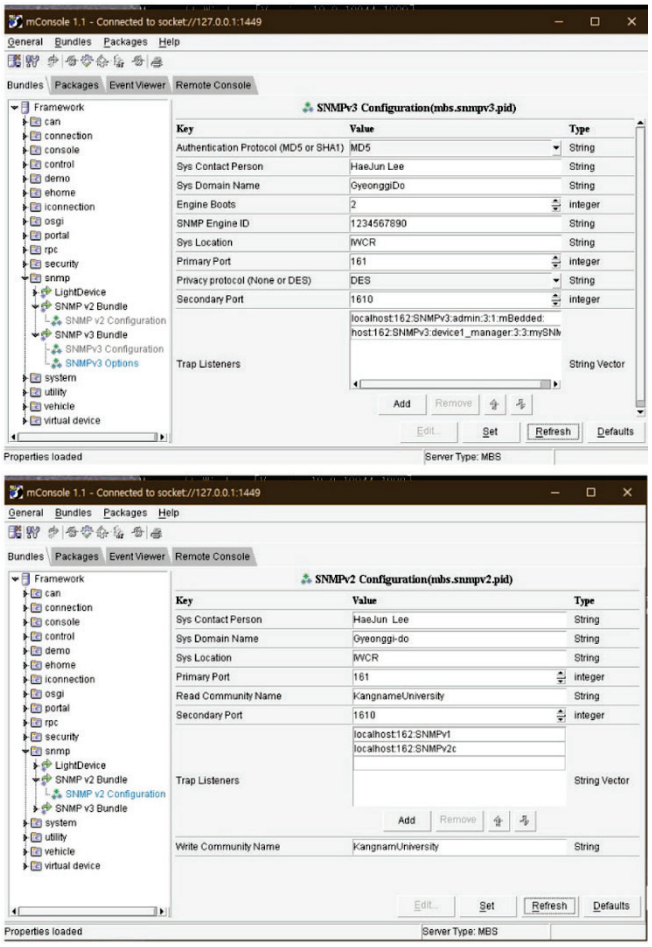


Figure 5 IoT SNMP multi-support of facility device interface

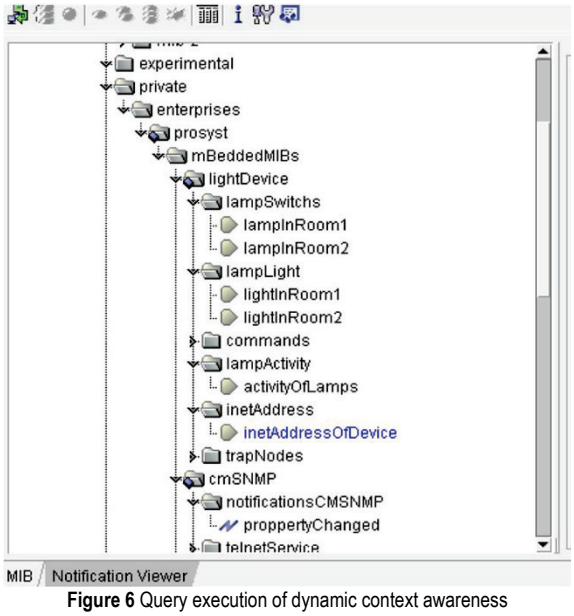


Figure 6 Query execution of dynamic context awareness

Fig. 7 shows one of result of property information about a device address with the object type. Supplementary, IP information, access information of the service, state, and additional system details may be confirmed.

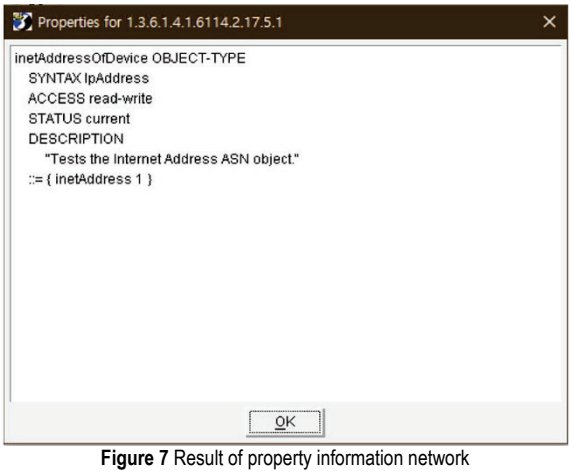


Figure 7 Result of property information network

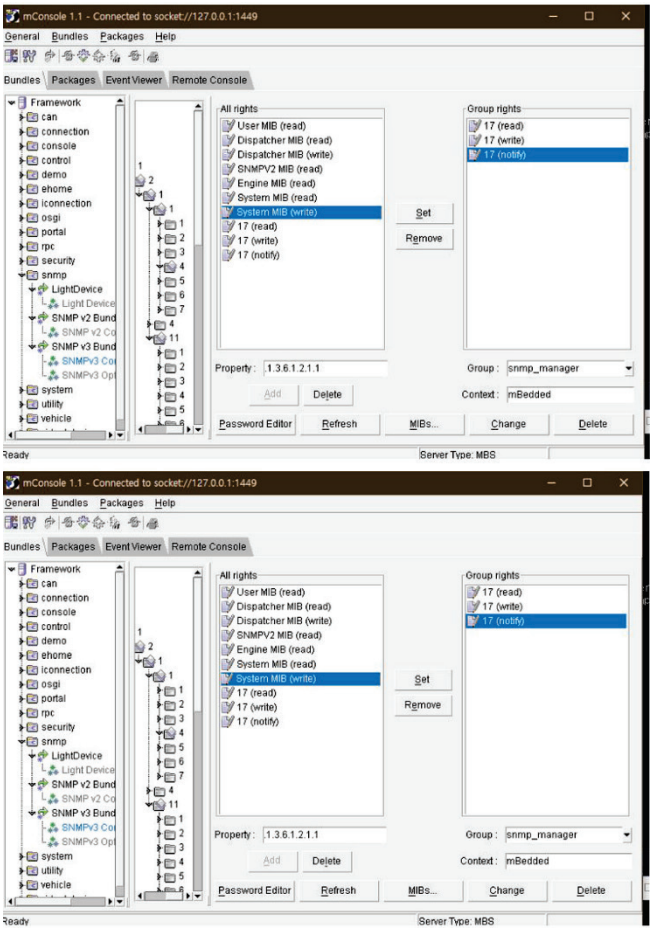


Figure 8 Query execution for dynamic recognition of situation

Fig. 8 shows the dynamic situation recognition query execution experiments and results configured based on IoT service device for lighting devices control [27]. IoT application service in this platform standard SNMP MIB information configured that based on constructed by assuming various dynamic situation recognition scenarios.

This facility device configuration should be installed as a common interface modularization for integrated network management. An application context registers the SNMP context object. The manipulator in a manner so that the SNMP service agent is notified of its existence and refers to the Light Service and registers it as a node in the MIB. In this way, SNMP Manager should be sending commands to the device, which the Light device represents. The Light device implements SNMP Manipulator. This interface is instantiated by listener services which then calls the MIB service object.

3 CONCLUSIONS

The IoT convergence service gateway is continuously expanding that should be supported for implementation of IoT facility devices and interface environments. Integration of IoT systems should be unified into a service represented in a dynamic situation recognition system based on distributed objects. A context awareness object configured based on Universal Middleware is easy to handle internal and external situations of a terminal with a standard interface for a communication module, a data module, and an application service module. Previously, various studies on contextual computing have been conducted, but part of the dynamic service linkage method based on Universal Middleware has not been performed. Practically, it is challenging to standardize the integration of module information provided by previous versions and contemporary device interfaces, in addition to integrating several network services. In response, in this study, a scalable platform was constructed based on the part that can objectify common resources for each SNMP version through Universal Middleware and the method that can be configured in the form of a life cycle process. Due to the constraints of the construction environment, this study could not demonstrate the use of service configurations with UPnP, DLNA, etc., which are part of the expansion of external systems, but if it is expected that usability for commercial devices can be secured, it is easy to configure convergence services.

According to the following, the standard interface of Universal Middleware could be handled with the event and tracking from context object through IoT service gateway. First, communication query, Second, data result, lastly, application service. First, like the lighting device shown in the standard communication module on the framework, it is possible to check SNMP configuration information at the same time as MIB information on the device's authority, dispatcher, engine, system, etc. Second, user tracking data may be used through Income information due to a query result on the framework. Third, it is possible to dynamically recognize situations between application service modules. Thus, by dynamically configuring these three standard service modules, you get the resources to configure the services for administrators and users as a simple and clear dashboard.

A context awareness object configured based on Universal Middleware is easy to handle internal and external situations of a terminal with a standard interface for a communication module, a data module, and an application service module. In particular, MIB Data information for each version based on SNMP enables efficient management of devices that require future scalability from basic information such as device settings, service configuration, domain configuration, and operator.

Through the evaluation of the system built based on the IoT SNMP platform design presented in this study, it was confirmed that instance objects that can be supported by dynamic system construction can be optimized and used as an enterprise-scale system. In this system, the SNMP agent service creates instances of the following assistant context objects. As the result, there are three kinds of context awareness classes. First, Device-Context properties maintain the property nodes of the facility device service in the configuration database. Second, Device-Base hosts all the nodes as fields that are inserted in the MIB tree. It provides methods for retrieval and modification of these fields. Third, Device-Trap, his is an SNMP trap, these reports changes in the value of the device nodes. Lastly, Device-Function such as turning on, off, and the other service functions.

In conclusion, the MIB information of dispatcher supports the dynamic incoming of numerous concurrent SNMP messages and manages the transfer of PDU system. As a future research plan, there is a need to assign the IP address of facility devices, solving the problem of returning to residential device service. Actually, dynamic awareness of context objects with mechanisms and gateway to optimization to publish residential information could be consisting of IoT interfaces to external service clients.

The configuration of dynamic modules that can be mounted on IoT service gateways discussed so far provides scalability to cover a wide range of service areas. In particular, it is expected that applying a platform designed based on Universal Middleware will contribute to the simplification of services suitable for the process of the IoT service. Optimization of dynamic service design is to grasp the propensity of service content, which was inherently complicated, rather than the expansion of network resources. It is necessary to use Universal Middleware to build situational recognition technology suitable for dynamic service design while maintaining hardware independence of network equipment, and it is expected to be applied as an example for the expansion of related research.

4 REFERENCES

[1] Müller, H. A., Villegas, N. M., Lau, A., & Jou, S. (2010). Dynamic context-aware applications: Approaches and challenges. *Proceedings of the 2010 Conference of the Center for Advanced Studies on Collaborative Research (CASCON '10)*, IBM Corp., USA, 404-406. https://doi.org/10.1145/1923947.1924016

[2] Wu, Y., Jia, J., & Wang, H. (2014). Using CORBA as a distributed extension solution for OSGi. *2014 IEEE Workshop on Electronics. Computer and Applications*, 178-181. https://doi.org/10.1109/IWECA.2014.6845587

[3] Morabito, R. L. (2019). Virtualization in edge computing for internet of things.

[4] Li, Y., Han, Z., Gu, S., Zhuang, G., & Li, F. (2021). Dyncast: Use dynamic anycast to facilitate service semantics embedded in IP address, 1-8. https://doi.org/10.1109/HPSR52026.2021.9481819

[5] Ishikawa, N. (2015). Middleware platform for smart home networks. *Ecological Design of Smart Home Networks*, 29-48.

https://doi.org/10.1016/B978-1-78242-119-1.00003-5
[6] https://docs.osgi.org/download/r6/osgi.cmpn-6.0.0.pdf. July (2015). Release 6.
[7] Ailisto, H. et al. (2011). Internet of things strategic research agenda (IoT-SRA). *Finnish Strategic Centre for Science, Technology, and Innovation: For Information and Communications (ICT) Services, businesses, and technologies*. https://www.researchgate.net/publication/258974575_Internet_of_Things_Strategic_Research_Agenda_IoT-SRA
[8] Doherty, J. (2014). Residential gateway system for automated control of residential devices.
[9] Wu, L., Sun, C., & Fan, F. (2022). Multi-criteria framework for identifying the trade-offs and synergies relationship of ecosystem services based on ecosystem services bundles. *Ecological Indicators*, 144. 109453. https://doi.org/10.1016/j.ecolind.2022.109453
[10] Kotsilieris, T., Zikos, P., Vlachos, E., Kalogeropoulos, S., Michalas, A., Karetsos, G., & Loumos, V. (2002). A prototype SNMP management framework for DiffServ linux routers. *Its Implementation and Performance*, 405-416. https://doi.org/10.1007/978-0-387-35620-4_35
[11] Saha, S., Eidmum, Md., Hemal, Md., Khan, Md., & Muiz, B. (2022). IoT based smart home automation and monitoring system. *Khulna University Studies*, 133-143. https://doi.org/10.53808/KUS.2022.ICSTEM4IR.0245-se
[12] Du, X., Cheng, C., Han, Z., Fan, W., & Ding, S. (2022). Hamiltonian properties of HCN and BCN networks. *The Journal of Supercomputing*. https://doi.org/10.1007/s11227-022-04723-w
[13] Suh, J. & Woo, C. -W. (2012). A Development of intelligent context-awareness middleware. *Journal of the Korea Society of IT Services*, 11, 165-176. https://doi.org/10.9716/KITS.2012.11.sup.165
[14] https://www.eclipse.org/equinox/framework/. (2022). Equinox Framework
[15] Petri, I., Rana, O. F., Bignell, J., Nepal, S., & Auluck, N. (2017). Incentivising Resource Sharing in Edge Computing Applications. In: Pham, C., Altmann, J., Bañares, J. (eds) *Economics of Grids, Clouds, Systems, and Services. GECON 2017. Lecture Notes in Computer Science, 10537*. Springer, Cham. https://doi.org/10.1007/978-3-319-68066-8_16
[16] Prakash, K., Lallu, A., Islam, F. R., & Mamun, K. A. (2016). Review of power system distribution network architecture. *The 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, 124-130. https://doi.org/10.1109/APWC-on-CSE.2016.030
[17] https://datatracker.ietf.org/doc/rfc2021/. (2022). IESG evaluation record
[18] IEEE Standard for Management Information Base (MIB) Definitions for Ethernet. (2011). In *IEEE Std* 802.3.1-2011, 1-474. https://doi.org/10.1109/IEEESTD.2011.5951710
[19] Lamaazi, H., Benamar, N., Jara, A. J., Ladid, L., & El Ouadghiri, D. (2013). Internet of thing and networks' management: LNMP, SNMP, COMAN protocols. *The International Conference on Wireless Networks and Mobile Communications*, At: Fes, Morocco.
[20] Sun, W., Feng, L., Honghui, L., & Chen, X. (2003). Implementing a CORBA/SNMP gateway with design patterns. *Proceedings of the Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies*, 855-857. https://doi.org/10.1109/PDCAT.2003.1236432
[21] Zeeshan, M., Siddiqui, M., & Rashid, F. (2019). Design and testing of SNMP/MIB based IoT control API. *IEEE 16th International Conference on Smart Cities: Improving Quality of Life Using ICT & IoT and AI (HONET-ICT),* 054-058. https://doi.org/10.1109/HONET.2019.8908111
[22] Hwang, G. -J., Yang, T. -C., Tsai, C. -C., & Yang, S. J. H. (2009). A context-aware ubiquitous learning environment for conducting complex science experiments. *Computers & Education*, 53(2), 402-413. https://doi.org/10.1016/j.compedu.2009.02.016
[23] Wang, L. (2010). Design and implementation of web service plug-in on Eclipse platform. *International Conference on Computer, Mechatronics, Control and Electronic Engineering*, 380-382. https://doi.org/10.1109/CMCE.2010.5610517
[24] Vidmachenko A. (2021). Features of Saturn's Equinox in 2010. *Kinematics and Physics of Celestial Bodies*, 37, 33-40. https://doi.org/10.3103/S0884591321010062
[25] Heo, G., Kim, E., & Choi, J. (2010). An extended SNMP-based management of digital convergence devices. *The 10th IEEE International Conference on Computer and Information Technology*, 2540-2547. https://doi.org/10.1109/CIT.2010.432
[26] Park, D., Cho, B., Park, T., & Lim, J. (2022). Process for identifying QoS requirements in the multi-domain operations environment. *Journal of the Korea Institute of Military Science and Technology*, 25, 177-186. https://doi.org/10.9766/KIMST.2022.25.2.177
[27] Javad Kalani, M. et al. (2022). Considering ambient sound variations to meet the requirements of smart cities. (2022). *Computers and Electrical Engineering, 102*, 10824. https://doi.org/10.1016/j.compeleceng.2022.108240
[28] Amila, S., Ling, L., & Shanika, K. (2021). Research on the application of an improved LEACH algorithm in smart home. *International Journal of Smartcare Home*, 1(1), 1-8. https://doi.org/10.21742/26531941.1.1.01
[29] David, T., Johan, B., & Lin, C. (2021). Research on real-time data transmission between IoT gateway and cloud platform based on two-way communication technology. *International Journal of Smartcare Home*, 1(1), 61-74. https://doi.org/10.21742/26531941.1.1.06
[30] Lu, L. (2021). Dynamic matrix clustering method based on time series. *Journal of Smart Technology Applications, 2*(1), 9-20. https://doi.org/10.21742/JSTA.2021.2.1.02
[31] Han, J. & Oh, S. (2018). A Study of IoT Home Network Management System Using SNMP. *International Journal of Control and Automation, NADIA, 11*(5), 163-172. https://doi.org/10.14257/ijca.2018.11.5.14

Author's contacts:

Hae-Jun Lee, Associate Professor
IWCR, Kangnam University,
40, Gangnam-ro, Giheung-gu, Yongin-si, Gyeonggi-do, 16979, South Korea
haejuna@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230220172540

Original scientific paper
Received: 2023-02-20, Accepted: 2023-04-06

AI Machine Vision based Oven White Paper Color Classification and Label Position Real-time Monitoring System to Check Direction

Hee-Chul Kim*, Youn-Saup Yoon, Yong-Mo Kim

Abstract: We develop a vision system for batch inspection by oven white paper model color by manufacturing a machine vision system for the oven manufacturing automation process. In the vision system, white paper object detection (spring), color clustering, and histogram extraction are performed. In addition, for the automated process of home appliances, we intend to develop an automatic mold combination detection algorithm that inspects the label position and direction (angle/coordinate) using deep learning.

Keywords: angle; color image processing; coordinate determination; machine vision; optimal operation; production management

1 INTRODUCTION

We support the process system of manufacturing companies to increase sales and promote employment through the successful technological advancement of companies. It is time to support the development of process automation systems in line with the 4th industrial revolution. It has secured significant technological accumulation and IoT solution development technology through government R&D projects of small and medium-sized enterprises that have created artificial intelligence convergence complexes. Appropriate government support is needed to overcome difficulties such as high risks in the subsequent commercialization stage and lack of self-commercialization capacity [1]. The driving force behind the machine vision market includes challenges that need to be addressed due to the lack of user awareness of the rapidly changing machine vision technology, such as growth accelerators, growth restraints, and market opportunities. Overcoming limitations in label position inspection and visual inspection of white paper color classification of products based on artificial intelligence. This is because it is important to increase the productivity by improving the reliability and speed of inspection. It is urgent to develop and commercialize low-cost machine vision systems and automation software solutions suitable for small and medium-sized enterprises (SMEs), as it is required to avoid such qualitative judgments and establish quantitative judgment standards. In order to improve domestic manufacturing productivity and competitiveness, it is necessary to make smart production sites [2].

Technologies such as vision, AI, and robot control are essential fields that must be promoted along with the wave of the global 4th industrial revolution. The government is also focusing on key areas such as the Ministry of Industry's smart factory supply project, smart industrial complex creation project, and the Ministry of SMEs and Startups' DNA (Data, Network, AI) Korea construction project. Although these technologies are currently showing great results in global companies such as Dassault Systèmes and Siemens, they are mostly applied to large corporations. For small and medium-sized enterprises (SMEs) to apply, it is a field that has limited entry due to size and cost limitations. In addition, although manufacturing innovation that converges each independent unit technology is essential for smart factories, it is not easy to apply in small and medium-sized enterprises (SMEs), and although government-level support measures are being implemented. They are still insufficient products to be applied to the manufacturing process using AI are being introduced competitively by various companies, but it is difficult to compare and evaluate them as products for general consumers as it requires customization and grafting of know-how to suit the situation of the target company. Currently, technologies for controlling robots by grafting vision are being developed and introduced by many companies and research institutes under the category of digital twin. In Korea, there is an attempt to build a production line that incorporates this in the actual manufacturing process, but it is very difficult for small and medium-sized enterprises to apply this technology without government support as they are built mainly by large companies.

By using AI-based object recognition algorithm, the position of the spring on the hinge is identified to increase the accuracy of product identification. Based on the histogram of the image obtained when identifying the corner, the reference value for judging the product was selected as the learned result. Smart factory-related technologies, such as product shape recognition, discrimination using AI, and robot control technology through simulation, are applied to actual processes and a system that can contribute to productivity improvement has been established.

This paper is the development of a real-time monitoring system that inspects artificial intelligence based home appliances.

In Chapter 2, the vision inspection system design and Chapter 3 develops a machine vision system for the automated process of home appliances, In Chapter 4, an automatic mold defect detection algorithm using deep learning was developed [3, 4, 17].

2 RELATED WORKS

2.1 Real-Time Monitoring Platform Design

We designed the infrastructure to configure the hardware of the real-time quality control monitoring system for home appliances based on AI machine vision. As shown in Fig. 1, design H/W for optical and real-time Auto-Focus inspection processing that inspects product appearance. We are going to implement an inspection image processing system for development and segmentation of product injection mold appearance and image processing [5].

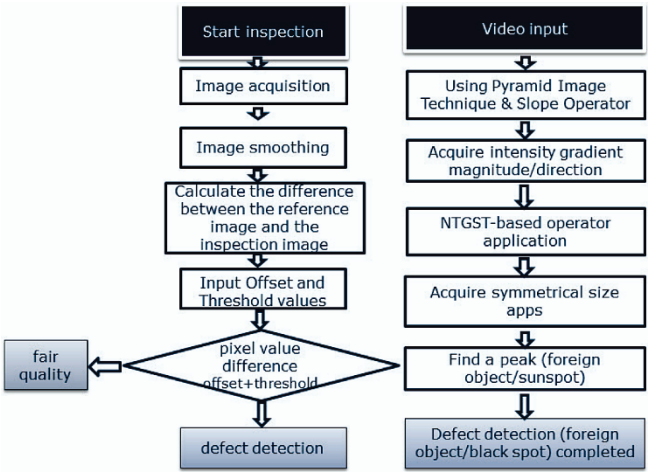


Figure 1 System logic

AI-based vision inspection algorithms learn and analyze images. We develop solutions to improve process efficiency and inspection accuracy by detecting difficult- to-detect defects, fast speed, and low errors. Design an integrated big data management platform architecture such as data collection, storage, exchange, and security for AI-based analysis by building a database for developing AI models for bad prediction [6]. Decision tree, random forest, naive Bayes, SVM, gradient boosting, etc. are used to utilize classification algorithm in the development of data set quality inspection module for AI-based automatic reading of quality information. PCA principal component analysis uses AI-based quality prediction modeling using quality inspection data and collected data [7, 8]. Defect occurrence prediction and predictive modeling results are linked in designing the monitoring system [9, 10]. Labeling is performed on various types of defects leaked from here.

2.2 Design of Vision Inspection System to Acquire Appearance Information

In Fig. 2, it is implemented as an integrated structure that can inspect both appearance inspection and Cabi curvature at the same time by designing and implementing the exterior vision inspection system infrastructure [11, 12, 18].

The inspection object is classified through the result of comparing/ analyzing/ processing images taken according to the vision inspection algorithm. As for the inspection object, we develop an injection mold exterior vision device that judges the status of good or defective products (scratches, dents, foreign substances detection, etc.) within 1 minute [13, 14, 19]. Optical and real-time Auto-Focus processing H/W technology is applied as shown in Fig. 3 to obtain results obtained by comparing, analyzing, and processing images taken according to the vision inspection algorithm. This is to develop an injection mold exterior vision equipment that judges the condition of the inspection object as good or defective (scratches, dents, foreign substances detection, etc.) within 1 minute. Here, the optical and real-time Auto-Focus inspection processing H/W design that inspects the curvature of home appliances is reflected.

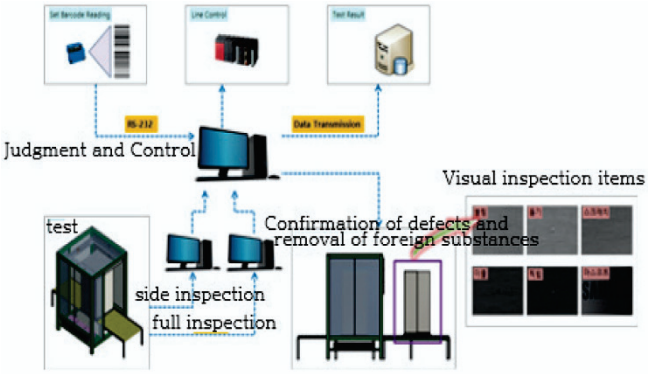


Figure 2 Facade Vision Inspection System Design

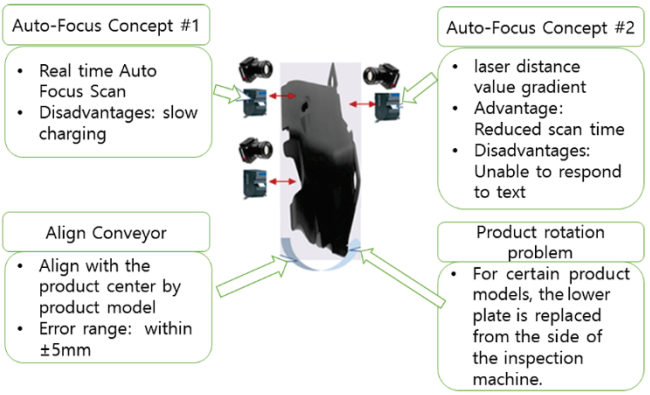


Figure 3 Optical and real-time Auto-Focus processing H/W technology

In Fig. 4, the defect calculation and judgment result are output using the blob area of the binarized image by removing the afterimage by adjusting the contrast from the image acquired by the scratch S/W algorithm.

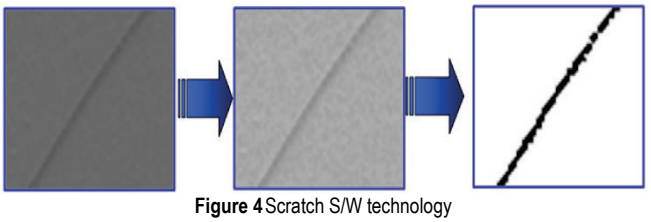


Figure 4 Scratch S/W technology

In Fig. 5, the afterimage is removed by adjusting the contrast from the image acquired by the foreign substance or sunspot S/W algorithm. By calculating the binarized image on the removed image, the binarized blob area is obtained,

and foreign substances are detected in the image with gradation, and the found foreign substances are displayed in real time [15].

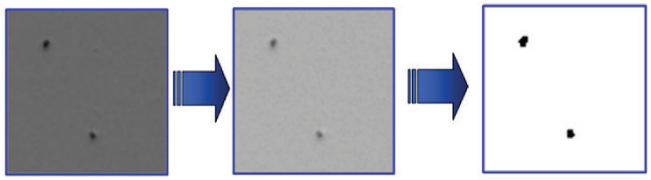


Figure 5 Scratch S/W technology

In Fig. 6, the standard deviation S/W algorithm calculates the standard deviation of the averages within the inspection area to set the range for non-defective products. Here, an area having an average value out of the range is judged as defective and displayed [16].

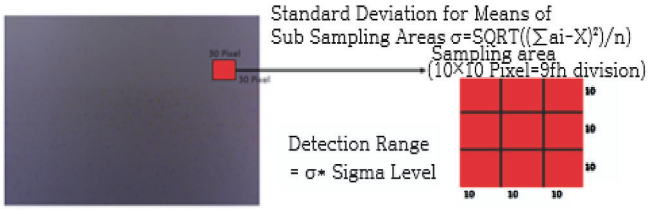


Figure 6 Standard deviation S/W technology

3 RESEARCH METHODS AND CONTENTS

3.1 Machine Vision Systems for Oven Manufacturing Automation Processes

Fig. 7 is a color classification system for classifying white paper models for oven doors. Oven doors are heavier than doors of general home appliances due to heat resistance and the like, and an internal spring is used to support them. The color and size of the white paper spring used for each model in the automated process are different.

After acquiring the image of the spring part using the machine vision system, we plan to identify the model by classifying the size and color of the spring. This is to detect the spring size of white paper by applying an object detection algorithm.

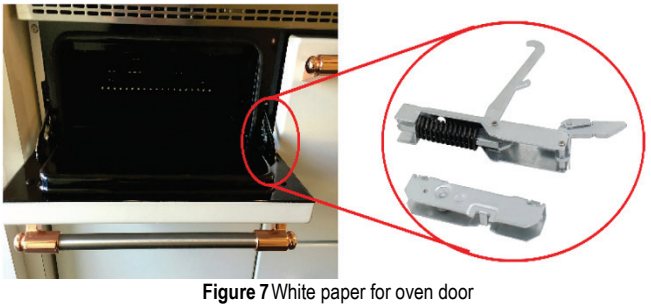


Figure 7 White paper for oven door

3.2 Vision Alignment System for Product Labeling

Object detection (spring) is to extract only the spring part from white paper by applying the image object detection algorithm. Color clustering removes distorted information caused by light reflections and the like through color clustering. Here, histogram extraction is to extract the main color range by obtaining an RGB-related histogram based on the clustered image. By combining these two algorithms, the color detection algorithms are listed in Fig. 8 to identify the whiteness of the product being worked on and to check whether the product is currently being put into assembly.

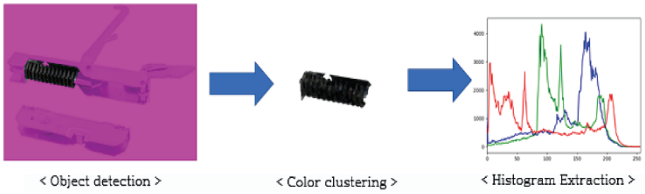


Figure 8 Color detection algorithm

3.3 Edge S/W Algorithm

In Fig. 9, FOV and Fixel values of the inspection effective part are given as an image processing S/W technology for image segmentation and image processing in the development of the interlocking system. The operating PC is designed to be able to respond immediately when abnormal conditions such as distributed processing speed and line scan inspection speed occur.

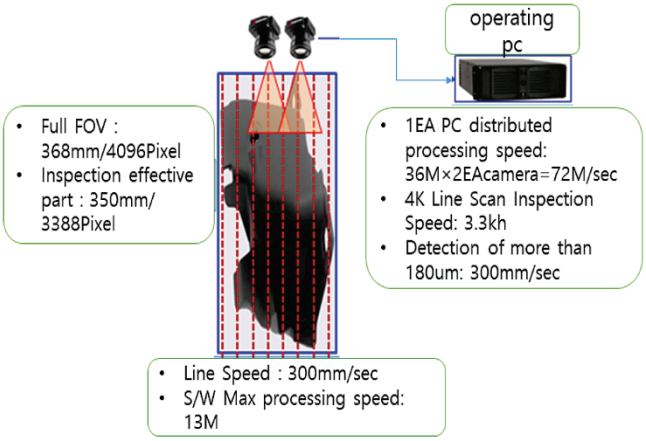


Figure 9 Image Processing Inspection image processing S/W technology

Fig. 10 shows the Edge S/W algorithm, and it is an image to which Sub Pixel is applied. In the primary differential method, the number of pixels in the original image is 5, and the edge is detected by the differential method in the enlarged image using sub pixels.

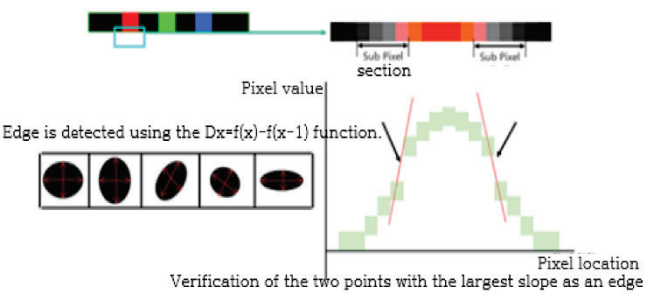


Figure 10 Edge S/W Algorithm

3.4 Big Data Management Platform Architecture

Design an integrated big data management platform architecture such as data collection, storage, exchange, and security for AI-based analysis. It is to analyze and store a large amount of structured and unstructured data generated in the manufacturing process in real time and provide it to users. Design and use a big data management platform as shown in Fig. 11 [11, 12].

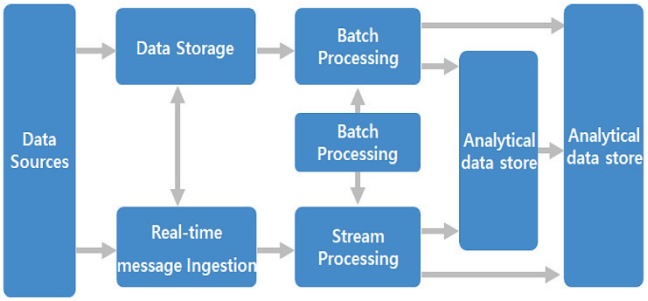


Figure 11 Big Data Management Platform Architecture

4 EXPERIMENTAL RESULTS

4.1 Development of Automatic Mold Defect Detection Algorithm Using Deep Learning

In order to predict product defects from the data acquired in the inspection process for vision inspection, the most suitable predictive model is implemented for the training data. The learning data is targeted for 24-hour cumulative data of the day when the defect occurred. It is estimated by using the decision tree model under which conditions a defect occurs, and the composition of the variables is shown in Tab. 1 below.

Table 1 Composition of variables

Variables	Description
Date	data generation time
Model No	Model Number
Statics	color, material, propose
Defect factor	Defective quantity Defect factor

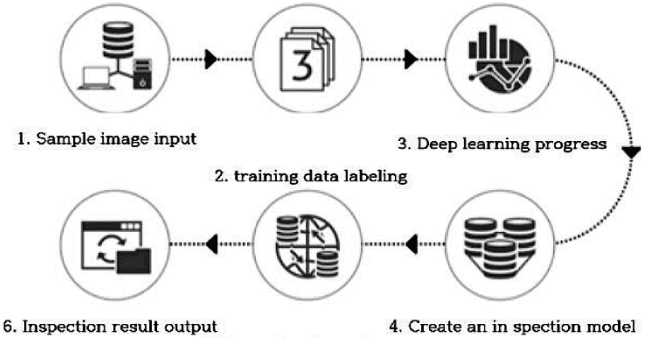


Figure 12 Automatic Mold Defect Detection Algorithm Using Deep Learning

In developing the methodology for deriving optimal control variables, log data generated from the process model is acquired through a vision sensor. The data manager pre-processed the data, and it was used as a model to suggest improved process variables based on empirical data and knowledge. For the process variable control model, an ensemble classification model that derives the optimal process variable region (process window) through voting of four classification models was applied. The optimal process variable region defines the margin border of the variable where defects occur. We use a regional ensemble classification model that combines optimal process variables that do not cause defects. First, if a defective product image is input, additional training data is generated as needed to label each defect type. A deep learning model with optimal conditions is created through the learning steps according to the selected mode and requirements. When the final model is created, the inspection image of the mass-produced product on the actual production line is input to determine whether it is a good product or a defective product. As shown in Fig. 12, the related information is an automatic mold defect detection algorithm using deep learning.

4.2 Edge Acquisition Algorithm

4.2.1 Hinge Recognition Unit

Fig. 13 shows the different types of oven hinges used. An algorithm was developed to classify the spring part of the equipment based on the size and color. After recognizing the spring using the object recognition algorithm, the length is measured. In addition, by analyzing the color distribution through the histogram, the oven hinge of the selected model is distinguished.

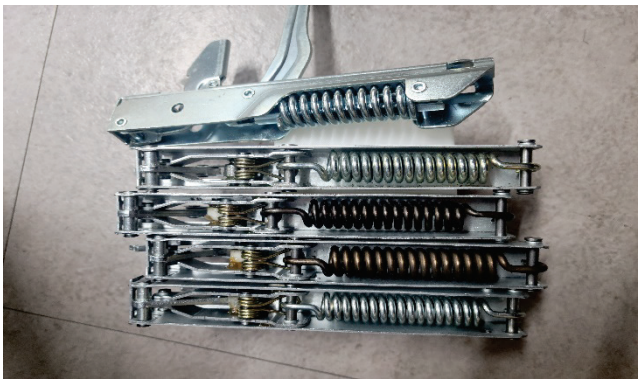


Figure 13 Oven hinges

4.2.2 Corner Recognition Unit

Fig. 14 is the result of recognizing the corners of the oven. After recognizing the corners of the top and sides of the oven, the exact location is derived on the image based on this. A black and white camera is used to acquire the image of the product, and the top and side are distinguished based on the black and white histogram. Based on the recognized line, two intersections are recognized as edges. Recognized edges calculate the exact product edge location based on the actual distance per pixel. Based on the value of the actual distance obtained in this way, a barcode label is later attached using a robot.

For edge recognition, the above figure and algorithm are implemented on the Region of Interest, the outline of the top and side of the product is recognized. The intersection point is calculated on the image, and the actual distance is calculated based on the calculated distance value per pixel by

performing calibration. Based on this actual distance, it is possible to attach a barcode at a desired location on the product in the real coordinate system.

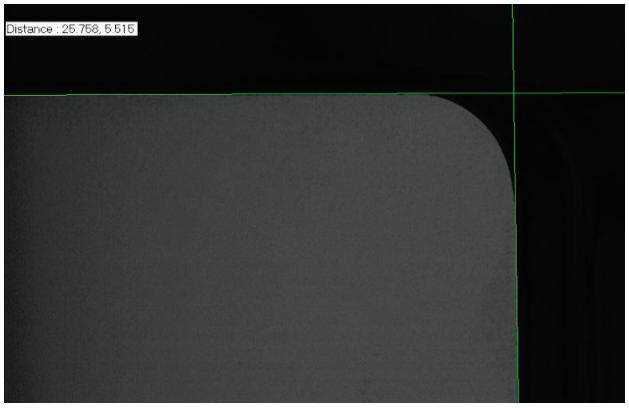


Figure 14 Edge recognition result

4.3 AI Exterior Vision Monitoring System Establishment

When a model that satisfies all the requirements for inspection is created through learning, a system for monitoring the quality of the exterior in real time at the actual manufacturing site is built. It provides name node and data node system information through a dashboard-type monitoring environment. In Fig. 15, the number of Map·Reduce jobs and program success/failure information are provided, and system progress monitoring and log events are tracked.

Figure 15 Application of AI injection mold exterior vision monitoring system

4.4 Color Classification and Label Attachment Position Alignment System

Fig. 16 is an oven hinge color classification and label attachment position alignment system applied with machine vision technology. Using two machine vision systems, images of hinges and edges for ovens are acquired, and image processing is used to recognize hinge models and edges for labeling. The oven hinge color classification and label attachment position alignment system to which this machine vision technology is applied is a system for deriving performance test results.

Figure 16 Oven hinge color classification and label attachment position alignment system

4.5 Hinge Model Grasp Speed

After preparing 5 different hinges according to the test conditions and methods, one or more hinges are put into the equipment. In addition, after acquiring the external image of the input hinge through the machine vision system, the object recognition algorithm measures the time to recognize the currently selected model among the input hinges. The criterion for decision was within 400 ms, and a result value of 282 ms was obtained as a vision response time. Detailed test results are shown in Tab. 2.

Table 2 Detailed test results

Exam conditions	Number	Vision response time (ms)
Hinge recognition speed	1 time	279
	Episode 2	285
	average	282

5 CONCLUSION

In this study, the final inspection of the manufactured product is carried out to improve the manufacturing process. By significantly lowering inspection time and cost, productivity and management convenience of the process line have been increased. By securing the base technology, a technological foundation was prepared that could dramatically lower the production cost of products, which had previously been a problem. This created the effect of further enhancing the degree of completion of commercialization.

These research results have improved quality and reliability, increasing quality satisfaction due to the introduction of a one-stop quality inspection system from the manufacturing stage to completion. The speed, stability, and accumulation of inspection results have made it possible to realize smart factories through big data in production and manufacturing processes. Through real-time process status monitoring, changes in the process capability index are immediately detected, and the process utilization rate is improved by shortening the processing time of abnormal factors such as registration, notification, and processing in case of abnormalities. Therefore, it has a relatively shortened inspection time and cost reduction effect and it is possible to conduct automatic discrimination inspection linked to the established product-specific database for future product line additions.

Based on the improvement of the quality of the products produced by the manufacturing company, it can be of great help to the growth of the company based on the increase in corporate reliability. In the inspection of parts such as automobile bearings, where safety is important, human error (human error) of the visual inspector was minimized and expectations were raised for improved accuracy in various inspection fields. In addition, healthcare, mail sorting, intelligent transportation system (ITS), intelligent security CCTV, autonomous vehicles, etc. It is expected that machine vision technology will be widely used in a wide range of fields, and eventually almost all inspections that can be done with the eyes can be replaced with machine vision.

In addition, through multidimensional analysis of process data, quality problems were detected in advance to increase production efficiency by improving the defect rate. By acquiring and analyzing mold data and defect data in real time, the optimal combination of process variables that do not cause defects was derived. In addition, the flexibility of inspection work for process variable recombination is improved and costs are reduced.

Acknowledgements

This research was supported by a research program, which was sponsored by GwangJu University in the 2023 school year.

6 REFERENCES

[1] Sarmah, S. S. (2020). An efficient IoT-based patient monitoring and heart disease prediction system using deep learning modified neural network. *IEEE Access*, 8, 135784-135797. https://doi.org/10.1109/ACCESS.2020.3007561

[2] Sung, K. K. (2017). A study on the IoT technology trend and utilization plan. *Journal of Next-generation Convergence Technology Association, 1*(3), 121-127.

[3] Pickell, P. D., Chavardes, R. D., Li, S., & Daniels, L. D. (2021). FuelNet: An artificial neural network for learning and updating fuel types for fire research. *IEEE Transactions on Geoscience and Remote Sensing, 59*(9), 7338-7352. https://doi.org/10.1109/TGRS.2020.3037160

[4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

[5] Masoumi, S., Baum, T. C., Ebrahimi, A., Rowe, W. S. T., & Ghorbani, K. (2021). Reflection measurement of fire over microwave band: A promising active method for forest fire detection. I*EEE Sensors Journal, 21*(3), 2891-2898. https://doi.org/10.1109/JSEN.2020.3025593

[6] Choi, S. Y. & Shin, S. J. (2021). A study on expansion proposal of data dividend qualification based on the contribution of platform workers. *The Journal of the Institute of Internet, Broadcasting and Communication (JIBC), 12*(6), 267-272. https://doi.org/10.7236/JII

[7] Jeon, J. S., Koo, J. K., & Park, C. M. (2015) Outlier detection in time series monitoring datasets using rule based and correlation analysis method. *Journal of the Korean Geo-Environmental Society, 16*(5), 43-53. https://doi.org/10.14481/jkges.2015.16.5.43

[8] Bor, M., Edward, J., & Roedig, U. (2016). LoRa for the Internet of Things. *Proc. International Conference on Embedded Wireless Systems and Networks,* Graz, Austria, 361-366.

[9] Roh, T. H. (2021). Accuracy analysis of road alignment reconstruction convergence technology based on mobile laser scanning technology. *Journal of Next-generation Convergence Technology Association, 5*(3), 426-436. https://doi.org/10.33097/JNCTA.2021.05.03.426

[10] Kang, G. C., Kim, B. J., Hong, S. W., & Yim, Y. C. (2017). Improving reliabilities of dam displacement based on monitoring given points by total station. *Journal of Korean Geosynthetics Society, 16*(1), 1-8. https://doi.org/10.12814/jkgss.2017.16.1.001

[11] Sharifi, M., Fathy, M., & Mahmoudi, M. T. (2002). A classified and comparative study of edge detection algorithms. *International Conference on Information Technology: Coding and Computing*, 117-120. https://doi.org/10.1109/ITCC.2002.1000371

[12] Lee, K. J., Kim, J. H., Ha, M. S., & Cho, K. H. (2020). Measurement management system using Lora, sensor node, and cloud platform. *Journal of the Korean Society of Hazard Mitigation, 20*(6), 143-150. https://doi.org/10.9798/KOSHAM.2020.20.6.143

[13] Yoo, C. H., Kim, I. H., Lee, S. J., Hwang, J. S., & Baek, S. C. (2018). Basic study on monitoring system of reservoir and leeve using wireless sensor network. *Journal of the Korean Geo-Environmental Society, 19*(1), 25-30. https://doi.org/10.14481/JKGES.2018.19.1.25

[14] Yoon, S. Y. & Lee, Y. W. (2022). UTOPIA smart city AI middle ware and person detection. *Journal of Next-generation Convergence Technology Association, 6*(5), 750-759. https://doi.org/10.33097/JNCTA.2022.06.05.750

[15] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, TPAMI-8*(6), 679-698. https://doi.org/10.1109/TPAMI.1986.4767851

[16] Prachi, P., Rashmi, G., & Ashwani, K. (2019). Neural networks for facial age estimation: A survey on recent advances. *Artificial Intelligence Review, 53*(5), 3299-3347. https://doi.org/10.1007/s10462-019-09765-w

[17] Al-Eiadeh, M. R. (2021). Automatic Lung Field Segmentation using Robust Deep Learning Criteria. *International Journal of Hybrid Information Technologies, 1*(1), 69-82. https://doi.org/10.21742/ijhit.2653-309X.2021.1.1.06

[18] Barthelemy, A. M. & Suter, G. (2021). Highly Intelligent Recommendation Algorithm based on Matrix Filling. *International Journal of Hybrid Information Technologies, 1*(1), 83-96. https://doi.org/10.21742/ijhit.2653-309X.2021.1.1.07

[19] Park, H.-K. (2016). Development of Machine Vision Monitoring System for Semiconductor Package Sorter. *International Journal of Control and Automation, NADIA, 9*(4), 63-72. https://doi.org/10.14257/ijca.2016.9.4.07

Authors' contacts:

Hee-Chul Kim, Professor, PhD
(Corresponding author)
Gwangju University,
277 Hyodeok-ro, Nan-gu, Gwangju, Korea, 61743
Tel./Fax: 062-670-2023/062-670-2187
E-mail: dangsali@gwangju.ac.kr

Youn-Saup Yoon, Professor, PhD
Gwangju University,
277 Hyodeok-ro, Nan-gu, Gwangju, Korea, 61743

Yong-Mo Kim, Professor, PhD
Division of Convergence Design, Major Visual Communication Design,
Gwangju University,
277 Hyodeok-ro, Nan-gu, Gwangju, Korea, 61743

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230224142711

Original scientific paper
Received: 2023-02-24, Accepted: 2023-04-06

A Study on the Effect of Quality Factors of Smartphone 5G Technology on the Reliability of Information and Communication Policy

Chil-Yuob Choo

Abstract: This paper analyzes the effect of the characteristics of 5G services on users' continuous intention to use, focusing on the technology acceptance model. With the start of the fourth industrial revolution in the 21st century, 5G is the best technology used in the Internet of Things, high-speed information and communication, artificial intelligence, big data, autonomous vehicles, virtual reality, augmented reality, robots, nanotechnology, and blockchain. The technical characteristics of 5G ultra-high-speed information communication are represented by ultra-high speed, ultra-high capacity, ultra-low delay, and ultra-high connectivity. 5G mobile communication technology is essential, and after the technology provided by 5G services is commercialized, it can play all its roles as a practical core new growth engine. 5G mobile communication (hereinafter referred to as 5G) is far superior to LTE, which is a 4G mobile communication, in terms of transmission speed, waiting time, and terminal capacity. 5G service is not just an axis of the process of developing mobile communication technology, but also the creation of innovative corporate value of technology. This is because higher network quality and innovation with 5G service technology will improve perceived usability, perceived ease of use, and perceived entertainment, which will ultimately have a positive impact on users' intention to use 5G services. Therefore, due to the lack of investment in information and communication bases, platforms, and applications, this paper can be used as the basis for establishing government policies.

Keywords: 5G; AI; big-data; ICT; IOT; smart phone quality

1 INTRODUCTION

This study examines the service quality factors of 5G smartphones, one of the 4th revolution projects that have recently become an issue, closely analyzes the effects of these quality factors on user satisfaction from the perspective of the technology acceptance expansion model, and identifies how user satisfaction affects government policy.

The effect of 5G service quality factors on perceived usability, perceived ease of use, and user satisfaction based on the technology acceptance model, and how this use satisfaction affects the trust of government policies are constructed by the technology acceptance research expansion model.

Modern information society is the 4th revolutionary industry, and hope for information transmission and utilization of economic value and future is growing as information transmission is rapidly transmitted anywhere and anywhere through high-speed, hyper-connected, and big data. In Korea, the world's first ceremony to declare the use of 5G was held in 2019, and the vision for the fourth revolutionary industry is gradually being promoted day by day. By early 2020, it is predicted that the era of 100 years of Smart Convergence Hundred, in which the era of 5G smart information that can lead the world, becomes common. By 2020, the government has designated 10 industries such as 5G smartphones, robots, and drones as 5G+ (plus) strategic industries, and the Ministry of Science and ICT has selected 5 service projects and 10 industries as strategic industries. This study should study the government's credibility in infrastructure and platform for high-speed information delivery industry, and establish support and supplement system for Korea to become a leader in global 4th revolutionary industry and improve people's lives and quality.

As the scope of the information system, such as the establishment of a customer relationship management (CRM) system for customer-centered management, gradually expands and various complexity increases, information data quality becomes the most important factor. Therefore, this study focuses on the information quality, service quality, and system quality of public data for the 4th revolution project, and reveals the relationship between the government's reliability and policy trust.

2 RELATED WORKS

2.1 Theoretical Background

The technology acceptance expansion model was introduced by Davis [1] by modifying Davis presented a technology acceptance expansion model as a basis using the variables perceived usefulness and perceived convenience as the main related variables of computer acceptance behavior. The perceived usefulness and perceived convenience variables can be said to be the core of the technology acceptance expansion model (TAM). Analyzing the definitions of these two variables, perceived usefulness refers to the difference in the degree to which individuals believe that working with a particular information system can improve work performance, and perceived convenience refers to the degree to which they believe working with it will be a little less [1]. The technology acceptance the rational behavior theory, and the purpose of the introduction was to model the user acceptance of the information system. The technology acceptance expansion model provides a sufficient explanation of the determinants of computer acceptance and is characterized by a wide range of end-user computing technologies and comprehensive enough to explain user behavior across the user population. The core purpose of the technology acceptance expansion model is a technology that provides a basis for tracking the integrated impact of external

factors on inner beliefs, attitudes, and intentions. In order to achieve these goals, the technology acceptance expansion model fully identified a small number of basic variables presented by continuous previous studies dealing with the determinants of computer acceptance and used rational behavior theory as a theoretical background to model them based on their theoretical relevance expansion model proposed by Davis [2] is shown.

2.2 Exploration of the Previous Research Model

In the field of information systems such as [5], models for expanding technology acceptance are being expanded and applied in various forms. As external variables suitable for mobile media, information system quality, social impact, self-efficacy, driving conditions, and trust were presented, and the causal relationship with mobile banking was verified from the overall perspective of TAM, trust, and influence variables. In addition, in a study on the success factors of the mobile video UCC service, the promotion of personal characteristics, system characteristics, and content characterization was expanded by expanding mobile content characteristics to analyze users' learning about e-learning. Empirical analysis of non-users and users of smartphone applications in the technology acceptance expansion model shows that they are willing to use the application in terms of perceived enjoyment and usefulness, such as repeated verification of the technology acceptance expansion model and adding external or belief variables. Although there are various studies on the extended technology acceptance model, website information quality has a significant effect on perceived pleasure, perceived usefulness, perceived ease of use, especially accuracy, clarity, sufficiency, and reliability among website information quality characteristics. The trend of research on the recently expanded technology acceptance expansion model is to supplement the explanatory power of information technology delivery technology by integrating the core variables, perceived ease, and perceived usefulness of the technology acceptance expansion model [6]. In order to explore previous studies related to 5G information technology, which is the main focus of this study, the study on the extended technology acceptance model centered on the Internet and mobile was examined. And Kim Soo-Hyun [7] expanded the effect of IPTV service on adoption intention.

2.3 Rational Theory of Behavior (TRA)

The theory of reasoned action (TRA) model is an extended model widely studied in social psychology, in which individual behavior is determined by behavioral intention, and behavioral intention is determined by individual attitude and subjective norm [3]. Behavioral intention is estimated to contain motivational factors that affect actual behavior. Attitude to behavior refers to expectations related to behavioral evaluation, and evaluation refers to evaluations related to desirable performance achievement, which are expressed as beliefs along with subjective norms. In addition, subjective norms are defined as perceived social pressures associated with performing actions. Normative beliefs also refer to beliefs related to meaningful behavioral expectations. Motivation for consent means motivation for pursuing consent of the subject of reference. However, there is a limitation that this expansion model has no choice but to apply only to actions within the scope of voluntary control. There must be a problem when applying information to areas such as the use of involuntary information technology that require the skill to access and utilize computer systems [4].

2.4 Planned Behavior Theory (TPB)

The theory of planned behavior (TPB) model is an extended theoretical model that adds the concept of a user's perceived behavioral control to the rational behavioral theory (TRA) model. This is a theoretical model for dealing with the behavior of people with incomplete voluntary control, and its scope of application is wider than the rational behavioral theory model. Perceived behavioral control can be seen as related to people's perception of the degree of ease (excessive resources, opportunities, proficiency, etc.) in performing actions. In other words, it is based on the assumption that the performance of behavior depends on both ability (behavior control) and motivation (behavior intention). Control beliefs can be said to be the degree to which an individual can control the resources, opportunities, and proficiency to perform an action. Perceived facilitation is a concept related to whether or not to promote control beliefs to prevent or promote action. However, empirical analysis of this series of mutual causality and interdependence between preceding variables is still insufficient [4].

3 RESEARCH AND DESIGN

Based on previous studies on the expanded technology acceptance model, a research model was established as shown in Fig. 1. Since the commercialization of smartphones using 5G information technology is not satisfactory yet, the results derived when applying 5G information technology to the smartphone quality sector based on existing data quality factors will be very significant.

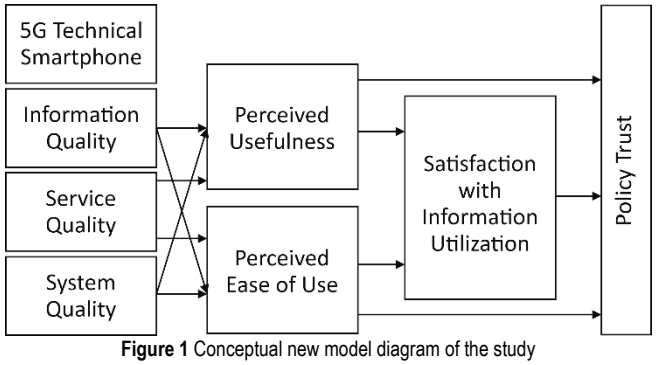


Figure 1 Conceptual new model diagram of the study

Information quality, service quality, and system quality were set based on the research of Lee Seong-Taek [8] as smartphone quality factors to which 5G information

technology was applied. To see how these factors finally affect government policy trust related to 5G information technology, perceived usefulness and perceived ease were introduced as parameters, and information utilization satisfaction was selected as parameters in the relationship between perceived usefulness and perceived ease of government policy trust.

For the design of this study, a survey was conducted on 310 people. As a result, there were 157 men (50.6%) and 153 women (49.4%), and 78 people (25.2%) in their 20s, 76 (24.5%) in their 30s, 76 (24.5%) in their 40s, and 80 (25.8%) in their 50s. There were 37 (11.9%) graduates from high school or below, 230 (74.2%) graduates from junior college or university, 43 (13.9%) graduates from graduate school or above, 6 (1.9%) public servants, 39 (12.6%) professionals, 161 (51.9%) office workers, 45 (14.5%) students, and 59 others (19.0%). There were 238 people (76.8%) in "yes" and 72 people (23.2%) in "no", and 92 people (29.7%) in "yes" and 218 (70.3%) in "no" in terms of 5G technology usage experience. The advantages of applying 5G technology to government public services were 78 (25.2%), 71 (22.9%) for disaster response and public safety services, 123 (39.7%) for big data-based intelligent services, 37 (11.9%) for VR and AR services, and 1 (0.3%).

As mentioned above, Jeong Man-Soo [9] while studying the intention to use 5G mobile communication services applied Amos 21 Structural Equation Modeling [10]. We constructed a new model by referring to a research model based on an extended technology acceptance model. It verifies the effect of belief variables on intention to use based on the research model.

Exploratory factor analysis was conducted to verify the validity of the measurement tool used in this study. Among the factor analysis methods, principal component analysis was used to minimize information loss while extracting factors explaining as much of the variance of the original variables as possible, and Varimax rotation was used to rotate the factor until the factor structure was most obvious.

As for the smartphone quality factors to which 5G technology is applied, one item (information quality number 4) that hinders validity was excluded, and the factor analysis was finally conducted with 11 items. As a result of the analysis, the *KMO* measure is. It was found to be 910, and Bartlett's sphericity test results were also significant ($p < 0.001$), the factor analysis model was judged to be suitable.

Factor classification was classified as a factor when factor loading exceeded 0.50, and smartphone quality factors applied with 5G technology were classified into three factors, and the three factors showed 64.743% factor explanatory power. The first factor was composed of three items, 'information quality', the second factor was four items, and the third factor was four items, 'system quality'.

In addition, the Cronbach's α coefficient was 0.783, the service quality was 0.785, and the system quality was 0.779. As a result, since all measurement tools showed a reliability index of 0.70 or higher, the reliability of the measurement tools was judged to be good.

The perceived usefulness and perceived ease of use were excluded from two items that hinder validity (recognized ease of use No. 1 and No. 2) and finally, a factor analysis was conducted with 12 items. As a result of the analysis, the *KMO* measure is. It was found to be 899, and Bartlett's sphericity test results were also significant ($p < 0.001$), the factor analysis model was judged to be suitable.

Table 1 Results of factor analysis of smartphone quality factors using 5G technology

Category	Main Cause		
	1	2	3
Information Quality 2	0.767	0.208	0.244
Information Quality 1	0.757	0.327	0.206
Information Quality 3	0.729	0.210	0.220
Service Quality 3	0.109	0.814	0.129
Service Quality 4	0.189	0.756	0.119
Service Quality 2	0.372	0.659	0.207
Service Quality 1	0.442	0.575	0.246
System Quality 1	0.128	0.080	0.796
System Quality 3	0.150	0.351	0.740
System Quality 2	0.474	0.061	0.663
System Quality 4	0.396	0.233	0.584
An Eigenvalue (Eigen value)	2.495	2.380	2.247
Co-Dispersion (%)	22.678	21.639	20.426
Cumulation Variance (%)	22.678	44.317	64.743
Cronbach's α	0.783	0.785	0.779
KMO = 0.910 Bartlett χ^2 = 1333.793 ($p < 0.001$)			

Factor classification was classified as a factor when factor loading exceeded 0.50, perceived usefulness and perceived ease of use were classified as two factors, and the two factors showed 58.173% factor explanatory power. The first factor consisted of seven items, 'cognitive ease', and the second factor consisted of five items, 'cognitive usefulness'.

In addition, the perceived usefulness of the Cronbach's α coefficient was 0.789, and the perceived ease was 0.878. As a result, since all measurement tools showed a reliability index of 0.70 or higher, the reliability of the measurement tools was judged to be good.

Information utilization satisfaction and policy trust were analyzed in nine items. As a result of the analysis, the *KMO* measure was 0.821, and Bartlett's sphericity verification result was also significant ($p < 0.001$), the factor analysis model was judged to be suitable.

The factor classification was classified as a factor when the factor loading exceeded 0.50, and information utilization satisfaction and policy trust were classified into two factors, and the two factors showed a factor explanatory power of 58.572. The first factor was composed of four items, 'policy trust', and the second factor was five items, 'information utilization satisfaction'.

In addition, the Cronbach's α coefficient was 0.756 for satisfaction with information use and 0.827. As a result, since all measurement tools showed a reliability index of 0.70 or higher, the reliability of the measurement tools was judged to be good.

The statistical analysis method conducted for this study is summarized as follows. First, frequency analysis was conducted to understand the demographic characteristics of the study subject.

Second, an exploratory factor analysis was conducted to analyze the validity of the measurement tool, and the

reliability of the items constituting the factor was analyzed using the Cronbach's alpha coefficient.

Third, descriptive statistical analysis was conducted to understand the level of major variables, and skewness and kurtosis were calculated to determine whether the normality assumption was satisfied. Fourth, Pearson's correlation analysis was conducted to understand the correlation between major variables. Fifth, an independent sample t-test and one-way variance analysis were conducted to verify the difference between major variables according to the demographic characteristics of the study subject. Sixth, structural equation modeling was analyzed to verify the research model. Seventh, bootstrap analysis was conducted using phantom variables to verify the indirect effect between variables. IBM SPSS 22.0 and AMOS 22.0 were used for statistical analysis, and statistical significance was determined based on the significance level of 5%.

4 RESEARCH RESULTS

4.1 Feasibility and Reliability Analysis of Measurement Tools

Smartphone quality factors using 5G technology Exploratory factor analysis was conducted to verify the validity of the measurement tool used in this study. Principle component analysis was used to minimize information loss while extracting factors that explain as much of the variance of the original variables as possible, and Varimax rotation was used to rotate factors until the factor structure was most pronounced while maintaining factor independence.

As for the smartphone quality factor to which 5G technology was applied, one item (information quality No. 4) that hinders validity was excluded, and factor analysis was finally conducted with 11 items. As a result of the analysis, the *KMO* measurement is. It was shown as 910, and Bartlett's sphericity verification results were also significant ($p < 0.001$), the factor analysis model was judged to be suitable.

Factor classification was classified as a corresponding factor if the factor loading exceeds .50, and smartphone quality factors using 5G technology were classified into three factors, and the three factors showed 64.743%. The first factor consisted of 'information quality' with three items, the second factor consisted of 'service quality' with four items, and the third factor consisted of 'system quality' with four items.

In addition, the Cronbach's coefficient was 0.783 for information quality, 0.785 for service quality, and 0.779 for system quality. As a result, the reliability of the measurement tool was judged to be good because all measurement tools showed a reliability index of 0.70 or more.

4.2 Perceived Usefulness and Perceived Ease

For perceived usefulness and perceived ease, factor analysis was finally conducted with 12 items, excluding two items (recognized ease No. 1 and No. 2) that hinder validity. As a result of the analysis, the *KMO* measurement is. It was 899, and Bartlett's sphericity verification results were also significant ($p < 0.001$), the factor analysis model was judged to be suitable.

The factor classification was classified as a corresponding factor if the factor loading exceeded 0.50, and perceived usefulness and perceived ease were classified into two factors, and the two factors showed 58.173% factor explanatory power. The first factor was composed of 7 items, 'perceived ease', and the second factor was composed of 5 items, 'perceived usefulness'.

In addition, the perceived usefulness of Cronbach's coefficient was 0.789, and the perceived ease was 0.878. As a result, the reliability of the measurement tool was judged to be good because all measurement tools showed a reliability index of 0.70 or more.

Table 2 Results of factor analysis of perceived usefulness and perceived ease of use

Category	Main Cause	
	1	3
Perceived Ease 9	0.866	0.075
Perceived Ease 8	0.816	0.205
Perceived Ease 7	0.719	0.328
Perceived Ease 3	0.703	0.067
Perceived Ease 5	0.652	0.346
Perceived Ease 6	0.642	0.399
Perceived Ease 4	0.558	0.423
Perceived Usefulness 2	0.028	0.805
Perceived Usefulness 4	0.217	0.786
Perceived Usefulness 1	0.204	0.770
Perceived Usefulness 3	0.286	0.577
Perceived Usefulness 5	0.383	0.528
An Eigenvalue (Eigen value)	3.894	3.087
Co-Dispersion (%)	32.451	25.722
Cumulation Variance (%)	32.451	58.173
Cronbach's α	.878	.789
KMO = 0.899 Bartlett χ^2 = 1663.229 ($p < 0.001$)		

4.3 Satisfaction with Information Utilization and Policy Trust

Factor analysis was conducted with nine items for information utilization satisfaction and policy trust. As a result of the analysis, the *KMO* measure was 0.821, and Bartlett's sphericity verification result was also significant ($p < 0.001$), the factor analysis model was judged to be suitable.

Table 3 Results of factors analysis of satisfaction with information utilization and policy trust

Category	Main Cause	
	1	2
Policy Trust 1	0.816	0.065
Policy Trust 2	0.804	0.146
Policy Trust 4	0.786	0.175
Policy Trust 3	0.770	0.195
Satisfaction With Information Utilization 2	0.082	0.784
Satisfaction With Information Utilization 3	0.069	0.746
Satisfaction With Information Utilization 1	0.233	0.689
Satisfaction With Information Utilization 4	0.379	0.618
Satisfaction With Information Utilization 5	0.215	0.610
An Eigenvalue (Eigen value)	2.778	2.494
Co-Dispersion (%)	30.866	27.707
Cumulation Variance (%)	30.866	58.572
Cronbach's α	0.827	0.756
KMO = 0.821 Bartlett χ^2 = 881.234 ($p < 0.001$)		

Factor classification was classified as a corresponding factor when the factor loading exceeded 0.50, and information utilization satisfaction and policy trust were

classified into two factors, and the two factors showed 58.572% of factor explanatory power. The first factor consisted of 'policy trust' with 4 items, and the second factor consisted of 'information utilization satisfaction' with 5 items.

In addition, the Cronbach's coefficient was 0.756 for satisfaction with information use and 0.827 for policy trust. As a result, the reliability of the measurement tool was judged to be good because all measurement tools showed a reliability index of 0.70 or more.

4.4 Correlation

Pearson's correlation analysis was conducted to understand the correlation between major variables.

Perceived usefulness is information quality ($r = 0.609$, $p < 0.001$), quality of service ($r = 0.663$, $p < 0.001$), system quality ($r = 0.582$, $p < 0.001$),There was a statistically significant positive (+) correlation with 0.001), and the perceived ease was information quality ($r = 0.398$, $p < 0.001$), quality of service ($r = 0.616$, $p < 0.001$), system quality ($r = 0.403$, $p < 0.001$)There was a statistically significant positive (+) correlation with 001).

Satisfaction with information utilization is information quality ($r = 0.541$, $p < 0.001$), quality of service ($r = 0.513$, $p < 0.001$), system quality ($r = 0.490$, $p < 0.001$). There was a statistically significant positive (+) correlation with 0.001, and perceived usefulness ($r = 0.680$, $p < 0.001$), perceived ease ($r = 0.528$, $p < 0.001$) also showed a statistically significant positive (+) correlation.

Policy trust is information quality ($r = 0.290$, $p < 0.001$), quality of service ($r = 0.436$, $p < 0.001$), system quality ($r = 0.182$, $p < 0.001$). There was a statistically significant positive (+) correlation with 001), and perceived usefulness ($r = 0.360$, $p < 0.001$), perceived ease ($r = 0.650$, $p < 0.001$). There was a statistically significant positive (+) correlation with 0.001), and a statistically significant positive (+) correlation with information utilization satisfaction ($r = 0.402$, $p < 0.001$).

Table 4 Correlation between key variables

Variable	1	2	3	4	5	6	7
1. Information quality	1						
2. Service quality	0.619***	1					
3. System quality	0.625***	0.532***	1				
4. Perceived usefulness	0.609***	0.663***	0.582***	1			
5. Perceived Ease of use	0.398***	0.616***	0.403***	0.584***	1		
6. Satisfaction with the use of information	0.541***	0.513***	0.490***	0.680***	0.528***	1	
7. Policy trust	0.290***	0.436***	0.182***	0.360***	0.650***	0.402***	1

4.5 Suitability of Measurement Model

Before conducting the structural equation model analysis, the suitability of the measurement model was verified through confirmatory factor analysis to confirm whether the observed variables explain the latent variables well. The maximum likelihood method was used to estimate the parameters of the measurement model, and the observed variables were composed of items constituting the latent variables.

Table 5 Goodness of fit of measurement model

Fitness index		Measured value	Criteria for determining suitability
Absolute fit index	$\chi^2(CMIN)p$	940.832 ($p < 0.001$)	$p > 0.05$
	$\chi^2(CMIN)/df$	2.124	$1.0 \le CMIN/df \le 3.0$
	RMSEA	0.060	≤0.08
	RMR	0.044	≤0.08
	GFI	0.827	≥0.80–0.90
Incremental fit index	*NFI*	0.817	≥0.80–0.90
	TLI	0.880	≥0.80–0.90
	CFI	0.893	≥0.80–0.90

Looking at the fit of the measurement model, the value 22 ($df = 443$) is 940.832 ($p < 0.001$). Although it was shown as 001, other goodness-of-fit indices were identified because the 2 value was greatly influenced by the size of the sample and the complexity of the model [11]. As a result, 22(*CMIN*)/*df* is 2.124, showing a value between 1 and 3, and *RMSEA* = 0.060, *RMR* = Suitable for 0.044, *GFI* = 0.827, *NFI* = 0.817, *TLI* = 0.880, *CFI* = 0.893. It was judged that the measurement model was suitable as all 893 showed good levels.

4.6 Structural Model Verification

In the structural model, the significance of the direct influence relationship between latent variables was verified.

As a result, quality of service ($\beta = 0.545$, $p < 0.001$) and system quality ($\beta = 0.342$, $p < 0.001$) was found to have a statistically significant positive (+) effect on perceived usefulness, and information quality did not have a significant positive (+) effect on perceived usefulness.

Service quality was found to have a statistically significant positive (+) effect on perceived ease ($\beta = 0.002$, $p < 0.001$), information quality and system quality did not have a significant positive (+) effect on perceived ease.

Perceived usefulness ($\beta = 0.746$, $p < 0.001$) and perceived ease ($\beta = 0.171$, $p < 0.05$) was found to have a statistically significant positive (+) effect on information utilization satisfaction.

Perceived ease ($\beta = 0.780$, $p < 0.001$) and satisfaction with information utilization ($\beta = 0.314$, $p < 0.05$) was found to have a statistically significant positive (+) effect on policy trust, and perceived usefulness did not have a significant positive (+) effect on policy trust.

Table 6 Results of direct effect analysis of structural model

Path			Non-standardized coefficient	Standard error	Standardization coefficient	Test statistics	Significance probability
Information quality	→	Perceived usefulness	0.057	0.121	0.059	0.472	0.637
Quality of service	→	Perceived usefulness	0.531	0.103	0.545	5.141***	<0.001
System Quality	→	Perceived usefulness	0.382	0.116	0.342	3.280**	0.001
Information quality	→	Perceived ease	–0.385	0.183	–0.360	–2.104*	0.035
Quality of service	→	Perceived ease	1.082	0.183	1.002	5.916***	<0.001
System Quality	→	Perceived ease	0.068	0.152	0.055	0.448	0.654
Perceived usefulness	→	Satisfaction with information utilization	0.620	0.078	0.746	7.973***	<0.001
Perceived ease	→	Satisfaction with information utilization	0.128	0.052	0.171	2.465*	0.014
Perceived usefulness	→	Policy trust	–0.448	0.180	–0.368	–2.481*	0.013
Perceived ease	→	Policy trust	0.855	0.110	0.780	7.756***	<0.001
Satisfaction with information utilization	→	Policy trust	0.459	0.231	0.314	1.990*	0.047
* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$							

4.7 In the Modified Model, the Significance of the Direct Influence Relationship between Latent Variables Was Verified

As a result, quality of service ($\beta = 0.548$, $p < 0.001$) and system quality ($\beta = 0.392$, $p < 0.001$) was found to have a statistically significant positive (+) effect on perceived usefulness, and service quality was found to have a statistically significant positive (+) effect on perceived ease ($\beta = 0.743$, $p < 0.001$).

Perceived usefulness ($\beta = 0.736$, $p < 0.001$) and perceived ease ($\beta = 0.177$, $p < 0.05$) was found to have a statistically significant positive (+) effect on information utilization satisfaction.

Perceived ease of use ($\beta = 0.790$, $p < 0.001$) and satisfaction with information utilization ($\beta = 0.324$, $p < 0.05$) was found to have a statistically significant positive (+) effect on policy trust, and perceived usefulness did not have a significant positive (+) effect on policy trust.

Table 7 Summary of hypothesis test results and mediated effect test results

Hypothesis								Result
1-1	Service quality will have a significant positive (+) effect on perceived usefulness.							Adopt
1-2	System quality will have a significant positive (+) effect on perceived usefulness.							Adopt
1-3	Service quality will have a significant positive (+) effect on perceived ease.							Adopt
2-1	The perceived usefulness will have a significant positive (+) effect on information utilization satisfaction.							Adopt
2-2	The perceived ease of use will have a significant positive (+) effect on information utilization satisfaction.							Adopt
3-1	The perceived usefulness will have a significant positive (+) effect on policy trust.							Dismiss
3-2	The perceived ease of use will have a significant positive (+) effect on policy trust.							Adopt
3-3	Satisfaction with information use will have a significant positive (+) effect on policy trust.							Adopt
Mediated effect path								Result
4-1	Service quality	Service quality	perceived usefulness			→	Policy trust	Dismiss
4-2	Service quality	→	Perceived usefulness	→	Satisfaction with the use of information	→	Policy trust	Dismiss
4-3	Service quality	Service quality	Perceived availability			→	Policy trust	Adopt
4-4	Service quality	→	Perceived availability	→	Satisfaction with the use of information	→	Policy trust	Dismiss
5-1	System quality	→	Perceived usefulness			→	Policy trust	Dismiss
5-2	System quality	→	Perceived usefulness	→	Satisfaction with the use of information	→	Policy trust	Dismiss

5 DISCUSSION AND CONCLUSION

This study was conducted using an expanded technology acceptance model to find out how quality characteristics and personal characteristics of smartphones with 5G information technology affect reuse intention and continuous use. First, the background and purpose of the study were described, the research method and the thesis composition of the previous paper were re-established, and before this study, the concept and theoretical characteristics of the independent major variables of the previous paper were studied, analyzed, and grasped.

Next, in previous studies, the relationship between independent variables was examined through a theoretical review, and based on this, 12 sub-factors constituting the independent variable were derived, measurement items were constructed through operational definition along with the dependent variable, and selected samples were verified. The objectivity of the study was secured through validity and reliability analysis of the measurement results, the study model was schematized with an expanded technology acceptance model, hypotheses were established, and major variables were empirically studied and analyzed with correlation. Exploratory factor analysis was performed to verify the validity and reliability of the measurement tool used in the study.

The principal component analysis was used to minimize information loss by extracting factors that account for as many of the original variables as possible, and to use a Verimax rotation analysis that rotates factors until the factor structure is most pronounced while maintaining factor independence. Factor classification consisted of one factor when the eigenvalue was 1 or more, and the factor load was. If it exceeded 40, it was classified as a corresponding factor.

To achieve the purpose of this study's extended technology acceptance model, technical standards and

terminology for 5G information technology were reorganized, and information quality was selected as independent variables to analyze the difference and effect of smartphone users' interests.

The trend analysis and improvement model of the expanded technology acceptance model were applied by referring to 10 overseas papers and 8 domestic papers. The effect of smartphone quality characteristics on perceived usefulness, perceived usability, and perceived playfulness was verified in detail.

Information data collection was surveyed on 519 5G users, mainly in groups using smartphones, with 259 men (49.9%), 260 women (50.1%), 130 people (25.0%) in their 20s and younger, 187 people (36.0%), 142 (27.4%) in their 40s and older (11.6%). There were 89 (17.1%) who graduated from high school or below, 338 (65.1%) who attended or graduated from vocational colleges and universities, 92 (17.7%) who attended or graduated from graduate schools, 25 civil servants (4.8%), 61 professionals (11.8%), 262 (50.5%), 71 (13.7%) and 100 others (19.3%).

The research model was set based on the extended technology acceptance model, and the results of verifying the model are as follows. The 5G user-centered research model has good exploratory factor analysis of 10 factors of network quality, cost of use, smartphone quality, social influence of individual characteristics, personal innovation, self-efficacy, perceived usefulness, perceived play, and continuous use intention. It was found to be 950, and Bartlett's sphericity test result was also significant ($p < 0.001$), so the factor analysis model was suitable, and the eigenvalue was 1.380 or higher, and the factors of 10 items showed 78.378% explanatory power, which was overall suitable and good. A path that affects continuous use intention was established by mediating the cognitive process and use satisfaction. The suitability of the model was found to be good, so it was judged that the research model was suitable.

As a result of path analysis of the research model, network quality and personal innovation were found to have a significant positive (+) effect on perceived usefulness, perceived ease, and perceived play, and user satisfaction had a significant positive (+) effect on user satisfaction.

It was verified that the positive (+) indirect effect of smartphone quality and social influence on continuous use intention was statistically significant, the positive (+) indirect effect of smartphone quality on continuous use intention was significant, and the positive (+) indirect effect of social influence on continuous use intention was verified to be significant.

In addition, smartphone quality and social influence were found to have a statistically significant positive (+) effect on perceived usefulness, smartphone quality and social influence were found to have a significant positive (+) effect on perceived play, and perceived usefulness, perceived ease, and perceived play had a significant positive (+) effect on continuous use.

5.1 Research Summary

This research was conducted using an extended technology acceptance model to investigate how the quality and personal characteristics of smartphones to which 5G information technology is applied have an effect on the intention to reuse and continuous use. First, the background and purpose of the study were described, the research method and the composition of the thesis were redefined, and prior to this study, the concepts and theoretical characteristics of the main independent variables of the previous thesis were studied, analyzed, and identified.

Next, the relationship between independent variables was examined through theoretical review in previous studies, and 12 sub-factors constituting independent variables were derived based on this, and measurement items were constructed through operational definition along with dependent variables, and verified for selected samples. The objectivity of the study was secured through the analysis of the validity and reliability of the measurement results, and after schematizing the study model into an extended technology acceptance model and establishing a hypothesis, the main variables were empirically studied and analyzed. Exploratory factor analysis was performed to verify the validity and reliability of the measurement tool used in the study. Among the factor analysis methods, principal component analysis was used to minimize information loss while extracting factors that explain as many parts of the variance of the original variables as possible, and to rotate the factors until the factor structure is most pronounced while maintaining factor independence. The factor classification was composed of one factor when the eigenvalue was 1 or more, and the factor loading amount was. If it exceeded 40, it was classified as a corresponding factor. In order to achieve the purpose of this study's expanded technology acceptance model, technical standards and technical terms for 5G information technology were reorganized, and in order to analyze the differences and effects of perceived ease of use, perceived usability, and perceived playability of smartphone users. A total of 18 previous papers were applied to the trend analysis and improved model of the extended technology acceptance model by referring to 10 overseas papers and 8 domestic papers, and the effect of smartphone quality characteristics on perceived usefulness, perceived usability, and perceived play was verified in detail. Information data collection targets 519 5G users were surveyed, centering on groups using smartphones, and the gender of the general characteristics of the study was 259 (49.9%) for men, 260 (50.1%) for women, 130 (25.0%) for those in their 20s and under, 187 (36.0%) for those in their 30s, 142 (27.4%) for those in their 50s and 60 (11.6% for those in their 50s. In terms of education, 89 students (17.1%) graduated from high school or less, 338 students (65.1%) enrolled or graduated from vocational colleges and universities, 92 (17.7%) enrolled or graduated from graduate schools, 21 (48%) professional workers, 262 (50.5%) office workers, 71 (13.7%) and 100 others (19.3%). The research model was set based on the expanded technology acceptance model, and the results of verifying the model are as follows. In the 5G user-centered research model, 5G service quality characteristic factors and personal characteristic factors are good, excluding 5 factors that hinder validity, network quality, smartphone quality, social influence of individual characteristics, personal innovation, self-efficacy, perceived

usefulness, perceived ease, perceived playfulness, and *KMO* measurement. It was found to be 950, and Bartlett's sphericity verification result was also significant ($p < 0.001$), the factor analysis model was suitable, the eigenvalue was 1.380 or higher, and the factors of 10 items showed 78.378% explanatory power, which was suitable and good overall. A path that affects the intention of continuous use was established by mediating the cognitive process and satisfaction of use. The fitness of the model was found to be at a good level, and the research model was judged to be suitable.

As a result of path analysis of the research model, network quality and personal innovation were found to have a significant positive (+) effect on perceived usefulness, perceived ease, and perceived playfulness, and user satisfaction had a significant positive (+) effect on user satisfaction. It was verified that the positive (+) indirect effect of smartphone quality characteristics and social influence on the perceived usefulness was statistically significant, the positive (+) indirect effect of smartphone quality on the perceived playability was significant. In addition, smartphone quality and social influence were found to have a statistically significant positive (+) effect on perceived usefulness, smartphone quality and social influence were found to have a significant positive (+) effect on perceived playfulness, perceived ease of use, and perceived play.

5.2 Discussions and Implications

The result that there was a significant difference in the path centered on the users of the 5G service in this study supports the research results of [9] Since the factors influencing the use of the service differ depending on whether or not the service is used, other variables should be considered in order to increase the intention to use the service for each user group. It has been verified that it is necessary to increase user satisfaction in order to induce 5G users to continue to use 5G services, and to increase user satisfaction, it is necessary to increase perceived usefulness, perceived ease, and perceived playability. In addition, in order to increase perceived usefulness, perceived ease of use, and perceived playability, improving network quality and personal innovation was suggested as a valid way. In other words, higher network quality applying 5G service technology will improve perceived usefulness, perceived ease, and perceived playability, increasing satisfaction with information utilization, and eventually positively influencing users' intention to use 5G service. Therefore, it is necessary to invest at the national level to improve the quality of online platforms. However, Korea, which commercialized 5G services for the first time in the world, also lacks foundation, platform, and application-related pools. In the future, it is necessary to improve service quality by investing in facilities related to 5G. The results of this study are expected to be used as evidence for improving smartphone quality applying domestic 5G service technology, intention to use subscriber information, and base data for information utilization. Therefore, for the development of the fourth industry at the national level, it is necessary to establish and actively support multimedia Big DATA Service that specializes in real-time interaction such as artificial intelligence (AI), augmented reality (AR), virtual reality (VR), and real-time online games. In order to improve the quality of information utilization, the information center, which is always managed, needs to be fully confirmed even if it is inconvenient to train and use certification and verification agencies along with AI and Big DATA security technology. In addition, an institution that can always monitor and control defective information by publicly authenticating the information and communication countries and platforms based on artificial intelligence networks will be needed, and short-term, medium- and long-term plans should be made every year. And it can be seen that this study contributed to the following aspects. First, the actual needs and phenomena were sufficiently reflected because the actual experiences of smartphone users related to actual information were analyzed. Second, the perceived usefulness, perceived ease, and perceived playability according to information quality were analyzed to derive the importance between factors of information quality as an expanded technology acceptance model. Third, by examining the satisfaction level of information use according to the perceived usefulness, ease of use, and perceived playability, it was confirmed with an expanded technology acceptance model which part to focus more on to increase the satisfaction level of information use. Fourth, by analyzing the impact on the intention to use information, it researched, analyzed, and confirmed what factors should be emphasized, analyzed, and nurtured in order to increase the intention of existing service users to use and secure new users in the future. This study is the world's first paper on 5G information technology built on April 3, 2019, and summarizes the latest trends and issues related to the opening of information and communication by using smartphone-applied information technology. Based on this, 5G information technology was defined as a broad concept from basic definition, and the information quality of 5G technology was analyzed and summarized by considering previous studies. It is also important to investigate how characteristic factors such as information quality, system quality, and service quality influence from the perspective of the expanded technology acceptance model and how these influencing factors affect perceived ease, perceived usefulness, and perceived play.

5.3 Contributions and Important Points of Research

It can be seen that this study contributed to the following aspects.

First, the actual information-related experiences of smartphone users were analyzed and the actual needs and phenomena were sufficiently reflected and compared.

Second, the importance between information quality factors was compared and derived with an expanded technology acceptance model by analyzing the perceived usefulness and perceived convenience of use according to information quality.

Third, we investigated the satisfaction level of information use according to perceived usefulness and ease of use, and compared and analyzed areas that should focus more on information use satisfaction through an expanded technology acceptance model.

Fourth, by analyzing the impact on the reliability of government policies, the research analysis confirmed what factors should be emphasized and fostered in order to increase trust in government policies in the future.

This study summarizes the latest trends and issues related to the opening of information policy as a research paper on how information use and user satisfaction will affect government policies using smartphone-applied information technology without existing research on 5G in the world. Based on this, 5G information technology was defined as a broad concept from the basic definition, and the information quality of information transmission was summarized through insufficient prior research.

It is also important to investigate how characteristic factors such as information quality, information system quality, and information service quality affect user satisfaction and policy trust in perceived ease of use and perceived ease of use.

As such, the public's requirements for new information quality, information application services, and product development continue to increase, but actual facility investment and support for it are still insignificant, and it is necessary to maximize the infinite value of use directly or indirectly. In order to maximize the value of using such information, it is necessary to establish a center that develops and provides compatible information systems for convenience of use and to connect to various infrastructure and platforms, and to continuously improve the use environment. Another problem of the domestic public information and communication centers is that information opening is still limited due to information security, and above all, the issue of establishing national standards for information formats such as standardization of information should be continuously researched and analyzed.

In addition, even if information delivery customer services are commercialized into the private sector, it is very important to ensure the quality, certification, and verification of actual information services as similar information services may become competitive.

From the perspective of raising the government's public awareness, improving people's understanding of government policies, and enhancing people's feelings, it is desperately necessary to actively implement public relations policies and develop more active public relations activities. In order to improve the public's sense of information service through 5G, it is important for all citizens to feel the public's public interest, and by holding ideas and competitions on a regular basis, various events and plans should be published, and public relations and educational activities should be systematically conducted. The use and ease of information services affect the satisfaction of use, and this satisfaction of use affects the trust of policy, so the government should establish a national strategy for the promotion of information use and the satisfaction of users and developers.

6 REFERENCES

[1] Davis, F. D. (1986). *A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results*. Sloan School of Management, Massachusetts Institute of Technology.

[2] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319-340, JSTOR 249008, S2CID 12476939. https://doi.org/10.2307/249008

[3] Fishbein, M. & Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior Addison-Wesley. Reading, MA.

[4] Xia, W. & King, W. (1996). Interdependency of the Determinants of User Interaction and Usage: An Empirical Test. *ICIS 1996 Proceedings*, 1.

[5] Koo, J.-C. (2008). User Selection Characteristics Analysis Technology Product: Development of Technology Product Selection Model. *Doctoral dissertation* at Graduate School of Kyung Hee University.

[6] Son, S.-H. (2011). A study on smartphone adoption behavior of early users using technology acceptance models. *Korean Journalism Society, 55*(2), 227-251.

[7] Kim, S.-H. (2016). A Study on Secure Data Management and Access Control in Cloud Computing Environments. *Doctoral dissertation* at Graduate School of Soonchunhyang University.

[8] Lee, S.-T. (2017), A study on the effect of using an online start-up education platform on college students' willingness to start a business: focusing on the moderating effect of the degree of self-determination. *PhD thesis* at Soongsil University Graduate School.

[9] Jeong, M.-S. (2020). Study on the factors affecting the usage intentions of 5G mobile communication service. *Doctoral dissertation* at Yonsei University Graduate School.

[10] Byeongryul, B. (2014). Amos 21 Structural Equation Modeling, 2014 Seoul: Cheongram. https://doi.org/10.1080/10705511.2014.919820

[11] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis* (Vol. 6).

Author's contacts:

Chil-Yuob Choo, Dr.
Donghyun Professional Engineers Office,
Kolon Science Valley 1st 11th floor, 43, Digital-ro 34-gil, Guro-gu,
Seoul, South Korea
Smartphone +82-10-7470-6688 / Tel 02-2025-2720 / Fax 042-271-9797
choocy@naver.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230416193006

Original scientific paper
Received: 2023-04-16, Accepted: 2023-04-20

The Assessment of Robotic Process Automation Projects with a Portfolio Analysis: First Step - Evaluation Criteria Identification and Introduction of the Portfolio Concept

Bernhard Axmann*, Harmoko Harmoko, Rahul Malhotra

Abstract: RPA's (Robotic Process Automation) usage in organizations has rapidly increased in recent years; as a result, companies have developed high expectations from this technology. However, according to Ernst & Young (E&Y), 30-50% of observed RPA projects initially fail and reveal several risks, which lead to investment losses. Consequently, the RPA project is prematurely retired, and the company is back to the manual process. This premature retirement is mainly because of wrong process selection and the not sufficient company automation (RPA) maturity. Therefore, this paper will introduce the concept of an RPA Portfolio, which will assess the complexity of business processes with a company's automation (RPA) maturity. The RPA Portfolio is a new innovative concept to simplify and visualize the business process selection for RPA projects, and will help to introduce successfully the right RPA projects.

Keywords: automation maturity; business process; portfolio; robotic process automation; technology assessment

1 INTRODUCTION

Robotic Process Automation (RPA) is a technology that emulates humans in executing repetitive and rule-based tasks in the office [1]. In comparison to other modes of business process automation, RPA bots act at the front-end level of the application such as understanding the text on the desktop screen, completing the keystrokes on the keyboard, identifying, and extracting data on the website or digital document, and other activities which require human-computer interaction [2, 3].

Currently, RPA is not wanted by large companies only, but also by SMEs (Small and Medium Sized Enterprises) [4]. Adopting RPA in SMEs is like hiring assistants for a fraction of the cost, and empowering employees to do high-value-added tasks [4]. SMEs can leverage the instant benefits of RPA such as increasing productivity, reducing operational costs, and improving the quality of customer services. In a nutshell, RPA can benefit SMEs to scale up their operations and revenues [5].

Once the foundation of RPA is started, any bot implemented will undergo a maintenance and monitoring phase [6, 7]. The Maintenance cost is sometimes higher than the implementation cost. It depends on the number of bots and the complexity of the process [8]. This condition burdens companies with a longer period of return on investment (ROI). On the other hand, without maintenance, the error rate of bots will increase, and more time will be spent fixing the errors [9]. This dilemma leads decision-makers to drop Bots before reaping the benefits and returning to manual processes [10, 11]. To avoid this circumstance, companies should select the right process by considering its complexity as well as the maturity of their automation. This research will provide a scientific approach that can help the company to select the best business process and assess its automation maturity, by answering the following research questions:

- How to identify the process selection criteria to evaluate the RPA project?
- How to identify the automation maturity criteria of an organization?

2 STATE OF KNOWLEDGE

2.1 The Challenges of RPA and Automation Maturity in the Organization

Companies have been increasingly adopting RPA in the last few years for automating their time-consuming, error-prone, and highly expensive business processes [12]. However, the challenges to adopting RPA are also enormous, which are defined into three dimensions:

- **Organizational:** (1) the inability to prioritize RPA initiatives related to process selection, (2) aversion to risk, (3) limited RPA skills and talents, and (4) little sense of urgency [10].
- **Technological:** (1) concerns about cybersecurity and data privacy, (2) difficulty in scaling applications, and (3) the difficulty of deciding on the best applications [13, 14].
- **Financial:** (1). High Implementation Costs, (2). Making better and more convincing business cases and (3) regulatory constraints [8, 15].

When all the challenges have been accounted for, the next step is to measure the automation maturity of the organization, so the company can understand where they currently stand and where they need to go in the future. According to Kumar (2016), The Company's automation maturity consists of five levels:

- **Level 1 Ad hoc**: the automation is not a planned activity, but it happens in small-small pockets across teams at different levels on a need basis. There is no management, and budget allocation.
- **Level 2 Opportunistic:** the automation is addressed to the specific issue, such as providing 24/7 service. At this level, the organization starts to create formal evaluation & governance tools and budget allocation.

- **Level 3 Systematic:** at this level, the organization evaluates the automation initiatives, which means exploring areas of automation to achieve defined levels of accuracy, productivity, and quality.
- **Level 4 Institutionalized:** at this level, automation initiatives are taken up at the organization with a portfolio of different solution providers and tools. Automation becomes the way of life and there is a documented automation strategy. It includes the usage of advanced technologies such as machine learning, artificial intelligence, optical character recognition, natural language processing, chatbots, and so on.
- **Level 5 Adaptive:** at this highest level of automation, the automated process becomes adaptive to the changing demands of the business. The selected automation technology has the capabilities of self-learning, self-healing, auto-scaling, auto-optimization, and so on [16].

2.2 The Overview of RPA Process Selection

Before implementing RPA, organizations must be able to identify what types of processes can be automated. Fung (2014) suggested that non-complex processes with high repetition and high human error should be prioritized. Meanwhile, Tripathi, (2018) required more detail, that the process should have (1) repetitive steps, (2) time-consuming, (3) high risk, (4) low-added value, and (5) involving many people in multiple steps. Nevertheless, the other process is deserving to be automated, if (1) it is well-defined by rule-based task, (2) it has logic steps, (3) the input of the process can be diverted to the software system, (4) the output process is accessible and (5) the benefit of automation is higher than the cost [1, 17].

Doguc (2020) suggests that volume is a considerable factor to prioritize which processes deserve to be automated by RPA. The process with high volume and repetition should be selected first. While Capgemini Consultant (2017) and Siderska (2020) argue that the process complexity has a significant role to determine the best process for RPA [18–20].

3 RESEARCH METHODOLOGY

3.1 Systematic Literature Research (SLR)

According to Huelin and Payne (2015), SLR is a type of scientific study that is designed to address a specific research problem or question by collecting all the available information on a topic that is being defined at the outset by absolute inclusion and exclusion criteria [21, 22]. This research used SLR to synthesize existing scientific contributions about the topic: *the failures and challenges of RPA projects*, with searching queries: *RPA, robotic process automation, RPA benefits, robotic process automation benefits, RPA used cases, robotic process automation used cases, RPA challenges, robotic process automation challenges, RPA failures, and robotic process automation failures*.

SLR started by focusing on Computer Science databases such as ACM Digital Library and IEEE Xplore. Later, the literature search was extended to multidisciplinary databases such as Science Direct, Research Gate, and SpringerLink. Following this strategy, 574 academic contributions were found. Through title, keyword, and abstract analysis, 56 contributions remained. Subsequently, through full-text analysis, SLR identified only 13 contributions that are relevant to answer the research questions (see Fig. 1).

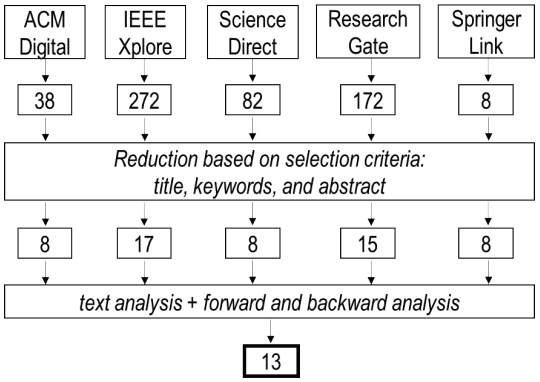


Figure 1 Result of the Systematic Literature Research

Since only $n = 13$ contributions dealing with challenges and failures were found, so, this was extended to Google search, where grey literature such as statistic reports, articles, and white papers from consulting firms and software development companies, describe real-world RPA challenges and failure reasons. As a result, in total $n = 35$ contributions were listed from academia and practitioner.

3.2 Expert Interviews and Questionnaires

Using RPA or Robotic Process Automation as a keyword, several experts were identified from the LinkedIn and LinkedIn groups such as "Robotic Process Automation (RPA), RPA (Robotic Process Automation) professionals Australia, Robotic Process Automation (RPA) / IPA / Artificial Intelligence / Machine Learning / Analytics, RPA Gurus Direct – India, RPA (Robotic Process Automation), VBA Excel & Access, Training, Jobs, Process Automation Solution" but only a few agreed to support this research topic that is 20 and out of it only 12 agreed for a detailed conversation over the telephone, but the expert questionnaires were filled by everybody. Due to the novelty of the topic and the low number of potential interviews locally, only telephonic interviews via Microsoft Teams and Zoom were opted to enable synchronous communication.

The interviews have not been recorded due to privacy constraints, but their ideas have been captured verbally and define the criteria for evaluating business process BP complexity and the criteria for measuring RPA's maturity model (see Tab. 1).

All the interviewees were from these backgrounds, some were Internal Consultants, some External Consultants, RPA Providers, RPA developers, and RPA leads. The average company size of every interviewee was more than 500 employees. The interviews were held from the time span of April to June 2022.

Table 1 Expert Interviewees

Interview	Date	Role	Company Size
I1	09.05.2022	Global Head - Intelligent Automation	48,500
I2	09.05.2022	Director Intelligent Automation	5,000
I3	13.05.2022	Principal Consultant	950
I4	16.05.2022	Internal Consultant	310,000
I5	18.05.2022	Senior Manager - Digital Transformation	55,000
I6	25.05.2022	RPA Lead	112,000
I7	31.05.2022	RPA Lead	29,800
I8	06.06.2022	Chief Transformation Officer - Intelligent Automation	130,000
I9	08.06.2022	RPA Lead	34,200
I10	15.06.2022	RPA Expert	1,000
I11	22.06.2022	External Consultant	8,000
I12	22.06.2022	Internal Consultant	500

4 RESULTS

4.1 General Overview

RPA is valuable to organizations that have many manual tasks typically performed on a daily, weekly, or monthly basis. However, not all organizations are successful to adopt RPA into their business process. It is a common tendency for organizations to shine on the R of Robotics and A of Automation but forget the P of Process in R-P-A).

According to SSON Analytics and Hofmann et al, 2019, "wrong process selection" is the leading cause of RPA project failures followed by "wrong tool selection", "insufficient change management support", and "lack of experience" [23, 24]. Shortlisting the wrong process to automate can cause the greatest headaches and is generally related to a highly complex, multi-step process that may contra productive with the goal of automation [25]. This condition will tarnish RPA's good name and be considered a technological "failure" which should have been avoided with more strategic anticipation in evaluating the process.

4.2 Criteria to Evaluate Business Process Complexity

Different RPA vendors have their own way of evaluating the business process complexity. Therefore, it is important that RPA should not simply be implemented to support the "as-is" process, without examining the current processes and identifying areas for improvement [26]. Examining the process from end-to-end with a clear focus on automation can help identify steps or sub-processes, which need to be simplified, standardized, and eliminated. Organizations should evaluate current business process's performance and review relevant business practices for immediate impactful improvements.

Table 2 Criteria to Evaluate Business Process Complexity

Evaluation Criteria	Definition	Effect on BP Complexity
Input Data Type [17] [Interview]	How much percentage of the data is structured and unstructured? It also specifies whether the data is digitalized or non-digitalized.	The higher the value of the Input data type, the lower will be the value of BP complexity (Inverse Trend)
Process Stability [3] [Interview]	How stable a process is in terms of its technicality, functioning, and reliability?	The higher the value of process stability, the lower will be the value of BP complexity. (Inverse Trend)
Number of Applications [1] [Interview]	How many numbers of applications or tools are involved in the process? The Applications can be MS Office, SAP, CRM and so on.	The higher the value of the number of applications, the higher will be the value of BP complexity. (Direct Trend)
Application Stability [14] [Interview]	How stable application(s) to run the process.	The higher the value of applications stability, the lower will be the value of BP complexity. (Inverse Trend)
Number of Process Steps [1, 24] [Interview]	How many steps are involved in a process for its completion?	The higher value of number of steps, the higher will be the value of BP complexity. (Direct Trend)
Complex Decisions [17, 19, 20] [Interview]	How is the decision complexity in each stage of the process? The simple decision process is the straightway process from one-step to another step (the complex decision process = the complex process).	The higher the value of complex decisions, the higher will be the value of BP complexity. (Direct Trend)
Standardization [1, 3] [Interview]	How much percentage of a rule-based or standardized process without individual justification?	The higher the value of standardization, the lower will be the value of BP complexity. (Inverse Trend)
OCR (Optical Character Recognition) [3, 14] [Interview]	How OCR involve in the process (the amount of document reading which is required in a process)?	The higher the value of OCR, the higher will be the value of BP complexity. (Direct Trend)
Process Frequency [1, 17, 18] [Interview]	How many times does the process repeat itself on a daily, weekly, or monthly basis?	The higher is the value of process frequency; the lower will be the value of BP complexity. (Inverse Trend)
Manual Hours [1] [Interview]	How many manual hours does a human workforce spend on that process?	The higher the value of manual hours, the lower will be the value of BP complexity. (Inverse Trend)

The right process for RPA is the one that follows clearly defined business rules with minimal interruptions/ manipulations or exception handling. In summation, the data should be structured, digitalized, easily accessible, and logically formatted. Moreover, when identifying perfect candidate business processes for automation, rigid applications with no integration capabilities or applications, which are not stable, should be avoided.

When starting an RPA journey, low-complexity processes provide quick wins (low-hanging fruits) to demonstrate the rationality of RPA and help out building confidence and internal buy-in for the technology [19]. RPA is naturally a "strategically lean step" for companies whose business processes are supported through Shared-services or Business Process Optimization (BPO).

There are certain identified criteria that will evaluate the complexity of a Business Process. These criteria are very different from each other, for some criteria, a higher value will mean lower BP Complexity (and vice-a-versa) and for some other criteria, a higher value will mean higher BP Complexity (and vice-a-versa) which is negative while selecting a process for automation. Tab. 2 describes all the criteria identified from the systematic literature review (SLR) and expert interviews. Mostly the criteria identification came from the literature and the definition was created by literature and interview. These criteria play a role in assessing the complexity of a business process.

Every time, lower complexity processes should be selected at least during the initial start of the RPA journey and when a certain level of automation maturity has been achieved, then the company can automate the higher complexity processes with consideration of process viability and Return on Investment (ROI).

4.3 Criteria to Measure RPA Maturity

One of the risks associated with scaling up the RPA after initial deployment is the underutilization or overutilization of the company's present automation maturity or RPA maturity. So, ***what exactly is meant by maturity automation?*** It is a maturity model that evaluates an organization's ability for continuous improvement in the automation area. It is done by accessing the business against a certain set of characteristics or criteria. Over a while, companies can re-evaluate their automation maturity to measure progress and continue to drive toward perfection. Three connected areas form the basis of any technique for achieving automation maturity:

1. **Process Viability** or the first step towards a high-quality process is to examine and map out how a particular process works i.e. who is involved, what is involved, what works well and what does not.
2. **Process Automation**: Robotic Process Automation works from the digitalization of manual and paper-based processes to more complicated processes or projects that impact the synergy of human labor, processes, and automation systems.
3. **Process Improvement:** All automation projects should be seen as iterative and continuous processes. Analyses and insights into process performance should be the fuel for continuous process improvement and investment [27].

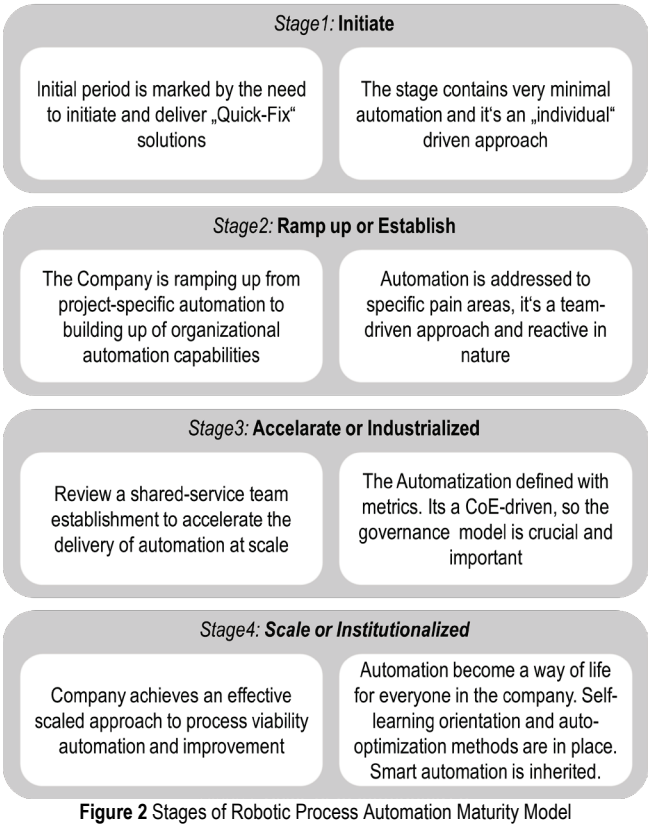


Figure 2 Stages of Robotic Process Automation Maturity Model

Companies using process automation more extensively establish an organization-wide, standardized, people-centric approach. However, less mature or "emerging" companies consider process automation to be impromptu, unmatched, process-centric projects predominantly geared around fixing a specific problem. The following is a further explanation of how established and emerging organizations differ in their view of the automation process.

Emerging Organizations: Organizations see process automation as a one-off project case, whether it is a "fire-fighting" meaning fixing a broken process, or an extemporary investment. It is an approach to fixing an immediate problem. It of no doubt that such kind of initiative produces a greater value to the organization and good ROI but still the company scratches only at the surface and doesn't produce a great benefit [27]. These types of organizations have a process-centric approach, where the automation happens on a process-by-process basis [27].

Established Organizations: Organizations sees process automation as an organization-wide, centralized, and standardized approach. The projects are planned strategically and systematically in order to scale up the technology. A culture is created in the organization where everybody is encouraged and empowered to pro-actively contribute to process automation. It is more like a people-centric approach as everyone in the organization is involved. This centralized body that takes ownership over the process automation decisions is called the Centre of Excellence (CoE). This body

helps to identify & update the processes, develops areas of improvement, and improve the skill and automation capability of people/worker in the organization [27].

In this research, four stages of the Robotic Process Automation Maturity model have been identified and evaluated against the criteria. Each stage is more advanced on the road to maturity than the previous stage, in terms of process viability, automation, and improvements (see Fig. 2). As the company matures, it moves from one stage to the next and evolves from a process-centric focus to people-centric culture [10, 27]. The Robotic Process Automation Maturity Model estimates and describes fourteen unique criteria to identify a company at a specific stage. Each of the four stages provides a measuring scale to determine each criteria's positioning.

Tab. 3 demonstrates the fourteen criteria along with their definitions. The criteria are defined by a systematic literature review and expert interviews. Some criteria have inferential or quantitative value for all stages, which means they can be easily measured in a specific value. However, there are also some criteria that have descriptive or qualitative values, which means that each stage must be quantified based on textual values so that it becomes measurable.

Table 3 Fourteen Criteria Measuring Robotic Process Automation's Maturity Model

Criteria	Definitions
Outlook on Process Automation Technology [3] [Interview]	The criteria measures to how a company views on this low-code automation technology.
View on Low-Code Automation Provider(s) [Interview]	The criteria refers to how RPA vendor(s) or supplier(s) is selected by the decision maker for the process automation.
Disposition on Process Automation needs [13, 24] [Interview]	The criteria evaluates how process automation is viewed and used within an organization.
Deployment Purpose [13, 24, 26] [Interview]	The criteria looks at the reasons why processes are being automated. Across all stages, the reason to why an organization adopts a low-code automation may differ. In the initial stage, a team or an individual may seek out a tool to get a specific job done fast. However, as an organization matures, this approach is then standardized across different departments. The objective becomes clear on functional and cross-departmental needs to enable an ongoing cycle of digital transformation.
Number of Processes Automated with Low-Code Automation [27] [Interview]	The criteria measures an organization's level of process automation adoption. It includes assessing an organization's use of RPA. It also measures how an organization views on process automation.
Number of Departments with Access to Process Automation outcomes [27] [Interview]	The criteria reflects the ability of an organization to scale robotic process automation across entire organization / department.
Process Mapping Initiatives Connected to Process Automation [27] [Interview]	The criteria measures how process mapping is available within an organization. Typically, process mapping occurs within the line of business and ranges somewhere between the use of enterprise-grade process mapping technology (mature) and the use of less scalable methods such as Excel or Visio (emerging).
Number of Processes Mapped with Process Mapping Technology [27] [Interview]	The criteria measures the number of end-to-end processes that a company has mapped and are currently managing it. The earlier stages include a learning curve on how best to map and understand the process. The later stages recognize the importance of discovering and mapping process.
Number of Departments with Access to Process Mapping Technology [27] [Interview]	The criteria measures the number of departments with access to process mapping technology. It indicates the ability of an organization to scale process mapping across many departments in the company. If a single department uses a process mapping technology for themselves that is a good start, however the ability to gain maximum ROI, comes with scaling process mapping technology across other areas in the company.
Enterprise Strategy on Right Mix of Low-Code Automation Tools, [Interview]	The criteria measures an organization's mix of Low-Code Automation providers to gain more out this initiative.
Deployment Model [Interview]	The criteria measures the quality of Process Automation Deployment strategies. Process Automation excellence initiative can start as a one-off solution and modify over time into larger strategic business solutions.
Process Excellence included in Hiring/ Training Plans [6, 27, 28] [Interview]	The criteria measures whether process excellence is influencing or being reflected in hiring and training programs in the company. Evolving into a process excellence culture means that process excellence is not the job of just a few, but of everyone. Maturing to a scalable people-centric culture requires regular training of existing personnel with access to collaborative process mapping tools.
Governance Model [7, 28, 29] [Interview]	The criteria measures whether an organization is empowered to solve its own problems related to Process Automation flow, and, therefore, scales it efficiently. Empowering the organization's workforce with automation technology can also present challenges. The most mature organization have a well-defined governance model, which ensures a secure and enterprise-grade path.
Process Excellence/ Automation Executive Sponsorship [10, 27, 28] [Interview]	The criteria surveys the amount of executive sponsorship within an organization. The most successful initiatives always have executive sponsorship to clearly signal their importance. Without strong executive sponsorship that is reinforced at every turn, many large-scale initiatives will fail.

4.4 Robotic Process Automation Portfolio

The Most common reason for failed automation initiatives is wrong process selection, and then the most common risk associated with it, is over or under-utilization of automation maturity. Our proposed innovation is a 2-Dimensional graph (see Fig. 3) with two variables, the Y-Axis which represents business process (BP) complexity, and the X-Axis which represents automation / RPA maturity.

The BP complexity (Y-Axis) is a dependent variable, which is divided into three levels: **Maximum, Average, and Minimum.** The Criteria to define BP complexity have already been listed in Tab. 2. While, The Automation /RPA maturity (X-Axis) is an independent variable, which is divided into four stages: **Stage 1, Stage 2, Stage 3, and Stage 4.** The Criteria to define automation / RPA maturity are listed in Tab. 3. The criteria on the variables X and Y are assessed, weighted, and averaged to the value from 0 to 1.

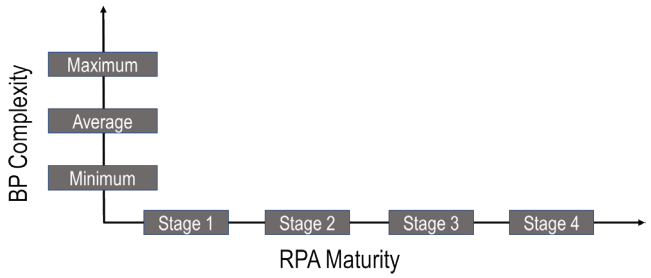


Figure 3 Robotic Process Automation Portfolio

The area between the X-axis and Y-Axis is divided into four quadrants: Long-Term Improvement, No Brainer, Must Do Improvement, and Quick Wins (see Fig. 4). The values or the proportion of the area covered by each of these quadrants is 25% each and equally. The maximum value that "No-Brainer" quadrant will take is $X = 0.5$ and $Y = 0.5$, so, any business processes whose X and Y values are under these maximum values will be treated as "No-Brainer" projects. "Long-Term Improvement" quadrant will take $X = 0.5$ and $Y = 1$, so, any business processes whose X and Y values are under these maximum values will be treated as "Long-Term Improvement" projects. "Must-Do Improvement" quadrant will take $X = 1$ and $Y = 1$, so, any business processes whose X and Y values are under these maximum values will be treated as "Must-Do Improvement" projects. Finally, "Quick-Wins" quadrant will take $X = 1$ and $Y = 0.5$, so, any business processes whose X and Y value are under these maximum values will be treated as "Quick-Wins" projects.

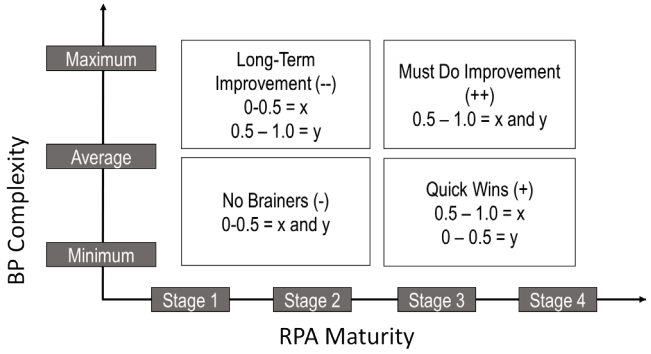


Figure 4 Robotic Process Automation Portfolio with Benefits

4.5 Robotic Process Automation Portfolio Benefits

As already described above there are four quadrants inside RPA Portfolio, starting from the bottom left which is "No-Brainers" and moving clockwise in order – "Long-Term Improvement". "Must-Do Improvement" and "Quick-Wins" (see Figure 4). From the results of the literature study and the interviews, we have created a draft description of the four quadrants for better orientation and understanding. This draft description needs to be verified and improved in future research.

No-Brainers (-): RPA Projects that are easily implemented without any complications and have low outcomes in terms of investments are referred to as "No-Brainers". Their benefits are expected to be less as compared with the "Quick-Wins", as the RPA Maturity of a company is not so advanced yet. Even though the complexity of a process is not much complicated in terms of, the flow development or its working but still the automated process will be a premature baby with no one to take care of or be responsible for its working and monitoring.

There can be frequent stoppages of the bot and every time the company has to be dependent upon external partners to fix it, as there is no expert yet in the company who can easily monitor the process execution and in times of breakages, they can fix it. Therefore, in terms of economics related to the implementation of that RPA Project, the maintenance cost would be rising significantly with no or little chance to jump over the total cost of RPA implementation and hence generate or give less ROI or no ROI at all.

Whenever a process has these kinds of characteristics, it is essential for the organizations to set their expectations and priorities right. Oftentimes, an RPA project becomes a "No-Brainers" project instead of a "Quick-Wins" project when the outcome is not much scalable. However, "No-Brainers" projects have specific advantages, like serving as a self-learning tool before more complicated and impactful processes are automated. Since "No-Brainers" are less impactful as compared to "Quick Wins", therefore, organizations with the low level of automation maturity can first move up the learning curve before moving on to the "Quick-Wins" projects. Just to get a glimpse of what RPA is what it can do, a company with little maturity in RPA can go on automating low-complexity processes and later on, when the automaton maturity is improved, the same process, which was under "No-Brainer" category, can easily be moved to "Quick-Wins" category.

Long-Term Improvement (--): Automation projects with high complexity and benefits similar to or even lesser than those of "No-Brainers" projects are referred to as “Long Term Improvement” projects. These projects bear even lesser benefits as compared to the "No-Brainers", so, companies should be vigilant enough to see that if their Business Process Complexity comes out to be high and on the other hand, the company is not much advanced in the RPA Maturity, then these types of projects are to be completely neglected or be left out.

The higher complexity of the process means the greater cost for development, server, consultancy, and so on. It covers around 50% or even more of the total RPA implementation cost. In addition, the benefits of projects are even lesser than the "No-Brainers" projects, so, in such cases or projects, the companies should not come under the perception of trying out the technology just for experiment or fun or if they did then very badly, they would be failing on RPA implementation with huge monitory losses. So, avoid automating such a category of projects. So, why "Long Term Improvement" project name given to such a quadrant or benefit is because without thinking anything long-term plans should be made to improve the company’s current level of automation maturity if it wants to automate such complicated processes and side by side making that process less complex and standardizing it.

Must Do Improvement (++): Must Do Improvement project bears high complexity of the business process in contrast to very high RPA Maturity. Such projects possess greater benefits than "Quick-Wins" projects as the more complex the process is, the higher the outcome in terms of quality, ROI, and efficiency of the process with no human error at all, while executing a process. Automating such processes which are highly complicated and possess overall greater benefits, should definitely be automated.

The company has a high level of RPA maturity where there's a capability of automatically arranging the unstructured data into a structured one, the frequency of the process is pretty high, and the process is pretty much impactful to the customers (whether internal or external). Of all the quadrants, "Must Do Improvement" projects possibly possess the greatest challenge to any RPA journey since project investments are high while their business impact is also high. So, why is "Must Do Improvement" the name given to such quadrant or benefit even if the outcome of the process is huge, the company or RPA in that company has such capabilities of improving the process, improving its bottlenecks and making it much cleaner and more straightforward with little time to invest on it and generating tremendous benefits.

Quick-Wins (+): RPA projects which are implemented fast and swiftly generate returns are referred to as Quick-Wins projects. They bear low complexity yet still have high reward potential. Low-complexity automation projects can be implemented easily and do not need much customization afterward. The development cost of the RPA project is very much low compared to other costs associated with RPA. The maintenance cost is also significantly very low as the company can solve the errors by itself during the execution of the program.

There is a proper centralized team (CoE) which is responsible for the RPA projects in that company. The "know-how" can be easily transferred from one department to another without any restrictions or anything. A proper Governance model is there with the capability of mapping out the processes that can be considered for automation. Some medium-complexity projects can still be defined as Quick-Wins where data transfers between applications are required. There is a positive ROI and no chance of getting an RPA project failed.

Success is being seen across all the departments of a company and automation is just like another daily routine thing for all the employees. RPA bots work non-stop through the pre-defined, consistent working method. There is scalable work force and an increased focus on value creation. RPA bears low risks by working on online servers without any help from a human. The processing speed is quite high.

5 CONCLUSION AND FUTURE RESEARCH

The design of the Robotic Process Automation Portfolio fulfills the objective of providing a relevant theoretical and practical understanding of RPA projects for company so that they can successfully select business processes (BP) that match their automation maturity, thus avoiding RPA early retirement. Literature reviews and expert interviews were used to identify criteria for the two portfolio variables (*X*) BP complexity and (*Y*) RPA automation/maturity. As a result, there are 10 criteria for business process complexity and 14 criteria for automation / RPA maturity.

The two-variable approach does not guarantee a comprehensive assessment, it is necessary to have other variables such as costs, benefits, usability, technology readiness, and company readiness, as suggested by Axmann and Harmoko (2020) [10]. However, to speed up decision-making as well as the suitability assessment of RPA to the organization, the initiation of two variables is sufficient.
Subsequent research is aimed at creating a broader portfolio while answering other emerging questions such as "How do organizations use automated portfolios?", and "How do organizations calculate the value of *X* and *Y* variables?".

6 REFERENCE

[1] Tripathi, A. M. (2018). *Learning Robotic Process Automation: Create Software robots and automate business processes with the leading RPA tool–UiPath*. Packt Publishing Ltd.

[2] Krakau, J., Feldmann, C., & Kaupe, V. (2021). Robotic process automation in logistics: Implementation model and factors of success. In Adapting to the Future: Maritime and City Logistics in the Context of Digitalization and Sustainability. *Proceedings of the Hamburg International Conference of Logistics (HICL), 32*, 219-256.

[3] Taulli, T. (2020). *The Robotic Process Automation Handbook*. Apress. https://doi.org/10.1007/978-1-4842-5729-6

[4] Pathak, A. (2019). RPA for Small Business. https://www.linkedin.com/pulse/rpa-small-business-avanish-pathak/ (Accessed Aug. 5, 2022).

[5] Axmann, B. & Harmoko, H. (2020). Robotic Process Automation: An Overview and Comparison to Other Technology in Industry 4.0. *The 10th International Conference on Advanced Computer Information Technologies (ACIT)*, Deggendorf, Germany, 559-562. https://doi.org/10.1109/ACIT49673.2020.9208907

[6] Harmoko, H. (2021). The five dimensions of digital technology assessment with the focus on robotic process automation (RPA). *Tehnički glasnik, 15*(2), 267-274. https://doi.org/10.31803/tg-20210429105337

[7] Noppen, P., Beerepoot, I., van de Weerd, I., Jonker, M., & Reijers, H. A. (2020). How to keep RPA maintainable? *The 18th International Conference, BPM2020*, Seville, Spain, September 13–18, *Proceedings, 18*, 453-470.

[8] Axmann, B., Harmoko, H., Herm, L.-V., & Janiesch, C. (2021). A framework of cost drivers for robotic process automation projects. *Business Process Management: Blockchain and Robotic Process Automation Forum: BPM2021 Blockchain and RPA Forum*, Rome, Italy, September 6–10, *Proceedings, 19*, 7-22. https://doi.org/10.1007/978-3-030-85867-4_2

[9] Schuler, J. & Gehring, F. (2018). Implementing Robust and Low-Maintenance Robotic Process Automation (RPA) Solutions in Large Organisations. Available at SSRN 3298036. https://doi.org/10.2139/ssrn.3298036

[10] Harmoko, H., Ramírez, A. J., Enríquez, J. G., & Axmann, B. (2022). Identifying the Socio-Human Inputs and Implications in Robotic Process Automation (RPA): A Systematic Mapping Study. In *Business Process Management: Blockchain, Robotic Process Automation, and Central and Eastern Europe Forum*, Cham, A. Marrella et al., Eds., 185-199.

https://doi.org/10.1007/978-3-031-16168-1_12
[11] Modliński, A., Kedziora, D., Jiménez Ramírez, A., & del-Río-Ortega, A. (2022). Rolling Back to Manual Work: An Exploratory Research on Robotic Process Re-Manualization. *Business Process Management: Blockchain, Robotic Process Automation, and Central and Eastern Europe Forum*, Cham, A. Marrella et al., Eds., 154-169. https://doi.org/10.1007/978-3-031-16168-1_10
[12] Agostinelli, S., Marrella, A., & Mecella, M. (2020). Towards intelligent robotic process automation for BPMers. arXiv preprint arXiv:2001.00804.
[13] Axmann, B. & Harmoko, H. (2022). Process & Software Selection for Robotic Process Automation (RPA). *Tehnički glasnik, 16*(3), 412-419. https://doi.org/10.31803/tg-20220417182552
[14] Herm, L.-V., Janiesch, C., Reijers, H. A., & Seubert, F. (2021). From symbolic RPA to intelligent RPA: challenges for developing and operating intelligent software robots. *Business Process Management: 19th International Conference, BPM 2021*, Rome, Italy, September 6–10, *Proceedings 19*, 289-305. https://doi.org/10.1007/978-3-030-85469-0_19
[15] Gunderson, C., Mahto, R., Abel, T., Harvle, J., & Wyatt, J. (2019). Taking RPA to the Next Level. https://econsultsolutions.com/wp-content/uploads/2019/05/2019-global-rpa-survey-protiviti.pdf (accessed Aug. 9, 2022).
[16] Kumar, R. R. (2020). Measure Your Enterprise Automation with Mindtree's Automation Maturity Model. https://www.mindtree.com/insights/blog/measure-yourenterprise-automation-mindtrees-automation-maturity-model (accessed Aug. 11, 2022).
[17] Fung, H. P. (2014). Criteria, use cases and effects of information technology process automation (ITPA). *Advances in Robotics & Automation, 3*.
[18] Doğuç, Ö. (2021). Robotic process automation (RPA) applications in COVID-19. *Management Strategies to Survive in a Competitive Environment: How to Improve Company Performance*, Springer, 233-247. https://doi.org/10.1007/978-3-030-72288-3_16
[19] Siderska, J. (2020). Robotic Process Automation—a driver of digital transformation? *Engineering Management in Production and Services, 12*(2), 21-31. https://doi.org/10.2478/emj-2020-0009
[20] Capgemini Consulting. (2016). Robotic Process Automation - Robots conquer business processes in back offices: A 2016 study conducted by Capgemini Consulting and Capgemini Business Services. (Accessed Feb. 3, 2022).
[21] Huelin, R., Iheanacho, I., Payne, K., & Sandman, K. (2015). What's in a name? Systematic and non-systematic literature reviews, and why the distinction matters. *The evidence*, 34-37,
[22] vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., & Cleven, A. (2015). Standing on the shoulders of giants: Challenges and recommendations of literature search in information systems research. *Communications of the association for information systems, 37*(1), p. 9. https://doi.org/10.17705/1CAIS.03709
[23] Hofmann, P., Samp, C., & Urbach, N. (2020). Robotic process automation. *Electronic Markets, 30*(1), 99-106. https://doi.org/10.1007/s12525-019-00365-8
[24] Wanner, J., Hofmann, A., Fischer, M., Imgrund, F., Janiesch, C., & Geyer-Klingeberg, J. (2019). Process selection in RPA projects-towards a quantifiable method of decision making. *International Conference on Interaction Sciences*.
[25] Michel, C. & Jameson. J. (2021). Discovering and Selecting the RIGHT Processes for Your Robotic Process Automation (RPA). https://tinyurl.com/2p9yrxc2 (accessed Sep. 2, 2022)
[26] François, P. A., Borghoff, V., Plattfaut, R., & Janiesch, C. (2022). Why Companies Use RPA: A Critical Reflection of Goals. *Business Process Management*, Cham, C. Di Ciccio, R. Dijkman, A. Del Río Ortega, and S. Rinderle-Ma, Eds., 399-417. https://doi.org/10.1007/978-3-031-16103-2_26
[27] nintex. (2022). Your guide to process automation maturity. https://www.nintex.com/resources/process-automation-maturity-model/ (accessed Sep. 4, 2022)
[28] Herm, L., Janiesch, C., Steinbach, T., & Wüllner, D. (2021). Managing RPA implementation projects. *Robotic Process Automation*, 27-46. https://doi.org/10.1515/9783110676693-002
[29] Ruiz, R. C., Ramírez, A. J., Cuaresma, M. J. E., & Enríquez, J. G. (2022). Hybridizing humans and robots: An RPA horizon envisaged from the trenches. *Computers in Industry, 138*, p. 103615. https://doi.org/10.1016/j.compind.2022.103615

Authors' contacts:

Bernhard Axmann, Prof. Dr.-Ing
(Corresponding author)
Faculty of Engineering and Management,
Technischen Hochschule Ingolstadt,
Esplanade 10, D-85049 Ingolstadt, Germany
+49 841 9348 3505, E-Mail: Bernhard.Axmann@thi.de

Harmoko Harmoko, M. Eng.
Faculty of Engineering and Management,
Technischen Hochschule Ingolstadt,
Esplanade 10, D-85049 Ingolstadt, Germany
+49 841 9348 6439, E-Mail: Harmoko.Harmoko@thi.de

Rahul Malhotra, M. Eng
Faculty of Engineering and Management,
Technischen Hochschule Ingolstadt,
Esplanade 10, D-85049 Ingolstadt, Germany
E-Mail: rahulmalhotra96@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230416204744

Original scientific paper
Received: 2023-04-16, Accepted: 2023-04-20

Design for Six Sigma Digital Model for Manufacturing Process Design

Elvis Krulčić*, Sandro Doboviček, Dario Matika, Duško Pavletić

Abstract: The transition to digital manufacturing has become more important as the quantity and quality of the use of computer systems in manufacturing companies has increased. It has become necessary to model, simulate and analyse all machines, tools, and raw materials to optimise the manufacturing process. It is even better to determine the best possible solution at the stage of defining the manufacturing process by using technologies that analyse data from simulations to calculate an optimal design before it is even built. In this paper, Design for Six Sigma (DFSS) principles are applied to analyse different scenarios using digital twin models for simulation to determine the best configuration for the manufacturing system. The simulation results were combined with multi-criteria decision-making (MCDM) methods to define a model with the best possible overall equipment effectiveness (OEE). The OEE parameter reliability was identified as the most influential factor in the final determination of the most effective and economical manufacturing process configuration.

Keywords: digital twin model; DFSS; multi-criteria decision-making methods (MCDM); overall equipment effectiveness (OEE); reliability

1 INTRODUCTION

Today, the application of digitalisation is making its way into all areas of our society. On the one hand, the pandemic COVID19 crisis is accelerating the process of digitalisation in education and business through the implementation of distance activities, the development of new digital models, methods, and skills. On the other hand, a significant energy crisis is accelerating electrification processes at a pace that was unimaginable until yesterday. Long-term strategies and plans need to be adapted to the new conditions, as rapid and significant changes occur every year. The biggest challenge in this process is maintaining competitiveness, speed, and quality. The right answer to this challenge can be the combination of known and proven methods and tools adapted to the new conditions in society. It is no longer sufficient to carry out optimisations during the product life cycle (PLC) [1]. It is imperative to look for solutions or partial solutions in the earlier stages of defining the product and its production process. An area that has developed and improved with unprecedented speed in the last decade is also the digitalisation of production processes. There are several tools at our disposal for creating product simulations and the associated production process. A clear methodology describing rules and norms in the form of standards has not yet been developed and defined, but there are numerous studies in the literature on digitalisation, digital twins (DT) and their application [2]. This paper proposes a model for manufacturing process design that combines digital models with Design for Six Sigma (DFSS) principles in studying the effects of different configurations of production equipment [3]. The digital model is validated by simulation in a Tecnomatix software package with analytical calculations on real machine data.

The paper is structured as follows: Chapter 2 contains an analysis of the current literature in the field of DFSS and DT. The aim of the literature review was to search for available tools for manufacturing process design to improve control over the impact of design on the final performance of the manufacturing process. The idea of combining several of the most useful DFSS tools with digital twins is explained in Chapter 3 with a figure of the proposed hybrid model. Chapter 4 presents the research findings resulting from the application of the proposed model to a case study. In the case study, two manufacturing systems with different configurations were analysed in terms of equipment reliability and productivity rate. The last chapter contains an overview, discussion, and conclusions.

2 THEORETICAL BACKGROUNDS

This section is a literature review of the concepts of DFSS, Digital Factory, Digital Twins and OEE from a manufacturing process design perspective.

2.1 Design for Six Sigma in the Life Cycle Environment

The digitalisation process to improve productivity and the economy is unstoppable and irreplaceable. It is a question of strategic level to what extent and when the digitalisation of certain areas and levels in our company or process will take place. One of the best ways is certainly to involve DT as early as possible, i.e., in the design of the product or production system. The desired end state is a digital factory in which all elements are integrated. According to the standard VDI 4499, Part1: 2008-02 of the Association of German Engineers, the digital factory is a generic term for a comprehensive network of digital models, methods, and tools - including simulation and 3D visualisation - that are integrated into an end-to-end data management system [4]. The goal of the digital factory is the holistic planning, evaluation and continuous improvement of all essential structures, processes, and resources of a real factory in connection with the product. In the development of the individual parts of the digital factory, the various digital tools and concepts emerge, the most important of which are: Internet of Things (IoT), Enterprise Resource Planning (ERP), Project Management (PM), Product Lifecycle Management (PLM), Advanced Planning & Scheduling (APS), Manufacturing Execution System (MES). The

Industrial Internet of Things (IIoT) brings together machines, advanced analytics, and people. The networked assets and devices that use communication technologies are creating systems that can monitor, collect, share, analyse and provide valuable new insights like never before [5]. These insights can then contribute to smarter and faster business decisions for manufacturers. According to a Cisco study from 2017, 74% of companies that start an IoT initiative fail because many of these projects were not successful for various reasons. In most cases, projects went over budget, implementation times were too long, there were problems with interoperability between existing platforms or planning, and resources were not allocated appropriately, leading to project abandonment [6]. An IoT platform is a multi-level technology consisting of software and hardware that includes an operating environment, storage, computing resources, security and development tools that support the management of smaller applications and IoT devices. Generic IoT platforms, like AWS, Azure IoT, SaS, Thing-Worx provides technologies and tools that support the management of IoT devices that are not used in an industrial environment. In this context, a home automation platform that manages smart household devices such as refrigerators, smart windows, temperature, etc. can be called a generic IoT platform. IIoT platforms, on the other hand, provide support for machines and smart devices used in an industrial environment. The IIoT platform provides customised software designed for industrial applications and analytics.

Since digitalisation requires significant resources, a certain level of knowledge and achieved the minimum requirement of Industry 4.0 concept, the question arises, especially for medium-sized and small companies, whether there is a simpler, faster, and cheaper solution with which such companies can maintain their competitiveness in the market. The answer lies in the combination of DFSS principles and DT in decision-making. Design for Six Sigma is a well-known element of product development in the Six Sigma quality programme [7]. The goal of DFSS is to prevent defects by integrating quality into the product, process, and system. Unlike other Six Sigma methods, DFSS is product-oriented and not process-oriented. Nevertheless, DFSS should be integrated into a product and manufacturing process development framework, so that it becomes repeatable and optimised to achieve sustainable success for any organisation. DFSS provides a range of techniques that engineers can use to improve their effectiveness in developing systems that reliably meet customer requirements. Talyor et al. [8] have demonstrated the integration of classical techniques to create a highly available system and provided an overview of some of the most useful techniques. Several of these techniques were used in presented case study. There are many available digital tools for product development, but DFSS is not one of them, it represents a cultural change within the various functions and organisations in which it is implemented [9]. It provides strong statistical tools to address weak or new processes and increase customer and employee satisfaction. To successfully implement these continuous methods, the objectives of DFSS and Six Sigma should be linked to the company's goals, vision, and mission. They are powerful tools that support the achievement of leadership in design, customer satisfaction and cultural change.

2.2 Digital Twin in the Life Cycle Environment

Grieves et.al [10], who first introduced the concept DT in 2003, defined DT as "a set of virtual information constructs that fully describe a potential or actual physical product from the microatomic to the macrogeometric level" from which all the required information "could be obtained by inspecting a physical product". DT has developed rapidly over the last few decades, and in recent years have seen an enormous scientific contribution to the exponential growth of scientific papers and patents filed on the subject DT. There are many analyses and attempts to summarise the different views and applications of DT [11]. Of particular interest are the reviews that attempt to define the gradation of seven main areas: Goals, User Focus, Lifecycle Focus, System Focus, Data Sources, Data Integration Level and Authenticity, as in Tab. 1 [12]. In other literature reviews, authors go into detail about the terms digital model, digital shadow and digital twin and the difference between them [13]. The definitions differ in the degree of data integration between the physical and digital parts. Some digital representations are independent models that are not connected to a physical object in real time, while others are fully integrated with real-time data exchange. Most publications are classified as Digital Shadow and Digital Model, only 18% of the papers use the term Digital Twin with bidirectional data transfer [13]. When modelling real-world scenarios in virtual and mathematical environments, it is obvious that the quality of the results depends on the quality of the model. Every production line and every manufacturing plant is different and there are classifications according to parameters and characteristics. Each of these classifications is approached in a particular way when modelling, as each has limitations, and a model is specific to the object of study and cannot be generalised [14].

Table 1 Digital Twin application dimensions [11]

Dimensions	Values			
Goals	Information acquisition	Information analysis	Decision and action selection	Action implementation
User focus	Single		Multiple	
Life cycle focus	One phase		Multiple phases	
System focus	Component	Subsystem	System	System of system
Data sources	Measurements	Virtual data		Knowledge
Data integration level	Manual	Virtual data		Knowledge
Authenticity	Low ⟷ High			

VDI Guideline 3633 defines simulation as the reproduction of a system, including its dynamic processes, in a model with which experiments can be carried out. The aim of simulation is to obtain a result that can be transferred to a system in a real environment. Furthermore, simulation defines the preparation, execution, and evaluation of

carefully controlled experiments in a simulation model [15]. Following the same guideline, the basic steps of any process simulation are as follows:

1) Definition of the problem.
2) Examination of the reasonableness: cost-effectiveness of creating a simulation.
3) Definition of the simulation objectives.
4) Verification of the process in a real environment and collection of data needed to create a simulation model.
5) Creation of a simulation model: a twin of the real model in accordance with the desired objectives of the simulation analysis.
6) Verification of the simulation model by comparing the individual results of the model with the actual results of the physical system in production.
7) Conducting experiments, i.e., performing simulations within the simulation model.
8) Analysing the results and interpreting the data - simulation results.
9) Documenting the results, i.e., management decision making based on the simulation results about process improvements in the real environment.

PwC [16] in their quantitative research conducted by 200 industry company in Germany, defines three different types of digital twin for industrial applications in "Digital Factories 2020": the digital twin of the product, the digital twin of the production facility and the digital twin of the factory. This work deals with the digital twin of the factory and is based on a discrete event simulation model representing the machining department in the automotive industry and aims to quantitatively evaluate the impact of different production configurations on lead time, production capacity and number of workers. A digital twin of one or more production lines is used for design, virtual commissioning, and ongoing operation. The focus is on simulating the operation of a plant to adjust and optimise its key parameters and enable concepts such as predictive maintenance or augmented reality. Simulation modelling has become an indispensable tool to analyse expected performance, validate designs, demonstrate, and visualise processes, test hypotheses and perform many other analyses. It is the preferred tool in a wide range of industries, and as mentioned earlier, in some industries it is even required before any major investment. However, there are some cases where simulation is not the best technique to find a suitable solution. Laguna et al. [17] have established 10 rules for when it is not appropriate to simulate.

By applying the basic principles of DFSS with DT in the form of using true-to-life digital models in simulations when designing manufacturing processes, it is possible to investigate most possible scenarios and make the best possible decision when determining the final configuration of the production process quickly and cost-effectively.

3 PROPOSED/NOVEL APROACH IN MANUFACTURING PROCESS DESIGN

Despite numerous articles about tools and models for the product development process, the literature review has not revealed a comprehensive model that proposes the use of methods and tools in the sense of DFSS in the field of manufacturing process design. This chapter presents the author's efforts to propose a hybrid model whose main application is the design of manufacturing systems.

3.1 DFSS and Classical Reliability Techniques

A good explanation of the DFSS tools and methods used to analyse and design highly available systems, with simple examples, is given by the Institute of Electrical and Electronics Engineers [8]. These techniques are listed in Tab. 2.

Table 2 DFSS tools and methods

Abbreviation	Tools / Methods
VOC	Voice of the Customer
KANO	KANO analysis
	Analysis of technical risk
QFD	Quality Function Deployment or House of Quality
CPM	Critical Parameter Management
	First principles modelling
DoE	Design of Experiments
DFMEA	Design Failure Modes and Effects Analysis
FTA	Fault Tree Analysis
	Pugh matrix
	Monte Carlo simulation
	Commercial DFSS tools
	Mathematical prediction of system capability
	Visualizing System Behaviour early in the life cycle
	Critical Parameter Scorecard

3.2 DT terms

As mentioned above, DT has evolved considerably over the past decades without a definitive definition of a standard. Due to the constraints for an adequate implementation of DT (Industry 4.0), it is important to understand the differences in the degree of data integration between physical and digital objects. These crucial differences make it possible to implement DT even without an achieved Industry 4.0 level in the company, at an early stage of the development of the manufacturing process. Uhlenkamp et al. proposes a classification of Digital Twins into three subcategories based on the degree of data integration [12]. The subcategories are shown in Fig. 1 and Fig. 2.

3.2.1 Digital Model (DM)

A DM is a digital representation of an existing or planned physical object that does not use any form of automated data exchange between the physical object and the digital object as shown on Fig. 1, left. Such models include, but are not limited to, simulation models of planned factories, mathematical models of new products, or other models of a physical object that do not use any form of automated data integration. Digital data of existing physical systems can be used to develop such models, but all data exchange is manual.

3.2.2 Digital Shadow (DS)

Based on the definition of DM, if there is also an automated one-way data flow between the existing physical

object and a digital object, one can speak of such a combination as DS, as shown on Fig. 1, right.

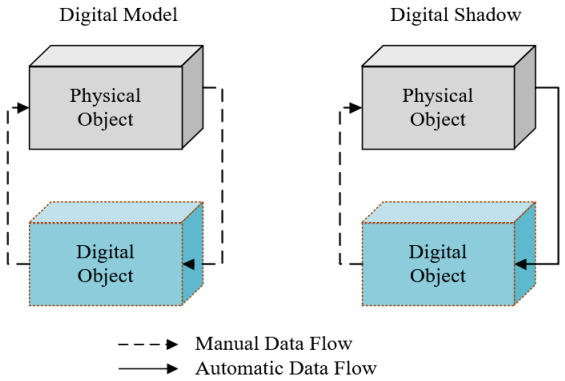


Figure 1 Data flow in a digital model and digital shadow

3.2.3 Digital Twin (DT)

The data flows between an existing physical object and a digital object are fully integrated in both directions, you could call it DT. In this case, the digital object can also act as a control instance of the physical object.

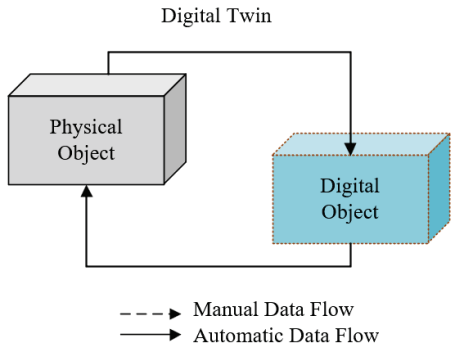


Figure 2 Data flow in a digital twin

3.3 DFSS & DM Novel Hybrid Approach

When developing a new manufacturing process or modifying part of an existing physical object, the authors propose combining some DFSS tools with digital models and MCDM tools. With digital models, it is possible to simulate the model of an existing physical system and analyse or compare alternative solutions for a new complete manufacturing process or part of the system. In manufacturing, it is often the case that several stages of prototypes are produced and tested in the introduction phase of production to improve the product design. Due to limited time and investment, it is not possible to apply the same principle to production facilities. Production optimisation is an ongoing process, both in the start-up phase and during the product life cycle. In some cases, experts have only realised the effects of a non-optimal design of the production process in the late phase of production preparation or later during production operation. The goal of the proposed hybrid Design for Six Sigma Digital Model (DFSSDM) approach is to avoid or significantly reduce these types of costs by using DM together with DFSS techniques. With the proper use of the tools, methods and DM shown in Fig. 3, it is possible to achieve these goals. The structure of the proposed hybrid model consists of using most of the original tools of DFSS and Digital Model in the presented grouped techniques. Certain tools have been replaced by more convenient tools. Design Failure Modes and Effects Analysis (DFMEA) has been replaced by Process Design Failure Modes and Effects Analysis (PFMEA), as the object of analysis is a production system and not a product. Similarly, Voice of Customer (VOC) has been replaced by Voice of Supplier (VOS) as the focus is on the supplier's product, the production system, and not the customer's product. The use of commercial DFSS tools such as Minitab, Matlab and others is complemented by digital twin models. The proposed model recommends the use of Plant Simulation or a similar software package for simulation, as this software offers much richer capabilities for representing the physical system in production and logistics.

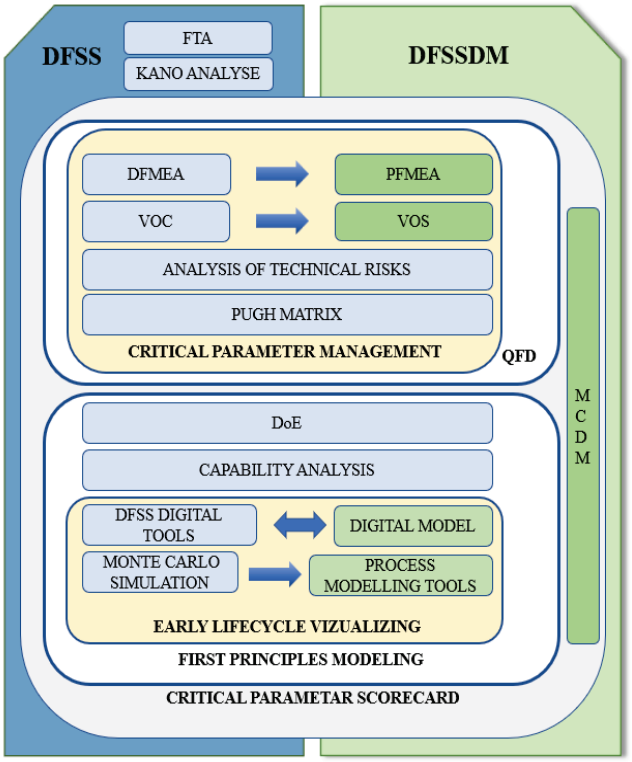


Figure 3 Transition model from DFSS to DFSSDM hybrid model

To properly define the Key Performance Indicators (KPIs) needed to determine the optimal solution for a design or analysis, it is necessary to use one of the many MCDM tools [18]. A model for selecting the most important KPIs and methods for selecting new assets in manufacturing is presented in a paper by the author [19], which can be used in combination with individual DFSSDM tools. Some tools such as the Pugh matrix [20], Design of Experiments (DoE) [8] are combined with other MCDM tools in their original form. Depending on the specificity of the object of analysis, the use of the defined hybrid DFSSDM needs to be adapted,

the model can be further extended with individual tools, or the use of the existing tools can be restricted.

3.4 Manufacturing KPI's

The goal of any manufacturing is to achieve all defined KPIs to realise the strategy. The definition of certain KPIs depends on many factors. Without going into the analysis of the KPIs for the design of the manufacturing process, they all have the function of designing an optimal system to ensure the achievement of the planned KPIs in manufacturing. In manufacturing, one of the most important KPIs that shows a real picture of a system's performance is Overall Equipment Effectiveness (OEE) [21]. The OEE value, which is the result of the calculation model that considers process availability, performance and quality, is a good indicator of process performance. The relationship is represented by the following Eq. (1).

$$OEE = Availability \times Performance \times Quality \tag{1}$$

Although all elements of OEE are equally important, practise shows that some companies have more problems with achieving the required quality level, others with the expected performance or plant availability. To improve its performance, each company must identify its weak points. At the stage of defining the manufacturing process, it is very important to know the expected OEE level as well as other KPIs. Since OEE and other KPIs are in a life cycle that is constantly being optimised, it is important to define the best resources in the right configuration to meet the company strategy. In many companies, availability is the most influential element. It depends on many factors, such as the quality of the equipment, the level of knowledge of the user, the quality of maintenance of the equipment, the level of knowledge of the maintenance staff. No less important are the type of maintenance and the organisation at company level. Determining the critical component of the production system is crucial to predicting the occurrence of a bottleneck. Bottleneck characteristics measure the success of the entire production system. This does not necessarily mean that the single most critical and worst element of the system is also the overall outcome of a system. In a production system, it is very important to define the configuration model of the system. However, even if assume that all the above factors are at a very high level, the configuration of the equipment in the production system can significantly affect the overall reliability of the production system. In production system configuration, recognize parallel, serial, and combined configuration models [21, 22]. According to the American Society of Quality [23], the reliability of a production system is defined as the probability that the product will perform its intended function under defined conditions without failure or error during a specified period. The most reliable factories tend to be more productive, have more stable processes that result in high quality and low costs, and have more skilled employees. There is a strong correlation between process KPIs and factory reliability. The reliability of a system can be cascaded to the reliability of its components. Mathematically the reliability of a serial system is expressed by Eq. (2), and for a parallel system by Eq. (3) [24].

$$R(t) = \prod_{i=1}^{n} \left[R_i(t) \right] \tag{2}$$

$$R(t) = 1 - \prod_{i=1}^{n} \left[1 - R_i(t) \right] \tag{3}$$

For simple systems, calculating reliability is straightforward, but often industrial systems involve complex, combined models. In this case, several key parameters need to be considered for a final decision, leading to a rather complex production system analysis, where digital models can be of great help.

4 DIGITAL MODEL VS CONVENTIONAL ANALYSIS – CASE STUDY

4.1 Creation and Validation of a Simulation Model

Prior to the creation of the digital model, the results of the PFMEA, the VOS, the technical risk analysis, and the capability analyses, which are part of the standard procedure of the company concerned regarding the input parameters for the simulation model, were reviewed. Following all the steps presented in chapter 2.2, the PROS1 and PROS2 models were created in the simulation software Siemens Plant Simulation for two production systems with different configurations. Tecnomatix Plant Simulation is a discrete event simulation tool that can be used to create digital models of logistics systems, e.g., in manufacturing, allowing the study of system characteristics and the optimisation of their efficiency. The Tecnomatix Plant Simulation software package can be used to model and simulate production systems and their processes. Material flow, resource utilisation and logistics can be optimised for all planning levels, from global production plants to local plants and specific production lines [25]. The models include CNC equipment for machining, a washing machine, and a leak testing station in various configurations for machining identical aluminium castings. The main difference between the two production systems is the CNC2 equipment (Tab. 3). The PROS1 configuration has four identical units in parallel configuration; PROS2 has one CNC2 unit with 3 serial modules and a robot cell for deburring after machining. The aim of the simulation is to validate the DFSSDM model and define a production system with better OEE performance. The production system PROS1 represents the existing physical equipment, PROS2 is an alternative solution considered to increase capacity. The analytical calculation of the finance department defined PRO2 as the more economical solution to produce a specific product. The purpose of applying the hybrid DFSSDM model is to confirm the hypothesis that the PROS2 system is more productive and economical than the existing PROS1 system. For practical validation of the novel DFSSDM model, the simulation is performed with a minimum number of process inputs. Tab. 3

shows the input data into the PROS1 and PROS2 production model after validation of the model, which was carried out with the same inputs for the same type of plant, and the comparison with the analytical results. The difference between the simulation model and the physical model for PRO1 was less than < 3,7%, so the model passed validation. The main process inputs in the digital model with the biggest difference defined with VOS, CPM and PFMEA are Reliability and Number of employees.

Table 3 Input data in simulation models

		Key process input			
	Assets	*Tc* (min)	*WF*	*R*	*MTTR* (min)
PROS1	CNC1	2,12	0,5	98%	12
	CNC1	2,12	0,5	98%	15
	CNC1	2,02	0,5	85%	11
	CNC1	2,02	0,5	85%	10
	CNC2	4,16	1	90%	5
	CNC2	4,16	1	90%	12
	CNC2	4,16	1	90%	10
	CNC2	4,16	1	90%	10
	WM	0,48	1	90%	20
	LT	0,48	1	98%	14
PROS2	CNC1	2,09	0,5	85%	15
	CNC1	2,09	0,5	85%	17
	CNC2	1,00	0,33	98%	45
	CNC2	1,06	0,33	98%	30
	CNC2	1,05	0,33	99%	33
	RC	0,56	0	90%	5
	WM	0,57	1	90%	20
	LT	0,56	1	98%	10

Work calendar: 24/7, 7,5h/shift, Internal logistic in workforce time
CNC1,2 – CNC machine for different machining operations
WM – washing machine, LT – Leak Test and package station.
RC – robotic cell for deburring
Tc – cycle time, *WF* – number of workers
R – Reliability, *MTTR* – Mean Time to Repair

The data for PROS1 are representative data collected during one production year for all unit types. The data for PROS2 is a combination of collected data for the same type of equipment as in PROS1 and planned data for new equipment provided by the equipment supplier. The system reliability results were calculated using the indirect method based on the results of the processed pieces. For additional validation of the model, the data of the quality level and productivity of the workers were blocked for this analysis, they were set to fixed optimal constant values so that they would not influence the result. The reliability data for the same type of equipment are average values for all equipment of the same type, as only this type of data was available for new equipment, while the mean time to repair (MTTR) values for PROS1 and for PROS2 are real for a similar application. The digital simulation models for production systems are shown in Fig. 4 and Fig. 5.

4.2 Research Results

To investigate the influence of time on the reliability behaviour of the system, simulations were carried out for different production periods, as shown in Tab. 4. From the results presented, for the specified production system with defined input parameters, it is necessary to create a simulation for a longer period, at least for 30 working days, in order to obtain repeatable results. The simulation results for the total produced part for a production period of 1 year and the reliability of the production system are shown in Tab. 4. The results of the resource statistics for PROS1 and PROS2 are shown in Fig. 6 and Fig. 7. These data are used to investigate the potential effects of quality level, workers, and balancing time between system elements on productivity.

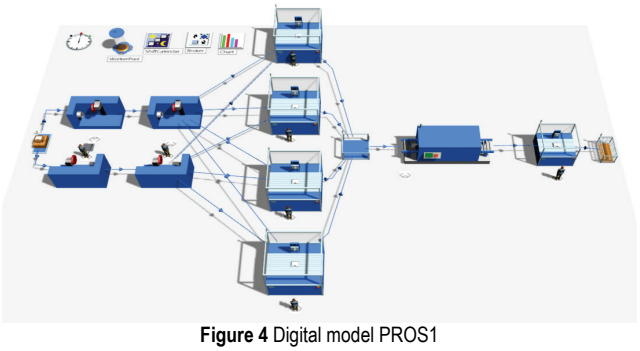


Figure 4 Digital model PROS1

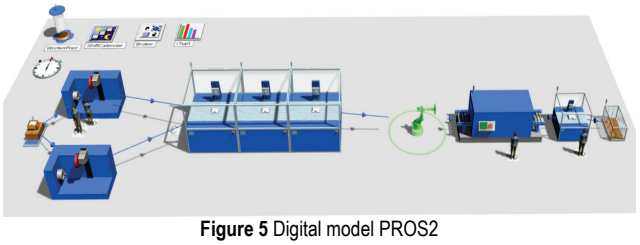


Figure 5 Digital model PROS2

Due to shorter cycle times and significantly higher individual availability in PROS2, the initial performance of PROS2 is better, but the overall result show that the PROS1 system has better reliability and productivity due to the different production configuration.

Table 4 Simulation result for productivity and availability

Production time	8h	1 day	10 days	30 days	330 days
Production (parts) PROS1	325	1045	10533	31605	348331
Production (parts) PROS2	357	1029	9145	27781	307894
Reliability (%) PROS1	93,7	87,8	88,7	88,5	88,7
Reliability (%) PROS2	88,8	84,3	75,0	75,9	76,5

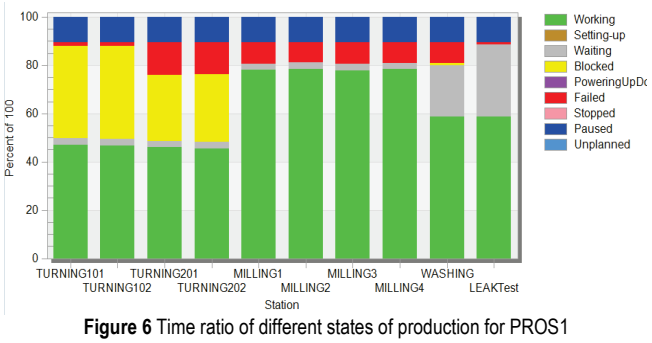


Figure 6 Time ratio of different states of production for PROS1

The PROS1 system can produce 40437 parts or 13% more on an annual basis. Therefore, from the point of view of availability, PROS1 is the better solution than PROS2. For

the final decision, the second important factor must also be considered - the number of production workers.

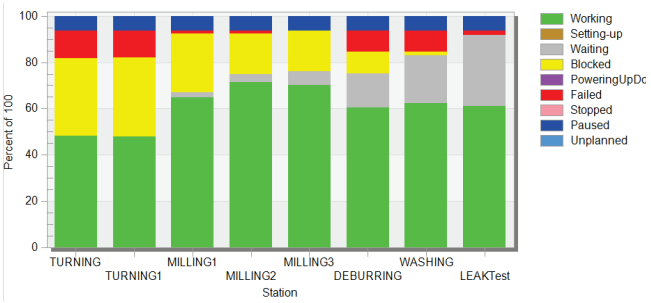


Figure 7 Time ratio of different states of production for PROS2

The initial setting of the number of workers in PROS1 is eight workers required, while PROS2 requires four workers. The results of the simulation models show an average worker utilisation of 94.6% for PROS1 and 72.5% for PROS2. Considering the difference of four workers in terms of number of workers, PROS2 is more favourable. Therefore, for the final decision between the two proposed production systems, additional factors should be considered, which can be done in different ways, e.g., by additionally applying an MCDM method and using the available data and expert groups to make a final decision.

Table 5 Pugh matrix

High Availability & Performance Attributes	Reference	Concept1 PROS1	Concept2 PROS2
Total investments	0	0	0
Maintenance - complexity of activity	0	+	-
Fast repair time	+	0	-
Overload protection	+	0	+
Low downtime	+	+	0
Cost/pcs	-	-	+
Flexibility	+	0	-
Clamping polyvalence	+	0	-
Set-up: complexity and duration	0	-	+
Maintenance - maintainer number	+	0	-
Worker number	-	--	++
Supplier Reference	+	+	-
Additional automatization possibilities	+	+	0
Quality level	0	0	+
OEE	0	0	--
Sum of pluses	8	4	6
Sum of minuses	2	4	7
Total score	6	0	-2

Legend: 0 neutral, + positive, - negative, ++ strong positive, -- strong negative

Another approach may be to define different technical solutions based on the impact of the process configuration on reliability. Alternatively, one can extend the scope of the simulation models to include other data such as quality, energy, logistics, additional analysis of different production configurations, etc. to confirm the analytical calculations of production costs. For the final decision in this study, the MCDM tool Pugh-Matrix was chosen to define the final solution as the period and data needed were limited. Other appropriate MCDM methods such AHP and TOPSIS or PROMETEE can also be used. The results of the Pugh matrix, conducted by a competent team of experts, can be found in Tab. 5. In the final decision based on the Pugh matrix, PROS1 was chosen as the better option for designing the production system for the machining process of the specific product. This means that the original hypothesis that PROS2 is the better solution than PROS1 to produce the identical product can be rejected.

5 DISCUSION AND CONCLUSIONS

This paper presents a model for the design of the manufacturing process that combines digital models with the principles of DFSS and MCDM tools.

Since DT requires significant capital, a certain level of knowledge and the minimum requirement of Industry 4.0 concept, the authors are looking for a solution that combines DFSS principles and DT in the decision-making process. Based on a literature review, good knowledge of DFSS methodology and experience with manufacturing process design, the authors propose a hybrid model for manufacturing process design. The model uses some original DFSS tools, some of them are replaced by more suitable ones, some are slightly adapted, all tools are integrated with digital models and MCDM methods. The use of a digital model is essential for the design of the manufacturing process, as the physical object does not yet exist. The final solution of the digital model is transferred to the digital twin after all networked devices are installed in the production.

The main purpose of the proposed model is to make the design of the manufacturing process professional and relevant by using simulation tools to find an optimal solution that avoids or significantly reduces the cost of starting the production of a new product. The application of the proposed model in a real manufacturing case confirmed the usefulness of using a model in the conditions of real manufacturing processes. The advantage of this model is that companies that are not yet at the level of Industry 4.0 can remain competitiveness on the market and further develop their business processes. With the new technological possibilities, flexibility, agility and lower costs, companies can start their journey to create a digital twin with lower capital investments and shorter time to value than ever before. Another advantage is that the model can be used continuously in optimisation activities in the PLC, so it can be applied to existing manufacturing processes and the development of new manufacturing processes.

In further research, the proposed model can be extended with another existing quality function development (QFD) tool to further define the framework for applying the MCDM method. Simulations with process modelling tools can be integrated with AI, especially in cases where the level of Industry 4.0 has been reached and digital twins are already present in the manufacturing area.

Acknowledgments

This work has been fully supported by the University of Rijeka (contract no. uniri-tehnic-18-33).

6 REFERENCES

[1] Semenkov, K., Promyslov, V., Poletykin, A., & Mengazetdinov, N. (2021). Validation of Complex Control Systems with Heterogeneous Digital Models in Industry 4.0 Framework. *Machines, 9*, 62. https://doi.org/10.3390/machines9030062

[2] Grieves, M. (2014). Digital twin: Manufacturing excellence through virtual factory replication. *White paper*, 1, 1-7.

[3] El-Haik, B. & Roy, D. M. (2005). *Service Design for Six Sigma: A Road Map for Excellence*. Willey-Interscience publication, 33-45. https://doi.org/10.1002/0471744719

[4] VDI 4499 Part 1. (2008-02). Digital Factory – Fundamentals

[5] Lewis, P. T. (1985). Correcting the IoT History. Chetan Sharma, 14 March 2016. Retrieved 1 June 2021.

[6] Wlodarczak, S. (2008). Don't Let Your Industrial Internet of Things Project Fail. *RFID Journal*. https://www.rfidjournal.com/view-from-the-top-5-cios-speak-out-on-rfid?16470%2F2

[7] Pavletić, D., Soković, M., & Krulčić, E. (2006). Six Sigma Process Quality Improvement in Automotive Industry. *Strojarstvo, 48*(3-4), 173-185.

[8] Taylor, Z. & Ranganathan, S. (2014). *Designing High Availability Systems: Design for Six Sigma and Classical Reliability Techniques with Practical Real-Life Examples*. First Edition. The Institute of Electrical and Electronics Engineer. John Wiley & Sons, Inc, 278-302. https://doi.org/10.1002/9781118739853

[9] El-Haik, B. & Shaout, A. (2010). *Software Design for Six Sigma: A Roadmap for Excellence*. John Wiley & Sons, Inc. 21-53. https://doi.org/10.1002/9780470877845

[10] Grieves, M. & Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behaviour in Complex Systems. In: Kahlen, J., Flumerfelt, S., Alves, A. (eds) *Transdisciplinary Perspectives on Complex Systems*. Springer, Cham. 85-113. https://doi.org/10.1007/978-3-319-38756-7_4

[11] Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital Twin in Industry: State-of-the-Art, *IEEE Transactions on Industrial Informatics, 15*(4), 2405-2416. https://doi.org/10.1109/TII.2018.2873186

[12] Uhlenkamp, J. F., Hribernik, K., Wellsandt, S., & Thoben, K. D. (2019). Digital Twin Applications: A first systemization of their dimensions. *IEEE (ICE/ITMC)*, 1-8. https://doi.org/10.1109/ICE.2019.8792579

[13] Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital Twin in Manufacturing: A categorical literature review and classification. *IFAC Papers Online*, 51-11.1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474

[14] Tao, F., Zhang, M., & Cheng, J. F. (2017) Digital Twin Workshop: A New Paradigm for Future Workshop. *Computer Integrated Manufacturing Systems, 23*, 1-9. https://doi.org/10.13196/j.cims.2017.01.01

[15] VDI 3633 (2018). Simulation of systems in materials handling, logistics and production - Terms and definitions, Verein Deutscher Ingenieure, Association of German Engineers.

[16] Geissbauer, R. Schrauf, S., Berttram, P., & Cheraghi, F. (2017). Digital Factories 2020. PricewaterhouseCoopers GmbH Wirtschaftsprufungggesellschaft (PwC). https://www.pwc.com/ca/en/industries/industrial-manufacturing/digital-factories2020.html

[17] Laguna, M. & Marklund, J. (2019). *Business process modelling, simulation, and design*. CRC Press. 281-287. https://doi.org/10.1201/9781315162119

[18] Zeleny, M. (1982). *Multiple criteria decision making*. New York. McGraw-Hill.

[19] Krulčić, E., Pavletić, D., Dobovičeк, S., & Žic, S. (2022). Multi-Criteria Model for the Selection of New Process Equipment in Casting Manufacturing: A Case Study. *Tehnički glasnik, 16*(2), 170-177. https://doi.org/10.31803/tg-20220407112829

[20] Pugh, S. (1981). Concept selection: a method that works. *Proceedings of the International conference on Engineering Design*, 497-506.

[21] Stamatis, D. H. (2017). *The OEE primer: understanding overall equipment effectiveness, reliability, and maintainability*. CRC Press Taylor & Francis, 19-20. https://doi.org/10.1201/EBK1439814062

[22] Ebeling, C. E. (1997). *An Introduction to Reliability and Maintainability Engineering*. McGraw-Hill, 2, 5, 9.

[23] American Society of Quality. (2022). https://asq.org/quality-resources/reliability (accessed 01.09.2022).

[24] Mangey, R. (2019). *Reliability Engineering – Methods and Application*. CRC Press Taylor & Francis Group. https://doi.org/10.1201/9780429488009

[25] Bigslow, S. (2020). *Tecnomatix Plant Simulation - Modelling and Programming by Means of Examples*. 2nd Ed. Springer. https://doi.org/10.1007/978-3-030-41544-0

Authors' contacts:

Elvis Krulčić, MEng. PhD Student
(Corresponding author)
University of Rijeka, Faculty of Engineering
Vukovarska 58, 51000 Rijeka, Croatia
ekrulcic@riteh.hr

Sandro Doboviček, PhD, Associate Professor
University of Rijeka, Faculty of Engineering
Vukovarska 58, 51000 Rijeka, Croatia
sandro.dobovicek@riteh.hr

Dario Matika, PhD, Full Professor
Mechanical Engineering, Zagreb University of Applied Sciences
Vrbik 8 10000 Zagreb, Croatia
dario.matika@tvz.hr

Duško Pavletić, PhD, Full Professor
University of Rijeka, Faculty of Engineering
Vukovarska 58, 51000 Rijeka, Croatia
dusko.pavletic@riteh.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230417145110

Original scientific paper
Received: 2023-04-17, Accepted: 2023-04-21

Artificial Neural Network System for Predicting Cutting Forces in Helical-End Milling of Laser-Deposited Metal Materials

Uroš Župerl*, Miha Kovačič

Abstract: When machining difficult-to-cut metal materials often used to make sheet metal forming tools, excessive cutting force jumps often break the cutting edge. Therefore, this research developed a system of three neural network models to accurately predict the maximal cutting forces on the cutting edge in helical end milling of layered metal material. The model considers the different machinability of individual layers of a multilayer metal material. Comparing the neural force system with a linear regression model and experimental data shows that the system accurately predicts the cutting force when milling layered metal materials for a combination of specific cutting parameters. The predicted values of the cutting forces agree well with the measured values. The maximum error of the predicted cutting forces is 5.85% for all performed comparative tests. The obtained model accuracy is 98.65%.

Keywords: cutting forces; helical end milling; layered metal material; linear regression; neural model

1 INTRODUCTION

Multilayer metal materials are increasingly used in tool shops to produce modern metal-forming tools for the automotive industry. The properties of individual layers of such material are determined according to the given requirements. There are quite a few processes available for producing such materials, the most widespread of which is the LENS (laser-engineered net shaping) process [1]. In this process, the workpiece is made layer by layer. If individual layers are produced with different process parameters, the machinability of these layers is variable. The machinability of such materials represents a significant challenge for the metal processing industry. The mechanical processing of these inhomogeneous materials is highly demanding and leads to frequent damage to the tool's cutting edges and excessive tool wear. Tool breakage and wear can be directly related to cutting forces. Knowledge of the maximum cutting forces on the tool enables the precise selection of cutting parameters, thereby reducing processing errors and increasing the tool's life.

Advanced machine tools enable cutting force control and adjust the parameters to current cutting conditions. Knowing the maximum cutting forces and advanced control of the cutting forces makes it possible to efficiently machine even difficult-to-cut non-homogeneous multi-layered materials.

In the literature, it is possible to find some research on processing multi-directional layered metal materials. The most common is research on the machinability of hard-to-machine nickel-based alloys [2, 3], titanium alloys [4] and laminates [5]. An overview of difficult-to-machine alloys, which are most often used in the aerospace, nuclear and medical industries, was given by researchers in work [6]. LI et al. [7] presented a literature review on the machinability of additively manufactured titanium alloys.

Altiparmak et al. [8] studied the machinability of high-strength aluminium alloys manufactured using additive technologies.

The researchers [9] studied the mechanism of chip formation in cutting laser-deposited materials: They found it to be complex and not the same as in the processing of homogeneous materials, where mostly form chips are formed.

Many models of cutting forces for end-milling can be found in the literature. Most mechanistic models [10] have proven robust, simple and effective. More sophisticated models consider radial cutter runout [11], tool deflection [12], system dynamics, cutter contact area [13], indentation of the cutting edge into the work material [14], dynamic chip thickness [15], chip forming and friction forces [16] and radius of curvature for sculptured surface machining [17]. In all these models, the cutting forces are calculated as the product of the uncut chip thickness and the specific cutting forces [18]. In mechanistic models, the helical end-mill is divided in the axial direction into differentially thin elements. The product of the specific cutting force and the uncut chip cross-section area determines the differential force on each differential element. By integrating the differential cutting forces over the entire height of the tool's cutting edge, which is in contact with the workpiece, the total cutting force on the tool is then determined. A few studies have been published that deal with the modeling of cutting forces in helical-end milling composites [19, 20, 21]. Models have been created that consider a unique method of determining the specific cutting force from orthogonal (cutting data) cutting data, the geometry of the chips, the geometry of the tool and the cutting parameters [22, 23]. The research [24] on creating a model of cutting forces for milling carbon fibre-reinforced polymer (CFRP) composites is essential. This model calculated the specific cutting forces by considering the current chip thickness and cutting speed. The main challenge in producing these models is the labor-intensive acquisition of specific cutting forces for shape cutting. Obtaining specific cutting forces for different combinations of tools and workpieces represents a great challenge, as it requires a lot of experimental and analytical work. This is particularly difficult in the case of multi-layer materials with different

machinability of individual layers. Therefore, in this research, we developed a methodology to automatically determine cutting forces for helical-end mills. The procedure is based on a short-term measure of cutting forces under certain cutting conditions. Immediately after the measurement, the obtained data and the corresponding cutting conditions are autonomously transmitted to the artificial intelligence. An artificial neural network automatically generates a model of the cutting forces in a few seconds. This type of modeling is fully automated and does not require human intervention. Obtaining specific cutting forces is not necessary. Many models of cutting forces based on neural networks have been published [25-27] and have been shown to be very accurate in prediction.

The article is organized as follows. Chapter 2 presents the modeling of cutting forces and helical-end milling of layered laser deposited materials. In chapter 3, three artificial neural network models for predicting the cutter's maximum cutting forces in the helical end-milling of laser-deposited metal materials are presented in detail. In chapter 4, the linear regression model of cutting forces is presented. In the next chapter, the results of the models are analyzed. Chapter 6 describes the experimental procedure of data acquisition. Section 7 gives the concluding remarks.

2 MODELING OF CUTTING FORCES IN HELICAL-END MILLING OF LAYERED LASER DEPOSITED MATERIALS

The mechanistic technique that divides the cutter in the axial axis into thin, small elements is often used to model cutting forces in helical-end milling (Fig. 1).

The differential cutting force on the thin element of the cutting tool is determined based on the cutting tool geometry by multiplying the cross-section of the chip and the specific cutting force for that part of the layered material that is in contact with the thin element of the tool. The total cutting force is determined by adding all differential cutting forces on all thin tool elements in contact with the workpiece. To calculate the differential cutting forces on a thin tool element, verified equations are available in the literature, and the calculation does not cause problems. A significant challenge is the experimental determination of specific cutting forces for each layer of a multilayer laser-deposited material. The specific cutting force for a material layer is determined as the ratio between the experimentally determined current component of the cutting force and the calculated cross-section of the chip. This process is time-consuming, demanding and labour-intensive.

Therefore, in this research, we developed a methodology to quickly determine cutting forces for helical-end mills. The procedure is based on a short-term measure of cutting forces under certain cutting conditions. Immediately after the measurement, the obtained data and the corresponding cutting conditions are autonomously transmitted to the artificial intelligence. An artificial neural network automatically generates a model of the cutting forces in 2-3 seconds. With further measurements, the already developed neural network model is supplemented with new data. This type of modeling is fully automated and does not require human intervention. Several artificial neural network architectures have been tested to achieve rapid model generation. Three of the most effective neural network architectures are presented, best adapted to modeling cutting forces in helical-end milling.

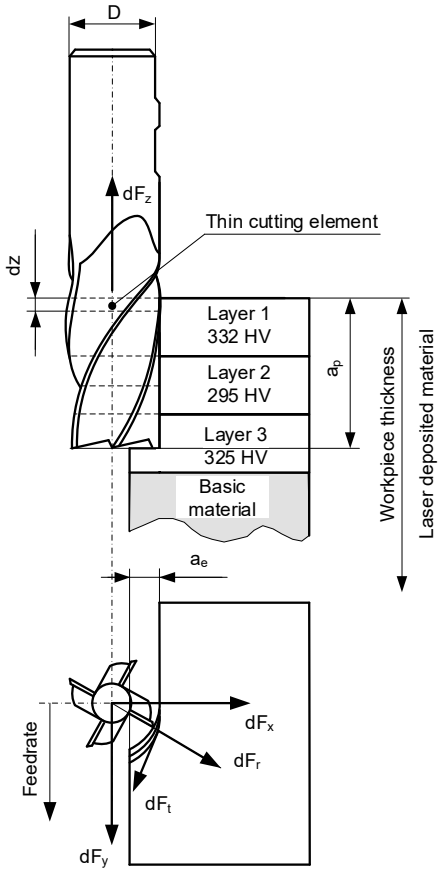


Figure 1 Cutting forces on a thin cutting disc in helical end milling of layered material.

3 ARTIFICIAL NEURAL NETWORK MODELS

The purpose of this research is to create a system of three artificial neural networks for predicting the maximum cutting forces on the cutter in the process of helical end-milling of laser-deposited metal materials. This chapter presents the architectures of three different neural networks adapted to the problem of predicting cutting forces based on a tiny sample of data for learning and testing.

Three different neural network architectures were constructed. The cutting parameters were entered into the neural networks as input data. The output data (result) was the maximum cutting force. Maximum cutting force is the maximum cutting force that occurs during one cutter rotation.

The popular multilayer feedforward neural network architecture was used for cutting force modeling. The neural networks were created in the Matlab software package using the nntraintool tool.

The neural model for predicting cutting forces is built in 4 steps. In the first step, the data for learning and validating

the neural model of cutting forces were obtained in the machining experiment.

This step made 36 measurements of the maximum cutting force on the milling cutter. Among the 36 measurements, 27 measurements were used to learn the neural networks, and 9 measurements were used to test the neural network's performance.

Tab. 1 shows 27 datasets for learning and 9 testing datasets. The test datasets are highlighted. The Tab. 1 consists of 5 columns.

The first column contains the serial number of the experiment - measurements, the second to fourth columns have the input data for the ANN, and the last column represents the output vector with the target value, i.e. the value of the maximal measured cutting force.

This step transferred the learning and testing data sets to the ANN.

Table 1 Datasets for model learning and testing.

Measurement No.:	Spindle speed (min^{-1})	Feed rate (mm/min)	a_p (mm)	Cutting force (N)
1	2000	390	0.27	174.63
2	2500	390	0.27	166.75
3	3000	390	0.27	149.38
4	3000	960	0.27	153.32
5	2500	960	0.27	186.45
6	2000	960	0.27	198.77
7	2000	2700	0.27	200.08
8	2500	2700	0.27	196.55
9	3000	2700	0.27	191.90
10	3000	6000	0.27	207.15
11	2500	6000	0.27	213.82
12	2000	6000	0.27	218.46
13	2000	390	0.72	180.89
14	2500	390	0.72	174.83
15	3000	390	0.72	151.20
16	3000	960	0.72	156.35
17	2500	960	0.72	188.77
18	2000	960	0.72	197.96
19	2000	2700	0.72	209.37
20	2500	2700	0.72	196.14
21	3000	2700	0.72	183.21
22	3000	6000	0.72	198.16
23	2500	6000	0.72	213.21
24	2000	6000	0.72	218.97
25	2000	390	1.40	188.16
26	2500	390	1.40	183.01
27	3000	390	1.40	159.68
28	3000	960	1.40	163.72
29	2500	960	1.40	191.90
30	2000	960	1.40	196.45
31	2000	2700	1.40	209.68
32	2500	2700	1.40	196.65
33	3000	2700	1.40	177.76
34	3000	6000	1.40	183.11
35	2500	6000	1.40	207.05
36	2000	6000	1.40	212.50

In step 2, the architecture of the neural network is defined.

Learning algorithms, activation functions, error for evaluating the predictive model's performance and learning parameters are defined.

The learning phase of the neural network is performed in step 3.

During the learning phase, the weights on the synapses between the neurons are set. In this way, the internal structure of the neural network adapts to the learning data and provides the correct predicted results according to the input features.

In the fourth step, the neural network model is completed and ready for use; prediction of cutting forces.

All three designed ANN architectures have three neurons in the input level: spindle speed n, feed rate f and axial depth of cut AD.

The radial depth of cut R_D was constant in all machining experiments.

The output from the ANN is the maximum cutting force generated in one tool rotation, so only one neuron is needed at the output level.

The first constructed neural network (ANN1) has one hidden level with three neurons. In one hidden layer, Sigmoid activation function is used. The scaled conjugate gradient learning algorithm (trsinscg) is used.

The performance of the learned network was determined based on Mean Squared Error (mse). The detailed architecture of the designed ANN1 with algorithm parameters and learning progress are shown in Fig. 2.

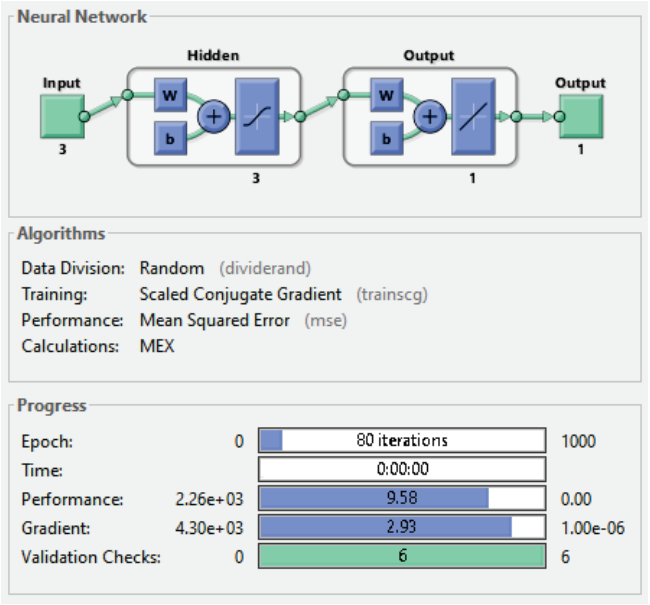


Figure 2 The detailed architecture of the designed ANN1 with algorithm parameters and learning progress.

Fig. 3 shows the learning flow of ANN1. The stopping condition on the validation set was triggered at the 72nd iteration when the MSE error value reached a minimum value of 4.2. At 72 iterations, ANN1 is the best learned and most accurately predicts cutting forces.

The gradient of the neural network is associated with the green curve. Learning stops when a stop rule is triggered, i.e. when the global minimum is reached (in the following six iterations the learning of the neural network does not improve).

The program then displays the learning results of the neural network.

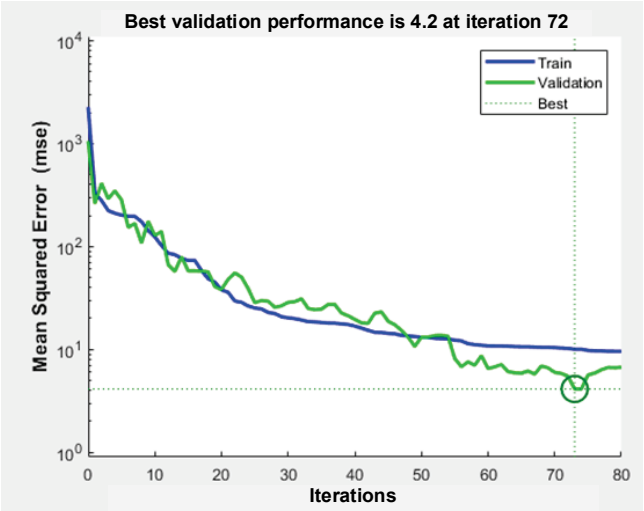


Figure 3 Learning and validation curves of the ANN1

Fig. 4 shows the match between the values of predicted (blue line) and measured cutting forces marked with circles. ANN1 compares the predicted cutting force values with the measured cutting force values. The dashed line (diagonal of the square) shows the ideal relationship between the measurements and the predicted values of ANN1. In Fig. 4, the predicted values with the dashed line almost overlap, this is shown by the value of $R = 0.98517$.

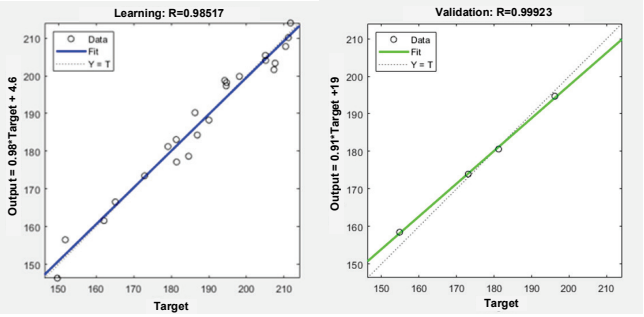


Figure 4. Learning and validation results of the ANN1 neural network.

Out of 27 measurements, four not used in the learning process were randomly selected for validation (Fig. 4). The linear regression results between the measured and predicted cutting forces is given with a value of $R = 0.99923$.

The second constructed neural network (ANN2) has one hidden level with six neurons. In the hidden layer, Sigmoid activation function is used. The Levenberg-Marquardt learning algorithm (trainlm) is used. The performance of the learned network was determined based on Mean Squared Error (mse). The detailed architecture of the designed ANN2 with algorithm parameters and learning progress are shown in Fig. 5.

Fig. 6 shows the learning flow of ANN2. The stopping condition on the validation set was triggered at iteration 12 when the MSE error value reached a minimum value of 17.4. The gradient of the ANN2 reached a global minimum at the sixth iteration. ANN2 learned with six iterations predicts cutting forces most accurately. The gradient of the neural network is associated with the green curve.

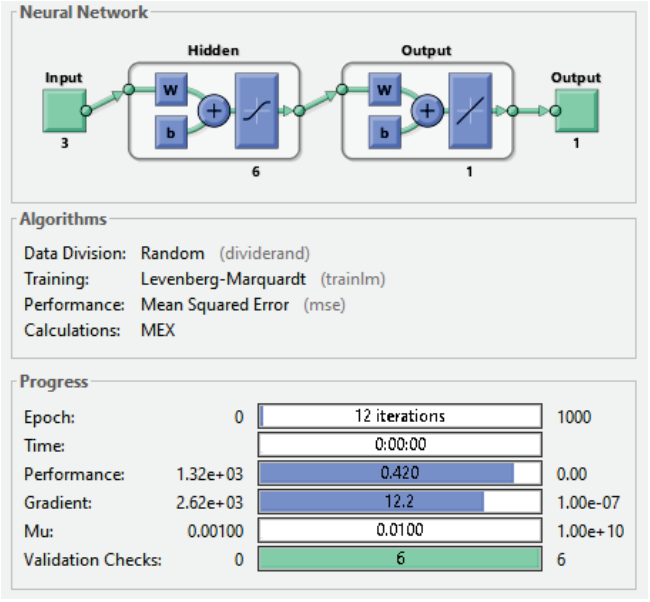


Figure 5 The detailed architecture of the designed ANN2 with algorithm parameters and learning progress.

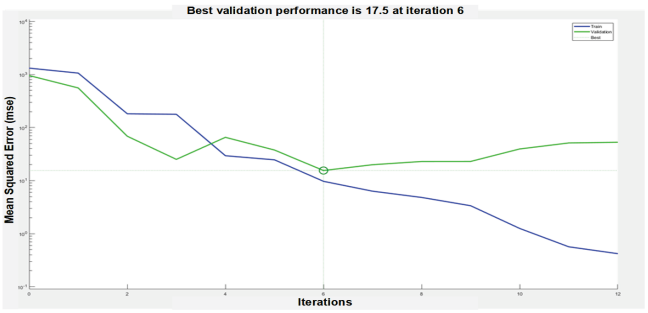


Figure 6 Learning and validation curves of the ANN2

Compared to ANN1, the match between predicted values and measured cutting forces was similar, with neural network learning achieving a ratio of $R = 0.974113$.

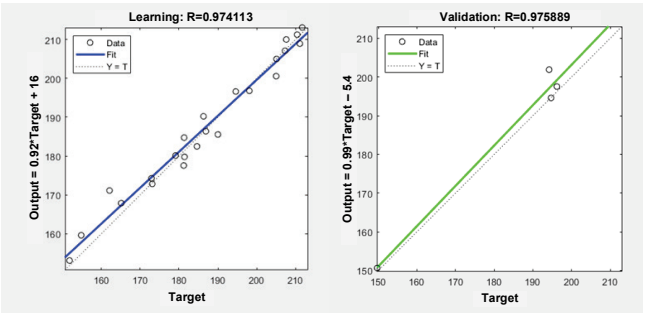


Figure 7. Learning and validation results of the ANN2 neural network.

In the validation, the linear regression results between the measured and predicted cutting forces are given with a value of $R = 0.975889$, which is similar to the results of ANN1 (Fig. 7).

The third constructed neural network (ANN3) has one hidden level with seven neurons. In hidden layer, Sigmoid activation function is used. The BFGS Quasi-Newton learning algorithm (trainbfg) is used. The performance of the

learned network was determined based on Mean Squared Error (mse). The detailed architecture of the designed ANN3 with algorithm parameters and learning progress are shown in Fig. 8.

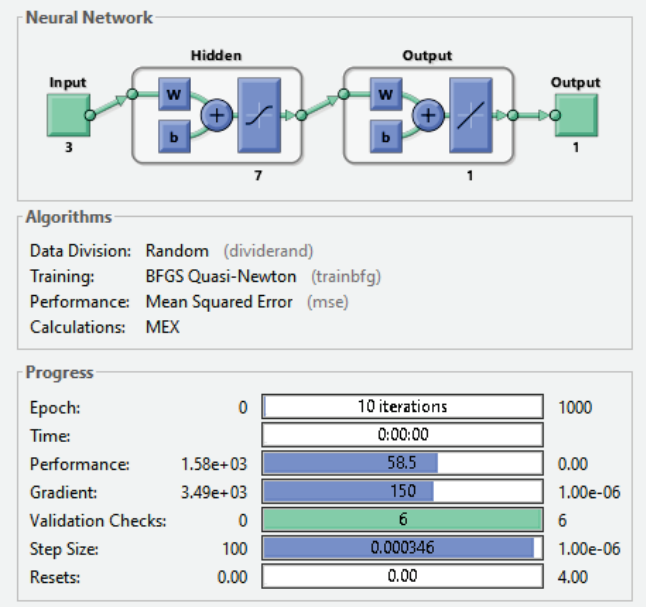


Figure 8 The detailed architecture of the designed ANN3 with algorithm parameters and learning progress.

Fig. 9 shows the learning diagram of ANN3.

The stopping condition in the validation set was triggered at the iteration 14 when the MSE error value reached a value of 37.31.

The gradient of the ANN3 reached a global minimum at 9th and 12th iterations. In the 14th iteration, the lowest MSE value was reached.

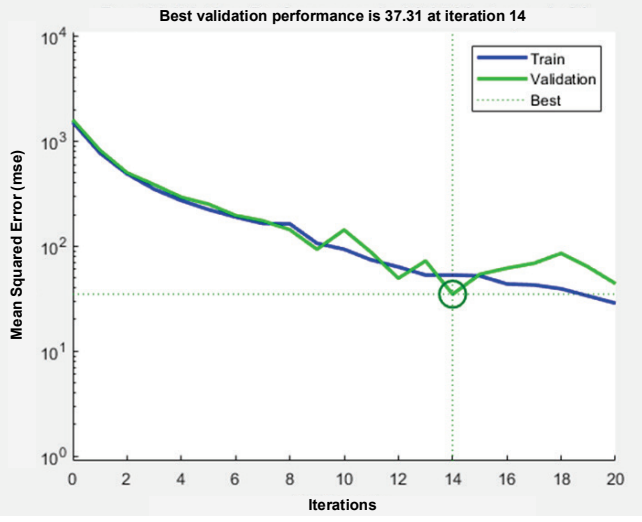


Figure 9 Learning and validation curves of the ANN3

The match between the predicted and measured values of the cutting forces during learning is given by the parameter R = 0.9359, which is the worst result of all the constructed neural networks (Fig. 10).

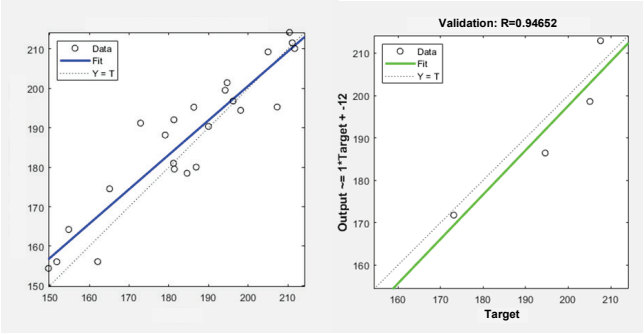


Figure 10 Learning and validation results of the ANN2 neural network.

The results of the linear regression between the measured and predicted cutting forces are given with a value of R = 0.94652, which is also the worst of all three neural networks.

4 A LINEAR REGRESSION MODEL FOR PREDICTING CUTTING FORCES

A simple linear regression model was constructed to predict cutting force as a function of cutting parameters. The designed linear regression model predicts with 90.3% accuracy.

A linear regression model for predicting cutting forces is given by:

$$F = +246.01 - 0.02901 \cdot spindle\ speed + 0.37101 \cdot feed\ rate - 0.065991 \cdot A_{\mathrm{D}} \quad (1)$$

The results of the linear regression model compared to the measured values are given in Tab. 2.

Table 2 Comparison of predicted and measured cutting forces.

No.:	Spindle speed (min^{-1})	Feed rate (mm/min)	a_p (mm)	Cutting force (N)	Predicted cutting force (N)
3	3000	390	0.27	149.4	164.84
6	2000	960	0.27	198.8	204.18
9	3000	2700	0.27	191.9	195.98
12	2000	6000	0.27	218.5	278.33
18	2000	960	0.72	198.0	203.84
24	2000	6000	0.72	219.0	277.99
27	3000	390	1.40	159.7	163.99
30	2000	960	1.40	196.4	203.33
33	3000	2700	1.40	177.8	195.13

5 RESULTS AND DISCUSSION

The research aimed to create a system of three neural network models for predicting cutting forces and to compare the results of the models with the measured values of the cutting forces and with the results of the linear regression model.

All three artificial neural networks and the linear regression model were trained with data from 27 measurements and tested with data from 9 measurements. The obtained results from neural networks and linear regression models, together with the measured cutting forces, are shown in Tab. 3 and Fig. 11.

Table 3 Comparison of model results with measurements.

No.:	Measured cutting force (N)	Cutting force (N) ANN1	ANN2	ANN3	Linear regression model
3	149.4	144.0	149.5	160.8	164.84
6	198.8	186.1	187.2	184.1	204.18
9	191.9	182.1	184.7	181.3	195.98
12	218.5	216.1	213.9	222.3	278.33
18	198.0	191.9	191.9	193.8	203.84
24	219.0	212.8	215.3	220.1	277.99
27	159.7	154.3	162.8	158.4	163.99
30	196.4	198.2	201.0	199.6	203.33
33	177.8	175.2	175.2	182.0	195.13

Table 4 Deviation of predicted values from measurements.

No.:	Measured cutting force (N)	Prediction error (%) ANN1	ANN2	ANN3	Linear regression model
3	149.4	3.60	–0.07	–7.71	–10.39
6	198.8	6.38	5.85	7.43	–2.73
9	191.9	5.15	3.75	5.54	–2.13
12	218.5	1.09	2.11	–1.78	27.54
18	198.0	3.07	3.10	2.12	2.98
24	219.0	2.84	1.68	–0.50	27.08
27	159.7	3.38	–1.93	0.83	2.70
30	196.4	–0.88	–2.33	–1.64	3.52
33	177.8	1.46	1.46	–2.39	9.81
	Model accuracy	97.13	98.65	97.21	92,52

The predicted cutting force values of the three neural networks are very similar to each other, but due to the small number of samples, the linear regression model predicted the results worse than the neural networks.

Fig. 12 and Tab. 4 show the average error of the predicted cutting force values compared to the measured cutting force.

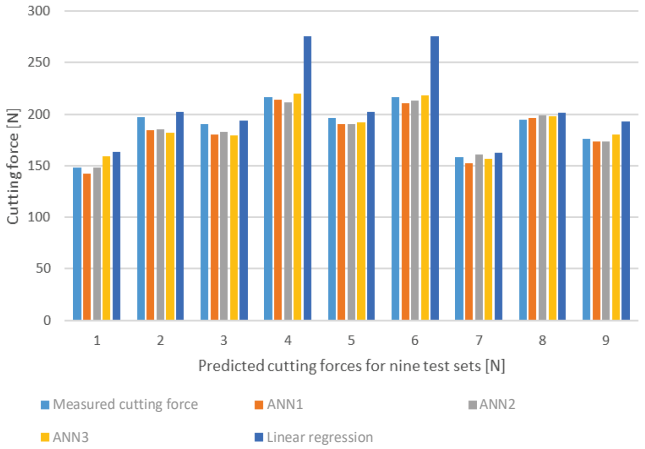


Figure 11 Comparison of model results with measurements

The results show that the neural network models are more accurate than the linear regression model and that the average error of the predicted cutting force is smaller than the model made by linear regression. Among all the neural networks, the ANN2 neural network showed the best results, with a maximum deviation of 5.85%.

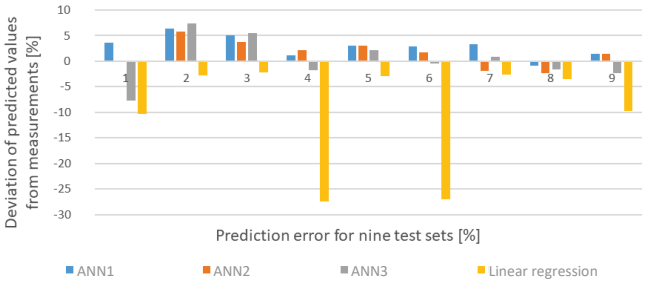


Figure 12 Deviation of predicted values from measurements

6 EXPERIMENTAL TESTS TO OBTAIN DATA

To create 3 neural networks and verify the predictive capabilities of the models, machining experiments were performed at the Heler BEA 02 machining centre. For the helical end milling process, carbide milling cutters with a diameter of 9 mm with two flutes, 28.1° helix angle and 3.80° rake angle were used. The tool material is sintered tungsten carbide with TiAlN coating and a hardness of 1800 HV.

The workpiece was produced on the Optomec LENS 850-R machine. The primary material of the workpiece is steel 20MnCr5. Three thick layers of stainless steel (316L) are applied to the 20MnCr5 steel base. The deposited layers have different properties (Fig. 1). The layers are made with 38% overlap. The thicknesses and harnesses of the individual layers are shown in Fig. 1 and range from 295 HV to 332 HV.

The maximum cutting forces that occur during one cutter rotation were measured with a Kistler 9257A piezoelectric dynamometer with a natural frequency of 3.5 kHz. A Dual Mode amplifier (NI 9215A) is used to amplify the charge. A low pass filter of 1.2 kHz cut-off frequency is used for filtering. The conversion of the charge into a voltage from 0-10V was performed by a data acquisition card manufactured by National Instruments. LabVIEW software was used to develop the measurement application. Each measurement was repeated three times.

Measurements were made at three milling depths: 0.27 mm, 0.72 mm and 1.40 mm. For each milling depth, 12 different measurements were made, which depended on feed and cutting speed. Three different tool frequencies were used: 2000, 2500 and 3000 rpm, and four different feed rates: 390 mm/min, 960 mm/min, 2700 mm/min and 6000 mm/min. All 36 measurements are shown in Tab. 1. Nine measurements were used to verify the manufactured models.

Fig. 1 schematically shows a machining experiment with a tool and a workpiece for the development and verification of models of cutting forces.

7 CONCLUSIONS

This article presents a system of three artificial neural networks models for accurate prediction of the maximal cutting forces on cutting edge of a cutter in helical end milling of layered metal material with different machinability of individual layers. The workpiece material is classified as a difficult-to-machine material. Artificial neural network models were developed in Matlab software.

The goal of the research is to replace the long-term analytical determination of maximum tool loads when machining materials with layers of different machinability by automatically creating models based on artificial intelligence from the smallest amount of experimental data.

The results provide a comparison between predicted cutting forces using linear regression and neural networks models and experimentally obtained values. It turns out that all three neural network models (ANN1, ANN2, ANN3) of cutting forces have an average error of less than 6.5% and are more accurate than the linear regression model. The maximal cutting forces predicted by the neural models are in good agreement with the experimentally obtained ones.

The second neural network (ANN2) proved to be the most accurate of the three with the largest deviation of 5.85% from all measurements and 98.65% model accuracy.

8 REFERENCES

[1] Mahmoud, E. R. I. (2015). Characterizations of 304 stainless steel laser cladded with titanium carbide particles. *Advances in Production Engineering & Management, 10*(6), 115-124. https://doi.org/10.14743/apem2015.3.196

[2] Zhang, S., Li, J. F., & Wang, Y. W. (2012). Tool life and cutting forces in end milling Inconel 718 under dry and minimum quantity cooling lubrication cutting conditions. *Journal of Cleaner Production*, *32*, 81-87. https://doi.org/10.1016/j.jclepro.2012.03.014

[3] Tabernero, I., Lamikiz, A., Martínez, S., Ukar, E., & Figueras, J. (2011). Evaluation of the mechanical properties of Inconel 718 components built by laser cladding. I*nternational Journal of Machine Tools and Manufacture, 51*(6), 465-470. https://doi.org/10.1016/j.ijmachtools.2011.02.003

[4] Jia, Z. Y., Ge, J., Ma, J. W., Gao, Y. Y., & Liu, Z. (2016). A new cutting force prediction method in ball-end milling based on material properties for difficult-to-machine materials. *The International Journal of Advanced Manufacturing Technology, 86*(9-12), 2807-2822. https://doi.org/10.1007/s00170-016-8351-8

[5] Dandekar, C. R., & Shin, Y. C. (2012). Modeling of machining of composite materials: A review. *International Journal of Machine tools and manufacture, 57*, 102-121. https://doi.org/10.1016/j.ijmachtools.2012.01.006

[6] Shokrani, A., Dhokia, V., Newman, S.T. (2012). Environmentally conscious machining of difficult-to-machine materials with regard to cutting fluids. *International Journal of Machine Tools and Manufacture, 57*, 83-101. https://doi.org/10.1016/j.ijmachtools.2012.02.002

[7] Li, G., Chandra, S., Rashid, R. A. R., Palanisamy, S., & Ding, S. (2022). Machinability of additively manufactured titanium alloys: A comprehensive review. *Journal of Manufacturing Processes, 75,* 72-99. https://doi.org/10.1016/j.jmapro.2022.01.007

[8] Altıparmak, S. C., Yardley, V. A., Shi, Z., & Lin, J. (2021). Challenges in additive manufacturing of high-strength aluminium alloys and current developments in hybrid additive manufacturing. *International Journal of Lightweight Materials and Manufacture, 4*(2), 246-261. https://doi.org/10.1016/j.ijlmm.2020.12.004

[9] M'Saoubi, R., Axinte, D., Soo, S. L., Nobel, C., Attia, H., Kappmeyer, G., Engin, S., & Sim, W. M. (2015). High performance cutting of advanced aerospace alloys and composite materials. *CIRP Annals-Manufacturing Technology, 64*(2), 557-580. https://doi.org/10.1016/j.cirp.2015.05.002

[10] Song, G., Sui, S., & Tang, L. (2015). Precision prediction of cutting force in oblique cutting operation. *The International Journal of Advanced Manufacturing Technology, 81*(1-4), 553-562. https://doi.org/10.1007/s00170-015-7206-z

[11] Sun, Y., & Guo, Q. (2011). Numerical simulation and prediction of cutting forces in five-axis milling processes with cutter run-out. *International Journal of Machine Tools and Manufacture, 51*(10-11), 806-815. https://doi.org/10.1016/j.ijmachtools.2011.07.003

[12] Omar, O. E. E. K., El-Wardany, T., Ng, E., & Elbestawi, M. A. (2007). An improved cutting force and surface topography prediction model in end milling. *International Journal of Machine Tools and Manufacture, 47*(7-8), 1263-1275. https://doi.org/10.1016/j.ijmachtools.2006.08.021

[13] Wei, Z. C., Wang, M. J., Zhu, J. N., & Gu, L. Y. (2011). Cutting force prediction in ball end milling of sculptured surface with Z-level contouring tool path. *International Journal of Machine Tools and Manufacture, 51*(5), 428-432. https://doi.org/10.1016/j.ijmachtools.2011.01.011

[14] Tuysuz, O., Altintas, Y., & Feng, H. Y. (2013). Prediction of cutting forces in three and five-axis ball-end milling with tool indentation effect. *International Journal of Machine Tools and Manufacture, 66,* 66-81. https://doi.org/10.1016/j.ijmachtools.2012.12.002

[15] Qu, S., Zhao, J., Wang, T., & Tian, F. (2015). Improved method to predict cutting force in end milling considering cutting process dynamics. *The International Journal of Advanced Manufacturing Technology, 78*(9-12), 1501-1510. https://doi.org/10.1007/s00170-014-6731-5

[16] Du, J., Li, J., Yao, Y., & Hao, Z. (2014). Prediction of cutting forces in mill-grinding SiCp/Al composites. *Materials and Manufacturing Processes, 29*(3), 314-320. https://doi.org/10.1080/10426914.2013.864402

[17] Cao, Q., Zhao, J., Li, Y., & Zhu, L. (2013). The effects of cutter eccentricity on the cutting force in the ball-end finish milling. *The International Journal of Advanced Manufacturing Technology, 69*(9-12), 2843-2849. https://doi.org/10.1007/s00170-013-5205-5

[18] Li, Y., Yang, Z. J., Chen, C., Song, Y. X., Zhang, J. J., & Du, D. W. (2018). An integral algorithm for instantaneous uncut chip thickness measuring in the milling process. *Advances in Production Engineering & Management, 13*(3), 297-306. https://doi.org/10.14743/apem2018.3.291

[19] Denkena, B., Boehnke, D., & Dege, J. H. (2008). Helical milling of CFRP–titanium layer compounds. *CIRP Journal of manufacturing Science and Technology, 1*(2), 64-69. https://doi.org/10.1016/j.cirpj.2008.09.009

[20] Karpat, Y., & Polat, N. (2013). Mechanistic force modeling for milling of carbon fiber reinforced polymers with double helix tools. *CIRP Annals, 62*(1), 95-98. https://doi.org/10.1016/j.cirp.2013.03.105

[21] Kalla, D., Sheikh-Ahmad, J., & Twomey, J. (2010). Prediction of cutting forces in helical end milling fiber reinforced polymers. International Journal of Machine *Tools and Manufacture, 50*(10), 882-891. https://doi.org/10.1016/j.ijmachtools.2010.06.005

[22] Budak, E., Altintas, Y., & Armarego, E. J. A. (1996). Prediction of milling force coefficients from orthogonal cutting data. *Journal of Manufacturing Science and Engineering, 118*(2), 216-224. https://doi.org/10.1115/1.2831014

[23] Gradišek, J., Kalveram, M., & Weinert, K. (2004). Mechanistic identification of specific force coefficients for a general end mill. *International Journal of Machine Tools and Manufacture, 44*(4), 401-414. https://doi.org/10.1016/j.ijmachtools.2003.10.001

[24] He, Y., Qing, H., Zhang, S., Wang, D., & Zhu, S. (2017). The cutting force and defect analysis in milling of carbon fiber-reinforced polymer (CFRP) composite. *The International Journal of Advanced Manufacturing Technology, 93*, 1829-1842. https://doi.org/10.1007/s00170-017-0613-6

[25] El-Mounayri, H., Briceno, J. F., & Gadallah, M. (2010). A new artificial neural network approach to modeling ball-end milling. *The International Journal of Advanced Manufacturing Technology, 47*(5-8). https://doi.org/10.1007/s00170-009-2217-2

[26] Al-Zubaidi, S., Ghani, J. A., & Haron, C. H. C. (2011). Application of ANN in milling process: a review. *Modelling and Simulation in Engineering, 9,* 1-7. https://doi.org/10.1155/2011/696275

[27] Kalla, D., Sheikh-Ahmad, J., & Twomey, J. (2010). Prediction of cutting forces in helical end milling fiber reinforced polymers. *International Journal of Machine Tools and Manufacture, 50*(10), 882-891. https://doi.org/10.1016/j.ijmachtools.2010.06.005

Authors' contacts:

Uroš Župerl, Assoc. Prof.
(Corresponding author)
University of Maribor, Faculty of Mechanical Engineering,
Smetanova ulica 17, 2000 Maribor, Slovenia
+38622207621, uros.zuperl@um.si

Miha Kovačič, Assoc. Prof.
ŠTORE STEEL. d.o.o., Štore
Železarska cesta 3, 3220 Štore, Slovenija
University of Ljubljana, Faculty of Mechanical Engineering,
Aškerčeva cesta 6, 1000 Ljubljana, Slovenia
College of Industrial Engineering Celje, Celje
Mariborska cesta 2, 3000 Celje, Slovenia
+386 (0)3 7805 262, miha.kovacic@store-steel.si

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230424191508

Original scientific paper
Received: 2023-04-24, Accepted: 2023-04-28

In-Crystal Dislocation Behaviour and Hardness Changes in the Case of Severe Plastic Deformation of Aluminium Samples

Zdenka Keran*, Amalija Horvatić Novak, Andrej Razumić, Biserka Runje, Petar Piljek

Abstract: The presence of dislocations significantly modifies the mechanical properties of crystalline solids. Severe plastic deformation (SPD) and the most used SPD process – the Equal Channel Angular Pressing (ECAP), affect the multiplication and localized accumulation of dislocations. This research is related to the observation of dislocation pile-up and significant reduction of the crystalline grain size caused by severe deformations in the ECAP process of the widely used aluminium material (Al 99.5%). Because of its lightweight, the application of Al 99.5 % can pose a challenge for the aviation and space industry, especially since its mechanical properties limit its application. Improving these mechanical properties can extend its applicability in cases of demanding constructions as well as influence the final product cost. As a confirmation of SPD in-fluence on mechanical properties, material hardness has been examined and described. Dislocation monitoring is enabled using the light and electron microscopy and AFM (Atomic Force Microscope) device. A numerical simulation of the Equal Channel Angular Pressing process using the ABAQUS software package determined the representative area of the most severe deformation.

Keywords: aluminium (Al); atomic force microscope (AFM); equal channel angular pressing (ECAP); hardness; severe plastic deformation (SPD)

1 INTRODUCTION

Each real crystal contains defects or irregularities in the crystal structure that can be dynamic or static. While dynamic defects, errors, irregularities of the crystal structure are caused by excitations of the crystal structure, static defects occur during the construction of the crystal structure or later processes such as mechanical deformations, heating, radiation, and others. Static defects can be point, line, surface or volume defects or irregularities. Many macroscopic properties of crystals can only be explained by the existence and change in the concentration of defects in crystals. From the perspective of mechanical properties, the dislocation is the most important crystal defect. It has been recognized that dislocations are the primary carriers of plastic deformation in crystalline solids, leading to ductility that makes metals workable, and through their manipulation and entanglement, hardening that makes the same metals stronger. [1].

Due to the action of external forces a glide of many dislocations occurs, which then results in a slip and leads to plastic deformation in crystalline solids. It can be thought of as sliding or successively moving one plane of an atom over another on so-called slip planes. This is uniquely defined as a plane containing both a line and a Burgers dislocation vector. Further deformation occurs either by more movement on existing planes of atoms or by creating new slip planes. The slip planes are usually the planes with the highest atom density, with the slip direction being the shortest translation of the vector. Often this direction is the one in which the atoms are closest. The sliding results in the formation of steps on the crystal surface. The direction of the slip is necessarily always parallel to the Burgers dislocation vector responsible for the sliding. The glide of one dislocation across the plane of the slip to the surface of the crystal produces a surface slip step equal to the Burgers vector [2].

The dislocation which is sliding through the crystal structure occurs within a single crystal. The remaining dislocations or grain boundaries present an obstacle for the movement of the dislocations. To overcome such obstacles, a large amount of energy and a very high force is required, which, in practice, prevents such movement. Therefore, during cold plastic deformation, dislocations accumulate within the grain and at the grain boundaries, thus making difficult further plastic deformation and increasing the strength of the deformed material.

Many scientific studies since the 1940s, up to now, relate to the experimental study and numerical calculation and simulation of the behaviour of dislocations within the crystal structure of metallic materials.

One of the first calculations can be dated back to 1940, when Burgers suggested that if an array of moving dislocations were stopped, a large local stress concentration would result [3]. This was the genesis of the dislocation pile-up, a concept which has proven to be helpful for analysing work hardening phenomena. Rosenfield and Hahn (1968) [4] have calculated the time dependent positions of dislocations emitted from a source into infinite homogenous medium and the positions of a moving array suddenly confronted with an obstacle. In 1969 Kanninen and Rosenfield [5] researched the numerical calculation of time dependent formation of single-ended dislocation pileups from the sequential emission of dislocations from the stress activated source. In 1974 Turnen [6] presented a method of numerical calculation for time dependent behaviour of several curved dislocations in a crystal. Devincre and Condat (1992) [7] made 3D computer modelling of plastic flow and model validation of a 3D simulation of dislocation dynamics. Also, several other authors were dealing with 3D computer modelling of dislocation movement in plastic deformations (Zbib 1998, Wang 2001, Yashiro 2006) [8-10].

Some techniques provided the possibility of a visual investigation of dislocations and their interactions in materials such as electron microscopy and X-ray diffraction measurement systems. Scientific research which relies on these techniques was conducted by Shen, Wagoner, and Clark (1987) [11] and Krause, Sylla and Oriwol (2016) [12].

This research work is related to the observation of dislocations pile-up caused by severe deformations of the aluminium material and, at the same time, a significant reduction of the crystalline grain size. Dislocation monitoring is enabled using the AFM (Atomic Force Microscope) device. Severe plastic deformation was performed using the ECAP (Equal Channel Angular Pressing) procedure. The ECAP procedure, as a method of achieving severe plastic deformation finds its application mostly for the hardening of titanium, copper, and aluminium alloys. ECAP application in the field of material hardening and several new interventions in the basic form of ECAP process has been described in scientific papers by Chia-Nan Wang (2016) [13] and Öğüt (2021) [14]. In this research, ECAP process has been applied on Al 99.5 % samples. This is a light weight, widely used, cheap material. Because of its light weight, the application of Al 99.5 % can pose a challenge for the aviation and space industry, especially because its mechanical properties limit its application. Improving these mechanical properties can extend its applicability in cases of demanding constructions as well as influence the final product cost.

2 SEVERE PLASTIC DEFORMATION

Since grain boundaries present an obstacle to the movement of dislocations, the fine-grained crystalline also results in an increase in the strength of the material. Hence, the modification of grain size can enable to design materials with desired properties. Physical, mechanical, and chemical properties can benefit greatly from the reduction of grain size [15, 16]. One of the possible ways for the microstructural refinement of metals is Severe Plastic Deformation (SPD).

The modern SPD technology originates from the work done by P.W. Bridgman who developed the techniques for materials processing through a combination of high hydrostatic pressure and shear deformation [17]. Severe plastic deformation leads to exceptional grain refinement of the material without introducing any significant changes in the overall dimensions of a specimen or workpiece. Materials produced by SPD techniques have grain sizes in the range of (50–1000) nm.

One of the most interesting, most frequently researched and most frequently used methods of severe plastic deformation (SPD) is Equal channel angular pressing (ECAP). This method has been proven to improve the mechanical properties of commercially pure metals, alloys, and composites [18, 19]. ECAP, like other SPD procedures, achieves extreme reduction of crystal grains by specific deformation of the material and brings their size to the nano level (even below 100 nm). Such an effect cannot be achieved or approximated by conventional heat treatment or conventional plastic processing. Most research on the angle extrusion process is devoted to relatively soft metals with Face centred cubic (FCC) crystal structure, such as aluminium and copper. However, ECAP finds its application in the case of processing more complex alloys and solid metals for which the number of slide planes along which their deformation takes place is limited [20].

2.1 ECAP Material Treatment

Plastic deformation by the ECAP process is achieved by extruding the material to be processed through two channels of the matrix that intersect at a certain angle, usually 90° [21-22]. The cross-sections of the channels are identical to the cross-section of the sample of material being processed and may be square or circular in shape. At the intersection of these channels there is a shear zone in which a large shear deformation is introduced. The amount of shear deformation is mainly defined by the geometry of the tool (inner and outer angle). A sample of the material is processed by metal removing operations to bring it to a measure where it fits the dimensions of the matrix channel before the sample is subjected to the ECAP process. The ECAP process can be repeated several times, which introduces greater plastic deformation into the workpiece without changing its dimensions.

In this specific case of experimental research work, an ECAP tool was designed (Fig. 1). The inlet and outcome channels of the die intersect at 90° (Fig. 2). They have identical, square cross sections, 14 mm in size. The ECAP process is conducted using high-capacity hydraulic press modified to slow down a ram speed. The obtained pressing speed is approximately 10 mm/s. Initial process temperature is 20 °C. Lubrication is provided using a homogeneous mixture of graphite and molybdenum disulphide. Two ECAP passes were conducted with so-called A route - when the orientation of the specimen remains unchanged after each pass. The initial material state is a cold rolled profile with an initial cold deformation of 15 %. Before the last step of cold deformation, it has been hot formed. Treated material is aluminium 99.5 %.

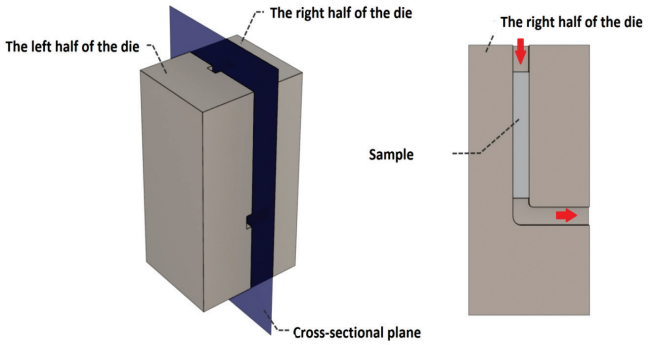


Figure 1 A scheme of the ECAP tool used in experimental research

To explore deformation status of the treated sample a numerical simulation of ECAP process was performed using Abaqus simulation software [23]. To determine the area of the highest severe plastic deformation influence, the first pass was observed. The same cross section is used in experimental monitoring of the second pass.

A three-dimensional case of the ECAP process was studied. The tool consists of the left and right halves of the die, which are symmetrical around the plane in which they are joined (Fig. 2). Process modelling is simplified with several assumptions which do not impair the accuracy of the results, but only shorten the time required to perform the

numerical analysis. The first assumption is that the entire tool and the sample itself are symmetrical around the joining plane; the possibility of setting a boundary condition of symmetry around that plane was used. In this way, the number of finite elements was significantly reduced in the discretization of the model. This directly means that the number of unknown parameters in the algebraic system of equations is also smaller, which in the end provides significant savings in the time required to solve the analysis, with unchanged results. However, to further simplify the three-dimensional analysis, the plunger was completely ejected and instead of it, an edge displacement condition of –140 mm path along the *Y*-axis was placed on the upper surface of the 140 mm long specimen. The next step before running the simulation was to define the finite elements. Three-dimensional, first-order hexahedral finite elements were selected for this analysis, using the Lagrange formulation and reduced integration - C3D8R. The number of elements on the sample is 2880 and the number of elements on the matrix is 13159. Defined friction coefficient was 0.15. Flow curve was chosen from the ABAQUS database.

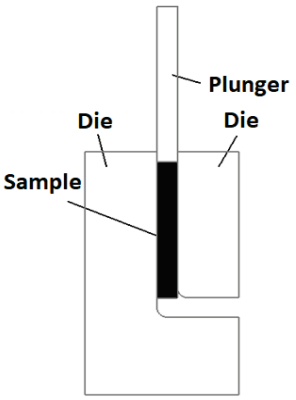


Figure 2 ECAP tool parts

According to the deformation scale, the largest equivalent deformation occurs at the cross-section A-A marked on Fig. 3. This cross section was used to examine structural changes in the sample material and changes in material hardness.

3 MATERIAL PROPERTIES – PRELIMINARY ANNOTATION

Two ECAP passes were performed on aluminium samples, using A route (the orientation of the specimen remains unchanged after each pass). Each pass resulted with changes in grain size and mechanical properties. The mechanical property observed was the hardness of the material. According to the deformation scheme of the numerical simulation, the cross section with the largest equivalent deformation was used. Testing samples for grain size and hardness examination were made by the cross cutting of ECAP deformed samples. Evaluation of the sample microstructure was made using light and electron microscopy. Equipment used in research was light microscope Olympus GX51F-5 with DP-25 CCD camera and image analysing software and scanning electron microscope VEGA TESCAN TS5136LS, Brno, Czech Republic with integrated SE and BSE detectors. The Figs. 4a), 4b) and 4c) present microstructure obtained by severe plastic deformation in ECAP procedure. Experimental results showed that ECAP deformation resulted in the grinding of crystalline grain. Grain size changes are presented in Fig. 4. After one ECAP pass the average grain size is 60 % smaller, and after two ECAP passes it is 87 % smaller than the initial grain size.

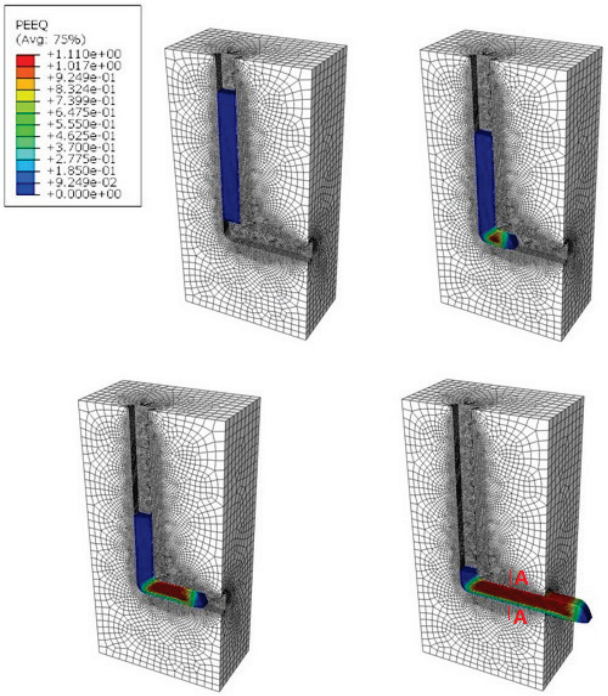
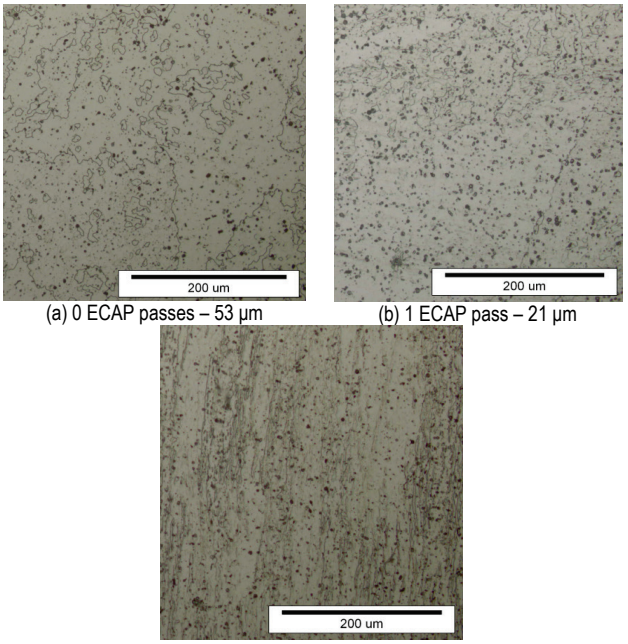


Figure 3 Equivalent plastic deformation after the first pass



(a) 0 ECAP passes – 53 μm (b) 1 ECAP pass – 21 μm

(c) 2 ECAP passes – 7 μm

Figure 4 Photo of cross section microstructure for three different deformation statuses. Grain size is calculated using the Linear Intercept Method.

Furthermore, besides changes in the crystal structure of the material, changes in mechanical properties have also occurred. The observed mechanical property was hardness, expressed as HV. It can be seen from the graph in Fig. 5 that the hardness after one ECAP pass has been increased by 50 % and after two ECAP passes by 110 % relative to the initial state of the material. Such hardness behaviour confirms the great influence of ECAP processes on the mechanical properties of the treated material.

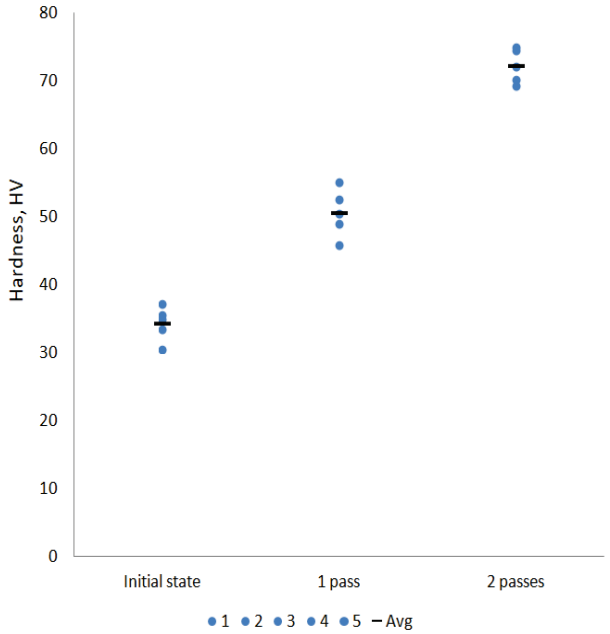


Figure 5 Hardness changes related to the number of ECAP passes

4 AFM INSPECTION OF CROSS SECTION

Dislocations on samples in origin state and after ECAP were detected using the AFM. All observations were taken on the surfaces with the largest deformation, marked as cross-section A-A (Fig. 3). Samples were prepared before scanning to adequate size 14 × 14 × 2 mm and polished using a 9 µm diamond paste. Surface analysis was conducted using the Oxford instruments model MFP-3D Origin atomic force microscope (AFM). AFM scanning was conducted using the AC mode (tapping mode). The tapping mode is a favourable inspecting mode for inspecting the materials whose surface is easily damaged. Contrary to the contact mode, where the tip is in constant contact with the surface, in the tapping mode the cantilever oscillates slightly below its resonant frequency causing lightly "taps" of the tip on the inspected surface.

The size of the investigated area spans 20 × 20 µm. Figs. 6, 7, 8, 9 and 10 show microstructures of aluminium samples. Also, reconstructed 3D images of surfaces for each sample are shown.

Arrows in the Figs. 6-10 mark dislocation areas. In the case of the initial state of the material that the dislocations are arranged relatively correctly within the structure. With the increase of the total deformation through one or two ECAP passes, the dislocations become broken and are irregularly distributed within the structure of the material. In the case of the two ECAP passes broken dislocations are pilled-up inside the small grain.

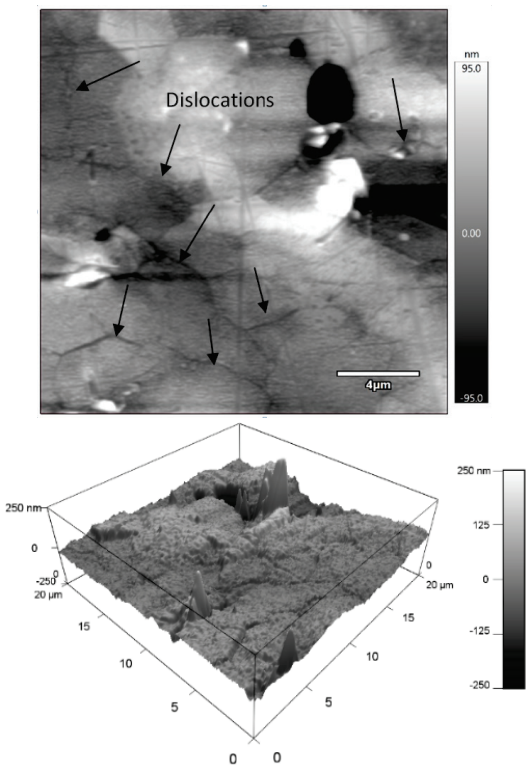


Figure 6 Microstructure of 99.5 % Al sample in origin state (before ECAP)

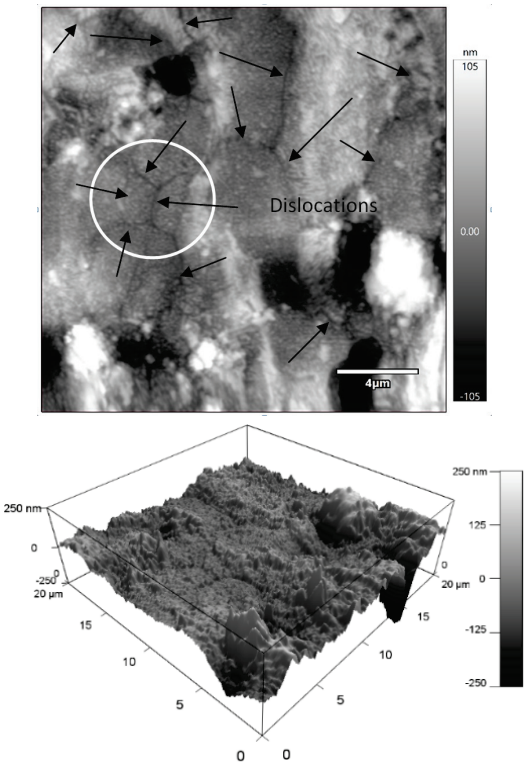


Figure 7 Microstructure of 99.5 % Al samples after 1 pass through ECAP

Since dimensions of the investigated area are 20 × 20 µm, the length and width are smaller than the average grain

diameter of the initial material state. In the case of one ECAP pass the investigated area dimensions correspond to the average grain diameter, and in the case of the two ECAP passes the investigated area dimensions are three times the grain diameter. The interior of the grain is thus clearly shown. It should be considered that in Fig. 7 and 8, some of the grain boundaries must be located within the observed area, with significantly more grain boundaries in the case of the two ECAP passes – Fig. 9. Those grain boundaries relatively disturb the correct view on dislocations because dislocations are located and pilled-up within the crystal grain. A different type of magnification of the detail marked on Fig. 10 gives a better perspective into dislocation pile-up caused by the Severe Plastic Deformation.

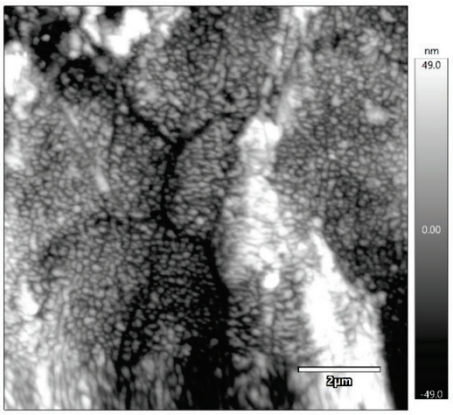


Figure 8 Detail from the microstructure of 99.5 % Al samples after 1 pass through ECAP marked with the white circle in Fig. 7 (note the different magnification)

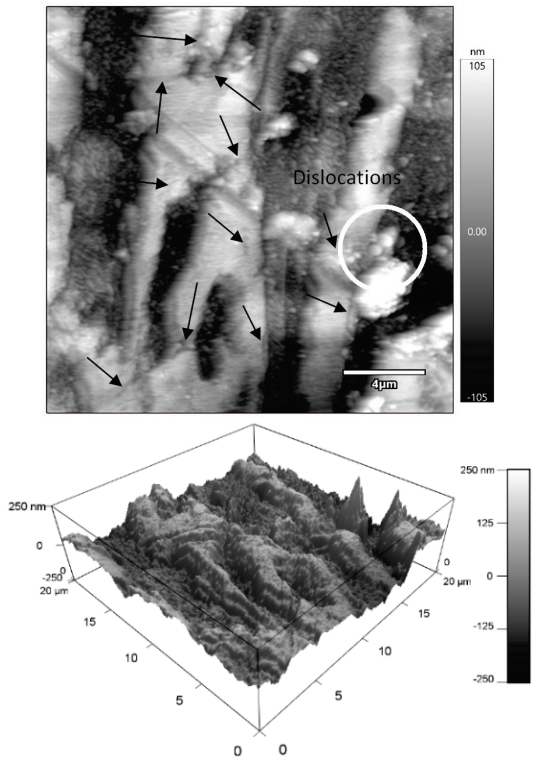


Figure 9 Microstructure of 99.5 % Al samples after 2 passes through ECAP

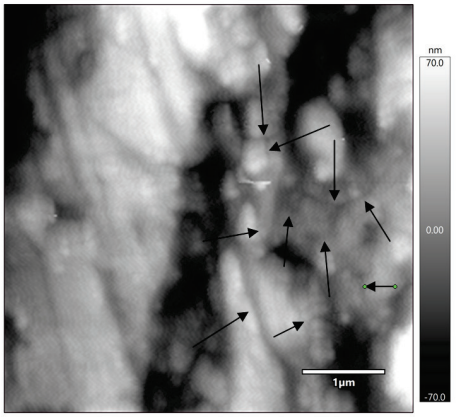


Figure 10 Details from the microstructure of 99.5 % Al samples after 2 passes through the ECAP marked with the white circle in Fig. 9 (note the different magnification)

To give a better visualization of the dimensions of individual structures encountered within the crystal structure, they are shown via diagram in Fig. 11. According to the diagram, the observed dislocations are within dimensions 10^{-5} - 10^{-8} m.

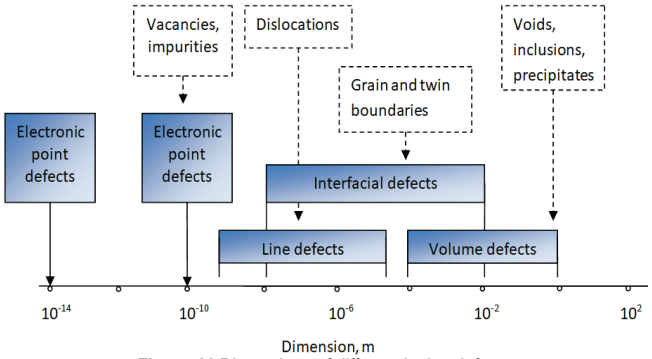


Figure 11 Dimensions of different lattice defects

5 CONCLUSION

The presence of dislocations can significantly modify the properties of crystalline solids. They are present in every crystalline material to a greater or lower extent. From the perspective of mechanical properties, the dislocation is the most influential crystalline defect. The greater the number of dislocations within the single grain, the greater the strength and hardness of the metallic material. Severe plastic deformation causes the diminishing of grain size, the increase of the number of dislocations and their accumulation within the smaller grain. In that way it influences the changes in mechanical properties of the material.

Through the presented research on the influence of the ECAP process on the structure of the aluminium Al 99.5 % material and consequently on its mechanical properties, the following can be concluded:

- After one ECAP pass the average grain size is 60 % smaller, and after two ECAP passes it is 87 % smaller than the initial grain size.
- The hardness after one ECAP pass has been increased by 50 % and after two ECAP passes by 110 % relative to the initial state of the material.

- The AFM observation showed that in the case of the initial state of the material the dislocations are arranged relatively correctly within the structure. With the increase of the total deformation through one or two ECAP passes, the dislocations are broken up and are irregularly distributed within the structure of the material. In the case of the two ECAP passes the broken dislocations are pilled-up inside the small grain.

Viewed from above it is clear that the ECAP process significantly changes and improves the structure and properties of the metallic material. Improving the mechanical properties of aluminium can extend its applicability in cases of demanding constructs as well as influence the final product cost.

6 REFERENCES

[1] Hull, D. & Bacon, D. J. (2011). *Introduction to dislocations*. The 5th ed., Oxford: Butterworth-Heinemann. https://doi.org/10.1016/B978-0-08-096672-4.00003-7

[2] Morris, J. W. Jr. (2001). Dislocation-controlled Plasticity of Crystalline Materials: Overview. In: Jürgen Buschow, K. H., Cahn, R.W., Flemings, M. C., Ilschner, B., Kramer, E. J., Mahajan, S., & Veyssière, P., editors. *Encyclopedia of Materials: Science and Technology*, Elsevier Ltd. 2245-2255.

[3] Burgers, J. M. (1940). Geometrical considerations concerning the structural irregularities to be assumed in a crystal. *Proc Phys Soc.*, 52, 23-33. https://doi.org/10.1088/0959-5309/52/1/304

[4] Rosenfield, A. R. & Hahn, G. T. (1968). Linear arrays of moving dislocations piling-up against an obstacle. *Acta Metall*. https://doi.org/10.1016/0001-6160(68)90147-8

[5] Kanninen, M. F. & Rosenfield, A. R. (1969). Dynamics of dislocation pile-up formation. *Philos Mag*. https://doi.org/10.1080/14786436908228729

[6] Turunen, M. J. (1974). Simulation of dislocation movements by a computer technique. *Philos Mag*. https://doi.org/10.1080/14786437408207256

[7] Devincre, B. & Condat, M. (1992). Model validation of a 3D simulation of dislocation dynamics: Discretization and line tension effects. *Acta Metall et Mater*. https://doi.org/10.1016/0956-7151(92)90332-9

[8] Zbib, H. M., Rhee, M., & Hirth, J. P. (1998). On plastic deformation and the dynamics of 3D dislocations. *Int J Mech Sci*. https://doi.org/10.1016/S0020-7403(97)00043-X

[9] Wang, Y. U., Jin, Y. M., Cuitiño, A. M., & Khachaturyan, A. G. (2001). Nanoscale phase field microelasticity theory of dislocations: model and 3D simulations. *Acta Mater*. https://doi.org/10.1016/S1359-6454(01)00075-1

[10] Yashiro, K., Kurose, F., Nakashima, Y., Kubo, K., Tomita, Y., & Zbib, H. M. (2006). Discrete dislocation dynamics simulation of cutting of γ′ precipitate and interfacial dislocation network in Ni-based superalloys. *Int J Plast*. https://doi.org/10.1016/j.ijplas.2005.05.004

[11] Shen, Z., Wagoner, R. H., & Clark, W. A. T. (1988). Dislocation and grain boundary interactions in metals. *Acta Metall*. https://doi.org/10.1016/0001-6160(88)90058-2

[12] Krause, A., Sylla, L., & Oriwol, D. (2016). Plastic deformation as an origin of dislocations in cast mono. *Energy Procedia*. https://doi.org/10.1016/j.egypro.2016.07.082

[13] Wang, C.-N., Lin, H.-S., Hsueh, M.-H., Wang, Y.-H., Vu, T.-H., & Lin, T.-F. (2016). The Sustainable Improvement of Manufacturing for Nano-Titanium. Sustainability. https://doi.org/10.3390/su8040402

[14] Öğüt, S., Kaya, H., Kentli, A. et al. (2021). Applying hybrid equal channel angular pressing (HECAP) to pure copper using optimized Exp.-ECAP die. *Int. J. Adv. Manuf. Technol*. https://doi.org/10.1007/s00170-021-07717-9

[15] Hassan, S., Sharma, S., & Kumar, B. (2017). A Review of Severe Plastic Deformation. *Int Refereed J of Eng and Sci*, 6, 66-85.

[16] Snopinski, P., Tanski, T., Matus, K., & Rusz, S. (2019). Microstructure, grain refinement and hardness of Al-3%Mg aluminium alloy processed by ECAP with helical die. A*rch Civ Mech Eng*. https://doi.org/10.1016/j.acme.2018.11.003

[17] Bridgman, P. W. (1946). The Effect of Hydrostatic Pressure on Plastic Flow under Shearing Stress. *J Appl Phys*. https://doi.org/10.1063/1.1707772

[18] Segal, V. M., Reznikov, V. I., Drobyshevskiy, A. E., Kopylov, V. I. (1981). Plastic Working of Metals by Simple Shear. *Russ Metall*. 1, 99-105.

[19] Venkatachalam, P., Ramesh Kumar, S., Ravisankar, B., Paul, V. T., & Vijayalakshmi, M. (2010). Effect of processing routes on microstructure and mechanical properties of 2014 Al alloy processed by equal channel angular pressing. India. https://doi.org/10.1016/S1003-6326(09)60380-0

[20] Zhao, X., Fu, W., Yang, X., & Langdon, T. G. (2008). Microstructure and properties of pure titanium processed by equal - channel angular pressing at room temperature. China, USA, Great Britain. https://doi.org/10.1016/j.scriptamat.2008.05.001

[21] Iwahashi, Y., Horita, Z., Nemoto, M., & Langdon, T G. (1998). The process of grain refinement in equal-channel angular pressing. *Acta Mater*. https://doi.org/10.1016/S1359-6454(97)00494-1

[22] Nakashima, K., Horita, Z., Nemoto, M., & Langdon, T. G. (2001). Development of a multi-pass facility for equal-channel angular pressing to high total strains. *Mater Sci Eng A*. https://doi.org/10.1016/S0921-5093(99)00744-3

[23] Programm Package ABAQUS 6.13-1 (Student Edition), Documentation & User's Guide, Dassault Systemes.

Authors' contacts:

Zdenka Keran, Assist. Prof., PhD
(Corresponding author)
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lučića 5, Zagreb 10000, Croatia
+385 1 6168 316, zdenka.keran@fsb.hr

Amalija Horvatić Novak, PhD
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lučića 5, Zagreb 10000, Croatia
+385 1 6168 486, amalija.horvatic@fsb.hr

Andrej Razumić, Mag. Ing. Mech.
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lučića 5, Zagreb 10000, Croatia
+385 1 6168 486, andrej.razumic@fsb.hr

Biserka Runje, Prof. PhD
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lučića 5, Zagreb 10000, Croatia
+385 1 6168 486, biserka.runje@fsb.hr

Petar Piljek, Assist. Prof. PhD
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lučića 5, Zagreb 10000, Croatia
+385 1 6168 383, petar.piljek@fsb.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230425171431

Original scientific paper
Received: 2023-04-25, Accepted: 2023-04-29

The Impact of Collaborative Robot on Production Line Efficiency and Sustainability

Aljaz Javernik, Robert Ojstersek, Borut Buchmeister*

Abstract: The global production tends towards more sustainable manufacturing, which forces manufacturers to constantly change and adapt. In our case, we are considering the FESTO CP LAB 400 production line, primarily designed for the training of personnel in the field of automation, which is essential for maintaining competitiveness in today's world. The production line consists of seven fully automated workstations and one manual assembly workstation, which represents the bottleneck of the production system. The paper presents a comparative study of the lines (human assembly vs. collaborative robot assembly) with an emphasis on economic and environmental aspects. The input parameters of the production line were obtained based on real-world measurements, while the assembly time of the collaborative robot was determined through simulation studies. The results of the study are presented graphically and numerically and confirm the contribution of the introduction of a collaborative robot to the assembly workstation in both financial and environmental terms.

Keywords: cobot; collaborative robot; manufacturing efficiency; production line; simulation modelling; sustainability

1 INTRODUCTION

The trend of the market has started shifting from mass production of identical products to mass-customization (low volumes and high variety). In order to maintain global competitiveness, companies strive for efficient and flexible systems that enable them to meet market demands [1]. Originally, line production was designed for mass production of standardized products, but with the emergence of new technologies and recent developments at automation and flexibility field, its use has expanded. The high cost of implementation, potential changes, and operation represent an important topic of discussion. Researchers have developed optimization models and methods to help manufacturers to achieve more effective systems. Despite the efforts of researchers, there is still a gap between real-world problems and the state of research [2]. The biggest problem in line production is bottlenecks, which determine product flow respectively system capacities. Fast and adequate identification of bottlenecks is essential to ensure adequate productivity of the system [3]. Therefore, engineers use both analytical and numerical approaches to identify bottlenecks. As the complexity of systems increases, human capabilities decrease. To cope with the complexity of the systems, the use of advanced computing environments is inevitable. The use of simulation tools allows us to model and analyse the system [4], and with the development of optimization methods and computer technology, it is now an important support and decision-making tool [5]. By using simulation, we can check the efficiency of system operation and the contribution of planned changes in a relatively short time and identify or analyse potential problem areas without interfering with the real system [6, 7]. The era of Industry 4.0 and the development of advanced technologies (IoT, Big Data, AI, remote monitoring, cloud computing) have contributed to the development of simulations. Researchers are particularly interested in the establishment of digital twins. The establishment of a digital twin not only helps to optimize operations, but also can help us in planning, maintenance, and sales by using advanced technologies [8]. The challenge of current globalization is to meet the needs of the market while ensuring all three aspects of sustainability: economic, environmental and social. The progress and development of Industry 4.0 provide various opportunities to introduce and implement sustainable manufacturing [9]. Research mentions various concepts, methods, and tools that contribute to a comprehensive treatment of sustainable manufacturing [10]. Increasingly, the development of standardized metrics for sustainable manufacturing is also being considered, which would enable different manufacturers to adopt a consistent approach [11]. To ensure efficient production and sustainable aspects at the same time, manufacturers must strive for changes at both the organizational and process levels. In recent years, the need for more flexible and efficient production systems has contributed to the increased interest in collaborative robots and their use in production processes. Collaborative robots differ from traditional industrial robots primarily in their ability to work with humans. There are also other differences between them, such as programming, manual guidance, shape, etc., but the collaboration allowed is the most important one [12]. Despite their ability to work with humans at the same time in the same space without safety barriers, collaborative robots still offer advantages such as speed, precision, and consistency [13]. The topic of collaborative robots is still relatively young, as the rise of the technology began with the development of Industry 4.0, so there is still much to explore in the world of research. However, there are already academic papers in research that showcase their contribution to efficiency and sustainability of the systems. Researchers cite various contributions of the implementation of collaborative robots into the production system and emphasize the importance of appropriate parameter settings to ensure optimal results [14]. Better support based on productivity analysis is needed for decisions on the acquisition and deployment of cobots (for single workstations and assembly lines) [15]. The role of system integrators could change in implementation projects for industrial collaborative robot applications [16]. The main challenges are in the areas of safety, knowledge, and

functionality. Collaboration with a cobot that adapts to human variability is possible and could lead to better performance and improve certain dimensions of the system's usability [17]. There is a lack of models that can evaluate the use of cobots and lead decision makers to choose the most cost-effective configuration [18]. In assembly systems, task allocation is fundamental to properly assign the available resources. It is fundamental to ensure the safety of the human operator while working with the cobot [19]. Studies show that workers are more likely to assign manual tasks to the cobot than cognitive tasks [20]. Proper implementation and parameter setting of the collaborative robot contributes to the elimination of bottlenecks, more efficient system operation [21], and system sustainability [22]. The introduction of collaborative robots also affects the social aspect, as the proximity of a collaborative robot affects both the mental and physical health of workers [23]. In particular, the use of a cobot is proposed to reduce ergonomic risks for workers assembling, for example, cable harnesses [24]. However, more and more positive aspects can be seen, as the collaborative robot takes on heavier tasks and workloads, which relieves the worker and has a positive impact on their health [25]. Due to their versatility and advantages, the use of collaborative robots is becoming more and more common.

The main goal of this paper is to compare production lines with manual assembly workstations and with collaborative assembly workstations. We wanted to show the readers how easily we can improve our production system (cost reduction, energy reduction, increase the number of finished products, etc.), even if the implementation of the cobot seems extremely expensive at the initial stage.

In this study, we aim to investigate in detail the impact of a collaborative robot on the efficiency and sustainability of the FESTO CP LAB 400 production line. With the use of real process times and simulation modelling methods, a comparative study of production lines, a line with a manual assembly workstation and a line with a collaborative robot introduced into the assembly workstation will be conducted. The results of the study will focus on the economic and environmental aspects of sustainable production.

2 PROBLEM DESCRIPTION

Ensuring efficient operations and sustainability are among the most important aspects of today's world. Companies with constant improvements strive to perfect their systems and ensure these aspects, which ultimately reflects in profit.

In this work we study the production line FESTO CP LAB 400, which is designed for training personnel in the field of automation. The line is almost completely automated, so there are few opportunities for improvement. However, despite all the automation, the line contains a manual assembly workstation that is a bottleneck in the system. The manual workstation among the fully automated workstations has a major impact on the efficiency of the line, so it makes sense to optimize the assembly workstation. The proposed optimization step is to introduce a collaborative robot to the existing manual assembly workstation.

In the research work, we first conduct a study of the existing line, in which we confirm the existence of a bottleneck (manual assembly workstation) using simulation methods. Then, with the proposed optimization step (introduction of a collaborative robot) a study of the newly created production line is performed. The result of the work is a comparative study of production lines focusing on efficiency and sustainability.

2.1 Production System Description

FESTO CP LAB 400 production line consists of 8 workstations (Fig. 1). Seven of them (①, ②, ③, ⑤, ⑥, ⑦, ⑧) are fully automated and one (④) is a manual workstation. The working process includes operations such as: inserting the front cover, measuring the height of the cover, drilling holes, assembly (inserting the circuit board and two fuses), inserting the back cover, pressing the covers, labelling and sorting the finished products.

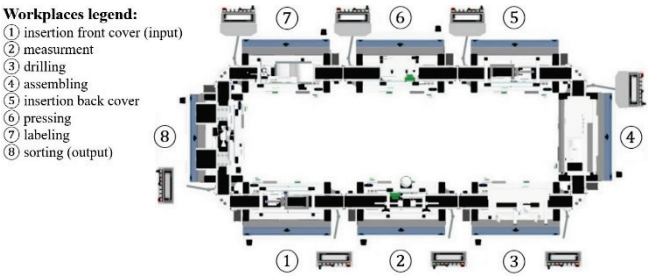


Figure 1 FESTO CP LAB 400 layout [26]

The final product is a "telephone", consisting of a front cover, a circuit board, two fuses, a back cover and a label (Fig. 2). It should be noted that the line FESTO CP LAB 400 is designed for training personnel in the field of automation, so the product complexity is low.

Figure 2 A set of components for the finished product

2.2 Sustainable Manufacturing

Ensuring aspects of sustainable manufacturing is an increasingly used concept in today's manufacturing. Sustainable manufacturing aims to improve the system in terms of environmental, economic and social aspects. All three aspects are equally important and together form a whole that allows companies to grow and have an advantage over their competitors. By ensuring sustainable manufacturing, companies strive to reduce their negative impact on the environment, improve their financial situation, and improve their social impact on people. The most common steps of sustainable manufacturing are to improve the efficiency of resources, to introduce the use of renewable resources, to implement a circular economy, to reduce excess materials, to reduce the use of hazardous substances, and to reduce the use

of energy. The mentioned steps not only address one aspect of sustainable manufacturing, but also have a significant impact on the other two aspects. Reducing energy consumption not only helps to protect the environment, but also allows companies to improve their financial situation. Reducing the use of hazardous substances not only helps preserve the environment, but also makes it more humane and less harmful to people. The introduction of modern technologies enables more efficient operation of workplaces and processes, which affects both the financial situation of the company and the well-being of its employees. There are still many steps and contributions to be made, but it must be recognized that in today's world companies need to develop in this direction to provide a better future for all of us.

There are three reasons why we decided to use a cobot rather than a robot. The first reason was safety. The FESTO CP LAB 400 production line is designed to train personnel in automation. The learning groups consist of several people and during the lessons it is difficult to observe/control all participants, so using a cobot was a safer option. The second reason is the open access to the cobot. Without a protective fence or additional safety sensors, participants can observe the operation up close and explore the functions of the FESTO CP LAB 400 in the middle of the process. The third reason is that FESTO CP LAB 400 enables the production of different products. By using the cobot, we gain flexibility. We are able to change applications quickly and easily.

3 SIMULATION MODELLING APPROACH

The research work includes two studies. The SIMIO simulation environment was used to model and simulate both scenarios. The first study examines the existing assembly line with a manual workstation, while the second study examines a newly created assembly line with a proposed collaborative robot at the assembly workstation. Real process times were used to conduct the studies. The assembly time of the worker was determined based on multiple repetitions, while the assembly time of the collaborative robot was determined using the simulation environment. The presence of a worker in the production line leads to variations in the assembly time, so based on the obtained measurements of the assembly time, a random triangular distribution (± 10 %) was proposed for the assembly time of the worker for the purpose of the simulations.

The simulation scenarios include the following assumptions:

- The simulation lasts one working day, three shifts.
- Semi-finished products are always available.
- The products are delivered in batches of 500 pieces.
- The size of the buffer is unlimited.
- The transfer time between workstations is 9 seconds.
- The simulation results are based on the average values of ten replications of the simulation model.
- The electricity price is 0.2 €/kWh.

To perform the simulation studies, it was necessary to define the cost of the workstations [27]. In previous cases the same method has given us accurate results. Tab. 1 contains only the cost analysis of the manual assembly workstation (MA) and the collaborative robot assembly workstation (CA). The other workstations remained unchanged during the simulation studies, so they are not shown in detail.

Table 1 Workstations cost calculation data

Cost calculation parameter	MA	CA
Purchase value of the machine (€)	31250	61250
Machine power (W)	97	143
Workplace area (m^2)	1.72	1.07
Depreciation period (year)	5	5
Useful capacity of the machine (h/year)	5049	5814
Machine write-off value (€/h)	1.238	2.11
Interest (€/h)	0.031	0.053
Maintenance costs (€/h)	0.062	0.105
Production system area costs (€/h)	0.041	0.022
Electrical energy consumption costs (€/h)	0.019	0.029
Machine operational costs (€/h)	1.238	2.319
Workplace total costs (€/h)	16.238	2.319
Workplace cost per batch (€/batch)	63.15	3.93

3.1 Production System Modelling

The MA workstation simulation model of the FESTO CP LAB 400 production line with an existing manual assembly workstation ④ is shown in Fig. 3. For the simulation study, the data from Tab. 2 and the assumptions from Section 3 were used. The simulation model simulated the operation of the production line of one working day, three shifts, with a useful operating time of 7.5 hours per shift, since the line had to be stopped during the lunch break. In the model itself, a utilization coefficient of 95 % was considered for fully automated workstations, while a utilization coefficient of 88 % was assumed for manual workstation due to the presence of a human.

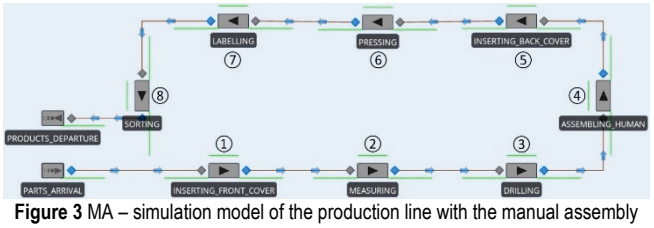


Figure 3 MA – simulation model of the production line with the manual assembly workstation

Table 2 Manual assembly-modelling parameters

Workplace	①	②	③	④	⑤	⑥	⑦	⑧
Process time (s)	1	3	4	T (25/28/31)	1	10	6	6
Operating costs (€/h)	1.27	1.27	1.27	16.24	1.27	1.27	1.27	1.27
Idle costs (€/h)	0.42	0.42	0.42	5.41	0.42	0.42	0.42	0.42
Energy consumption (W/h)	97	97	97	97	97	97	97	97

The process times in Tab. 2 were determined based on real measurements. The assembly time of the worker is the average value from several measurements. To obtain high quality results from the simulation model, a triangular random distribution was assumed for the assembly time of

the worker. The other data in Tab. 2 come from the characteristics provided by the equipment manufacturer and the assumptions made for research purposes.

The manual assembly workstation ④ represents the bottleneck of the production line, which is confirmed numerically in Section 4. The introduction of a collaborative robot was proposed as an optimization step. Fig. 4 shows the CA workstation simulation model of the newly proposed production line with the introduced collaborative robot at the assembly workstation ❹. The simulation model used the data from Tab. 3 and the same assumptions as the previous model. With the introduction of the collaborative robot, the production line became fully automated, so a utilization coefficient of 95 % and a useful operating time of 8 hours per shift were assumed for all workstations. Again, a working day with three shifts was simulated.

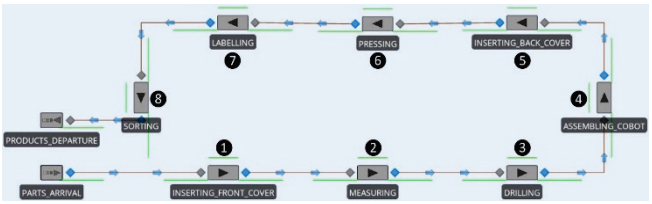


Figure 4 CA – simulation model of the production line with the collaborative robot assembly workstation

Tab. 3 shows the process times, operating and idle costs and energy consumption of each operation. Except for the process time of the assembly workstation ❹, process times of remain workstations did not change. Despite the introduction of a collaborative robot, the assembly workstation still represents the bottleneck of the line.

Table 3 Collaborative robot assembly-modelling parameters

Workplace	❶	❷	❸	❹	❺	❻	❼	❽
Process time (s)	1	3	4	12.2	1	10	6	6
Operating costs (€/h)	1.19	1.19	1.19	2.32	1.19	1.19	1.19	1.19
Idle costs (€/h)	0.40	0.40	0.40	0.77	0.40	0.40	0.40	0.40
Energy consumption (W/h)	97	97	97	143	97	97	97	97

3.2 Collaborative Workplace Modelling

To determine the assembly time of the collaborative robot, it was necessary to model the assembly workstation with the collaborative robot. The model of the collaborative robot workstation and the assembly simulation were created in Siemens Tecnomatix Process Simulate. The simulation model included the UR3e collaborative robot, the Robotiq 2F-85 collaborative gripper, a worktable, a conveyor belt, and the components required for assembly (front cover, circuit board, fuses), as shown in Fig. 5.

The collaborative robot assembly operation was performed in the same order as the manual assembly operation. In the initial stage, the collaborative robot (I) waited for a signal that the front cover (V) had arrived at the assembly location. The arrival of the front cover (V) at the specified position triggered a signal to start the assembly process. The collaborative robot (I) first inserted the circuit board (VI) into the front cover (V) and then placed two fuses (VII) on the circuit board (VI). After the assembly was completed, the collaborative robot triggered the continuation of the process.

Legend:
I – collaborative robot
II – collaborative gripper
III – worktable
IV – conveyor
V – front cover
VI – circuit board
VII – fuse

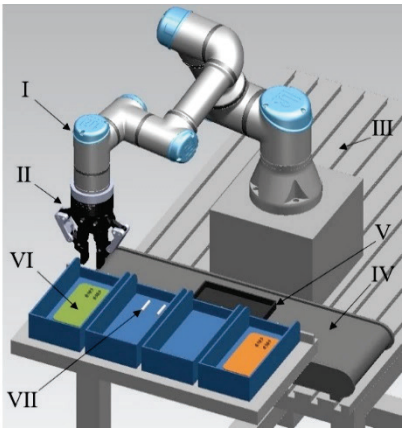


Figure 5 Simulation model of collaborative robot assembly workstation

In the simulation, the speed and acceleration of the collaborative robot from Tab. 4 were used. Move L is used to move the tool centre point linearly to a given destination at a certain speed and acceleration. All cobot joints will rotate as needed, to ensure the tool centre point stays on path with a consistent orientation. Move J is similar to Move L with the exception that the tool centre point does not move along a straight path. The cobot (tool centre point) will move to the destination along a non-linear path. The defined speed and acceleration correspond to 1/3 of the maximum value specified by the manufacturer (Universal Robots) for safety reasons. If the study were not purely academic, higher values would be used for research, but these values should be tested on a real collaborative robot to ensure proper and safe operation. Higher speed is not recommended for training.

Table 4 Collaborative robot speed and acceleration data

	Move L	Move J
Speed	333 mm/s	120 °/s
Acceleration	833 mm/s^2	267 °/s^2

4 RESULTS

In this section, the results of the MA and CA workstations scenarios are presented numerically and graphically. First, the results of each study are explained individually then a comparative study between the scenarios MA and CA is conducted. In a comparative study, we analyse the utilization of the workstations, the number of finished products, the cost per batch, and the energy consumption.

4.1 Manual Assembly

Tab. 5 shows the results of the production line (MA). Workstations ①, ②, ③ and ④ are fully utilized. Workstation ① is 100 % utilized because the study assumes that the input components are always available. Workstations ② and ③ are almost 100 % utilized, which is due to the short processing times of the previous operation. Workstation

④ represents a bottleneck of the line, which is evident from the results, as the utilization of the line from workstation ④ to workstation ⑤ drops from almost 100 % to 3.57 %, which corresponds to only 0.8 hours of operation time out of a possible 22.5 hours. The utilization of the subsequent workstations is relatively low, which is of course a consequence of the bottleneck and a low further input quantity of components. Simulation results based on ten replications of the scenario predict that the line will produce 2890 finished products and consume 12831 W of electrical power in a working day with a useful time of 22.5 hours. The cost of the batch is 71.4 €, and 2220 W of electrical energy is used to produce 500 finished products.

Table 5 Results of production line with manual assembly workstation

Workplace	①	②	③	④	⑤	⑥	⑦	⑧
Utilization (%)	100	99.99	99.97	99.96	3.57	35.70	21.42	21.41
Time processing (h)	22.50	22.50	22.49	22.49	0.80	8.03	4.82	4.82
Operating costs (€)	28.60	28.59	28.59	365.20	1.02	10.21	6.12	6.12
Idle costs (€)	0	0.001	0.003	0.05	9.20	6.13	7.50	7.50
Cost per batch (€)	0.2	0.5	0.5	65	0.2	2	1	1
Energy consumption (W)	2183	2182	2182	2182	779	1247	1039	1038

4.2 Collaborative Assembly

The introduction of a collaborative robot has contributed to the complete automation of the production line. With a fully automated line, we were able to increase the number of useful hours to 24 hours per working day. The utilization of workstations ❶, ❷, ❸ and ❹ remains unchanged compared to the previous scenario (MA), while the utilization of the other workstations in the line has increased. However, despite the introduction of the collaborative robot, the assembly workstation ❹ remains a bottleneck in the line as it is still fully utilized, followed by a significant decrease in the utilization of workstation ❺, which illustrates a lack of input semi-finished products compared to the available capacity of workstation ❺. In Tab. 6, we see that the utilization of workstation ❺ increased from 3.57 % to 8.19 % compared to workstation ⑤, which contributed to an increase in the number of finished products to 7074. The power consumption of the production line in a working day is 16467 W, which is higher than in the previous scenario (MA), since the presence of the collaborative robot must be considered. Based on the quantity of finished products, the cost of the batch is 9.05 € and 1164 W of electrical energy is consumed.

Table 6 Results of production line with collaborative robot assembly workstation

Workplace	❶	❷	❸	❹	❺	❻	❼	❽
Utilization (%)	100	99.99	99.98	99.96	8.19	81.91	49.13	49.13
Time processing (h)	24	24	24	24	1.97	19.66	11.79	11.79
Operating costs (€)	28.44	28.44	28.43	55.63	2.33	23.29	13.97	13.97
Idle costs (€)	0	0.001	0.002	0.008	8.70	1.72	4.82	4.82
Cost per batch (€)	0.15	0.5	0.5	4	0.15	1.5	1	1
Energy consumption (W)	2328	2328	2355	3431	902	2047	1538	1538

4.3 Results Comparison

Fig. 6 graphically shows the utilization of workstations in MA and CA scenarios. By introducing a cobot into the production line (CA), we were able to increase the line utilization by 13.3 % compared to the production line (MA) with the existing manual assembly workstation. However, despite the introduction of a cobot, the assembly workstation remains a bottleneck. The capacity of the cobot assembly workstation is still insufficient, considering the process times of previous operations. As shown in Fig. 6, by reducing the assembly time in the CA scenario, the utilization of the workstation ❻ increases rapidly, which would become a new bottleneck after further optimization of the cobot assembly workstation and ensuring an assembly time of less than 10 seconds.

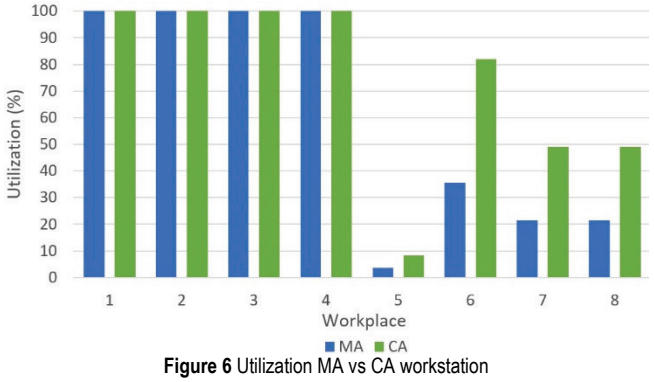


Figure 6 Utilization MA vs CA workstation

The introduction of a cobot enabled 7074 finished products in a working day, which is 144.8 % more than can be produced with the MA production line (Fig. 7).

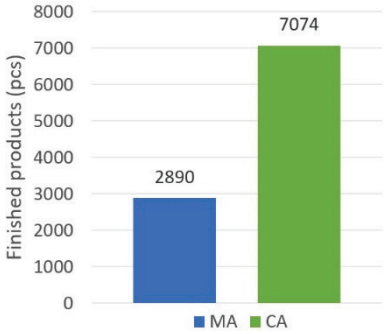


Figure 7 Finished products MA vs CA workstation

The cost per batch comparison is shown in Fig. 8 and is extremely interesting as costs drops rapidly. With the introduction of the cobot, we were able to reduce the batch cost by 87.3 %, which is difficult to imagine before running simulations given the high investment costs associated with the introduction of a cobot.

In today's world, companies are looking to reduce energy consumption as it relates to both financial and environmental aspect. The energy consumption of the collaborative robot was 3431 W per working day (Tab. 6), while the energy consumption of the manual assembly workstation was only 2182 W (Tab. 5). However,

conclusions cannot be made immediately, it is important to consider the useful number of hours per day for each production line and the number of finished products. Fig. 9 shows the energy consumption per batch. Based on further calculations, the comparison of the lines shows that the energy consumption per batch with the use of the cobot is 47.6 % lower than the energy consumption of the MA production line.

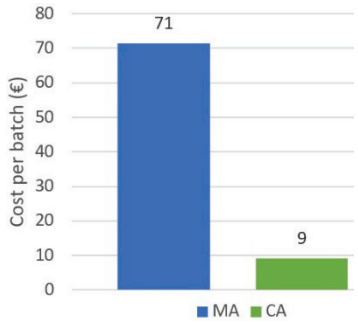


Figure 8 Cost per batch MA vs CA workstation

Figure 9 Energy consumption per batch CA vs MA workstation

5 CONCLUSIONS

In the research work, we study the production line FESTO CP LAB 400 using simulation modelling methods. In this study, we focus on the efficiency of the production line and aspects of sustainable manufacturing. The FESTO production line was developed to train staff in automation and to demonstrate the values of Industry 4.0. Automation of production is becoming increasingly widespread as companies try to achieve global competitiveness, which is difficult to achieve in today's world. The featured line demonstrates modern automation approaches very well, as it contains seven fully automated workstations and only one manual assembly workstation. Manual workstations are usually a problem in well-automated production lines, as human skills are not comparable machines.

A simulation study of the existing production line showed that the manual assembly workstation was a bottleneck in the line. The introduction of a collaborative robot at the assembly workstation was proposed as an improvement. The results of the comparative study showed that the introduction of a collaborative robot has a positive impact on the production line. With the help of a collaborative robot, the overall line utilization was increased by 13.3 %, productivity increased by 144.8 %, which meant 4182 more finished products at the same time, operating costs and batch costs decreased, and the negative environmental impact in the form of electricity consumption was reduced by 47.6 %.

Many questions remain open for further research. From the viewpoint of sustainability, the social aspect should be studied. The social aspect is definitely present due to the introduction of a collaborative robot at the worker's workplace and the removal of the worker. From the point of view of the efficiency of the production line, the parameters of the collaborative robot need to be optimally defined and not only based on the assumptions of the manufacturer of the collaborative robot.

Although the research study was only informative and based on assumptions that real production systems cannot provide, it can be concluded that the proper implementation of a collaborative robot and the adjusted parameters help to eliminate bottlenecks and have a positive impact on the efficiency of the production line FESTO CP LAB 400 and the considered aspects of the sustainable manufacturing.

Acknowledgement

The authors gratefully acknowledge the support of the Slovenian Research Agency (ARRS), Research Core Funding No. P2-0190. The authors acknowledge the use of research equipment system for development and testing cognitive production approaches in industry 4.0: Collaborative robots with equipment and Sensors, hardware and software for ergonomic analysis of a collaborative workplace, procured within the project "Upgrading national research infrastructures – RIUM", which was co-financed by the Republic of Slovenia, the Ministry of Education, Science and Sport and the European Union from the European Regional Development Fund.

6 REFERENCES

[1] Maria, A. (1997) Introduction to modeling and simulation. *Proceedings of the 29th Winter Simulation Conference*, 7-13. https://doi.org/10.1145/268437.268440

[2] Bisen, A. S. & Payal, H. (2022). Collaborative robots for industrial tasks: A review. *Materials Today: Proceedings, 52*(3), 500-504. https://doi.org/10.1016/j.matpr.2021.09.263

[3] Boysen, N., Fliedner, M., & Scholl, A. (2007). A classification of assembly line balancing problems. *European Journal of Operational Research*, *183*(2), 674-693. https://doi.org/10.1016/j.ejor.2006.10.010

[4] Cherubini, A., Passama, R., Crosnier, A., Lasnier, A., & Fraisse, P. (2016). Collaborative manufacturing with physical human–robot interaction. *Robotics and Computer-Integrated Manufacturing*, *40*, 1-13. https://doi.org/10.1016/J.RCIM.2015.12.007

[5] Grznar, P., Gregor, M., Gola, A., Nielsen, I., Mozol, S., & Seliga, V. (2022). Quick workplace analysis using simulation. *International Journal of Simulation Modelling*, *21*(3), 465-476. https://doi.org/10.2507/IJSIMM21-3-612

[6] Haapala, K. R., Zhao, F., Camelio, J., Sutherland, J. W., Skerlos, S. J., Dornfeld, D. A., Jawahir, I. S., Zhang, H. C., & Clarens, A. F. (2011). A review of engineering research in sustainable manufacturing. *ASME 2011 International*

Manufacturing Science and Engineering Conference, 2, 599-619. https://doi.org/10.1115/MSEC2011-50300

[7] Jin, W., He, Z., & Wu, Q. (2022). Robust optimization of resource-constrained assembly line balancing problems with uncertain operation times. *Engineering Computations*, *39*(3), 813-836. https://doi.org/10.1108/EC-01-2021-0061

[8] Li, L., Chang, Q., & Ni, J. (2009). Data driven bottleneck detection of manufacturing systems, *International Journal of Production Research*, *47*(18), 5019-5036. https://doi.org/10.1080/00207540701881860

[9] Lu, T., Gupta, A., Jayal, A. D., Badurdeen, F., Feng, S. C., Dillon, O. W., & Jawahir, I. S. (2011). A framework of product and process metrics for sustainable manufacturing. *Advances in Sustainable Manufacturing: Proceedings of the 8th Global Conference on Sustainable Manufacturing*, 331-336. https://doi.org/10.1007/978-3-642-20183-7_48

[10] Melesse, T. Y., Di Pasquale, V., & Riemma, S. (2020). Digital twin models in industrial operations: A systematic literature review. *Procedia Manufacturing*, *42*, 267-272. https://doi.org/10.1016/j.promfg.2020.02.084

[11] Ojstersek, R., Buchmeister, B., & Javernik, A. (2023). The importance of cobot speed and acceleration on the manufacturing system efficiency. *Procedia Computer Science*, *217*, 147-154. https://doi.org/10.1016/j.procs.2022.12.210

[12] Ojstersek, R., Javernik, A., & Buchmeister, B. (2021). The impact of the collaborative workplace on the production system capacity: Simulation modelling vs. real-world application approach. *Advances in Production Engineering & Management*, *16*(4), 431-442. https://doi.org/10.14743/APEM2021.4.411

[13] Ojstersek, R., Javernik, A., & Buchmeister, B. (2022). Importance of sustainable collaborative workplaces – simulation modelling approach. *International Journal of Simulation Modelling*, *21*(4), 627-638. https://doi.org/10.2507/IJSIMM21-4-623

[14] Schmidtler, J., Knott, V., Hölzel, C., & Bengler, K. (2015). Human centered assistance applications for the working environment of the future. *Occupational Ergonomics*, *12*(3), 83-95. https://doi.org/10.3233/OER-150226

[15] Cohen, Y., Shoval, S., Faccio, M., & Minto, R. (2022). Deploying cobots in collaborative systems: major considerations and productivity analysis. *International Journal of Production Research*, *60*(6), 1815-1831. https://doi.org/10.1080/00207543.2020.1870758

[16] Andersson, S. K. L., Granlund, A., Bruch, J., & Hedelind, M. (2021). Experienced challenges when implementing collaborative robot applications in assembly operations. *International Journal of Automation Technology*, *15*(5), 678-688. https://doi.org/ 10.20965/ijat.2021.p0678

[17] Fournier, E., Kilgus, D., Landry, A., Hmedan, B., Pellier, D., Fiorino, H., & Jeoffrion, C. (2022). The impacts of human-cobot collaboration on perceived cognitive load and usability during an industrial task: An exploratory experiment. *IISE Transactions on Occupational Ergonomics & Human Factors*, *10*(2), 83-90. https://doi.org/10.1080/24725838.2022.2072021

[18] Barravecchia, F., Mastrogiacomo, L., & Franceschini, F. (2023). A general cost model to assess the implementation of collaborative robots in assembly processes. *International Journal of Advanced Manufacturing Technology*, 125, 5247–5266. https://doi.org/10.1007/s00170-023-10942-z

[19] Faccio, M., Granata, I., & Minto, R. (2023). Task allocation model for human-robot collaboration with variable cobot speed. *Journal of Intelligent Manufacturing*. https://doi.org/10.1007/s10845-023-02073-9

[20] Schmidbauer, C., Zafari, S., Hader, B., & Schlund, S. (2023). An empirical study on workers' preferences in human-robot task assignment in industrial assembly systems. *IEEE Transactions on Human-Machine Systems*, 1-10. https:// 10.1109/THMS.2022.3230667

[21] Schou, C., Andersen, R. S., Chrysostomou, D., Bøgh, S., & Madsen, O. (2018). Skill-based instruction of collaborative robots in industrial settings. *Robotics and Computer-Integrated Manufacturing*, *53*, 72-80. https://doi.org/10.1016/J.RCIM.2018.03.008

[22] Stevanov, B., Sremcev, N., Lazarevic, M., Anderla, A., Sladojevic, S., & Vidicki, P. (2022). Optimization of the subassembly production process using simulation. *International Journal of Simulation Modelling*, *21*(4), 663-674. https://doi.org/10.2507/IJSIMM21-4-633

[23] Stock, T. & Seliger, G. (2016). Opportunities of sustainable manufacturing in Industry 4.0. *Procedia CIRP, 13th Global Conference on Sustainable Manufacturing – Decoupling Growth from Resource Use*, *40*, 536-541. https://doi.org/10.1016/J.PROCIR.2016.01.129

[24] Navas-Reascos, G. E., Romero, D., Rodriguez, C. A., Guedea, F., & Stahre, J. (2022). Wire harness assembly process supported by a collaborative robot: A case study focus on ergonomics. *Robotics*, *11*(6), 131. https://doi.org/10.3390/robotics11060131

[25] Tekin, E. & Sabuncuoglu, I. (2010). Simulation optimization: A comprehensive review on theory and applications. *IIE Transactions, 36*(11), 1067-1081. https://doi.org/10.1080/07408170490500654

[26] Festo Didactic SE (2021). *Festo Didactic – Technical Education Solutions (Doc. No. 123.664)*, 1-35, Denkendorf.

[27] Warnecke, H. J., Bullinger, H. J., Hickert, R. & Voegele, A. (1990). *Cost Accounting for Engineers*. Carl Hanser Verlag, München.

Authors' contacts:

Aljaz Javernik, MSc.
University of Maribor, Faculty of Mechanical Engineering,
Smetanova 17, 2000 Maribor, Slovenia
aljaz.javernik@um.si

Robert Ojstersek, Asst. Prof. Dr. Sc.
University of Maribor, Faculty of Mechanical Engineering,
Smetanova 17, 2000 Maribor, Slovenia
robert.ojstersek@um.si

Borut Buchmeister, Full Prof. Dr. Sc.
(Corresponding author)
University of Maribor, Faculty of Mechanical Engineering,
Smetanova 17, 2000 Maribor, Slovenia
borut.buchmeister@um.si

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230426184514

Original scientific paper
Received: 2023-04-26, Accepted: 2023-04-29

Topology Optimization of Electric Train Cable Carrier

Damir Godec*, Mario Brozović, Tomislav Breški

Abstract: In the area of mobility (transport), great efforts are being made to optimize means of transport. While reducing their mass the main goal is to maintain or even improve the performance of the means of transport. Additive manufacturing allows the production of very complex geometric shapes of products that cannot be made by traditional production processes; therefore it has become one of the significant tools in the design of modern vehicles or their parts. Additive manufacturing, combined with appropriate computer tools for numerical design analysis, also allows optimization of product design by topology optimization. In this paper, concept of topology optimization is applied to the roof carrier of a high voltage cable of electric train. Design optimization process was conducted based on two different optimization criteria: minimized part mass and maximized part stiffness.

Keywords: additive manufacturing; cable carrier; energy consumption; static load; topology optimization; weight reduction

1 INTRODUCTION

The motivation to reduce the energy consumption in transport is not only environmental in nature but has a direct correlation with the financial costs of organizations. Given that each organization aims to minimize costs and maximize profits, it is logical that the transport sector will turn to the goal of reducing the energy consumption. One of the main directions for implementing improvements in terms of reducing energy consumption has long been recognized by engineers working on the development of new aircraft, cars, trains or some other means of transport and that is the reduction in vehicle mass [1].

There have been computer tools on the market for many years that allow engineers to determine the loads and fixed points of some crucial components that they want to reduce mass and perform a numerical calculation on their own designs. The principles that allow these calculations are mainly based on numerical methods, and as the main numerical method in today's mechanical engineering in the context of reducing the mass of the structure is certainly the finite element method (FEM) [2].

Using FEM, it is possible to determine the peak stresses of the structure, as well as the displacements in some critical parts. Since conventional production processes are still largely limited by the shapes they can achieve, their use does not always make it possible to achieve a technology optimal shape of the structure. [3]

The optimal shape of the structure would be one that is often impossible to produce using conventional production processes, and structural optimization is a special branch of engineering aimed at making a structure that best stands imposed loads using minimum amount of the material. Therefore, it can be concluded that structural optimization is the main tool of engineers in the process of reducing the total mass of the structure, while the production of optimized parts is increasingly achieved using modern additive manufacturing (AM). [3, 4]

One of the most widespread methods of structural optimization in today's mechanical engineering is topology optimization. Topology optimization is still used as a kind of tool to get an idea during the conceptual phase of the product development, but with the introduction of additive manufacturing, it is increasingly used as a tool that gives the final form of the product. [3]

This paper presents an example of topology optimization design of the electric train cable carrier, based on two optimization criteria: minimizing cable carrier mass and maximizing cable carrier stiffness (minimizing compliance).

2 TOPOLOGY OPTIMIZATION

Optimizing structures is a way of obtaining structures that meet certain requirements, at imposed restrictions. The range of design optimization is usually divided into dimensional (size) optimization, shape optimization, and topology optimization (TO). [4, 5]

Topology optimization is basically the determination of the optimal distribution of materials in the design space that minimizes (or maximizes) the target function, while meeting the set limits [4]. The target function, for example, can be to minimize compliance, i.e. to maximize stiffness, for static problems, or to maximize the base frequency or frequency range for dynamic problems [6]. In the continuous approach, the design variables are the number of the gaps, their connection, shape and location (Fig. 1a), while in case of discrete approach, the variables are the thickness or cross-sectional surfaces of the design elements (Fig. 1b) [7].

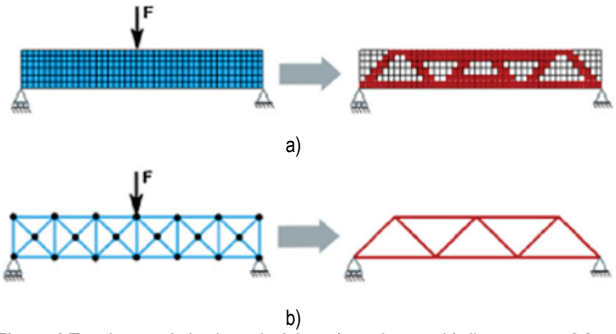


Figure 1 Topology optimization principles: a) continuous, b) discrete case [7]

Topology optimization, as a pre-processor for shape and size optimization, in its most general setting, should consist of finding out within the design domain Ω the best material distribution that minimizes an objective function f (Fig. 2).

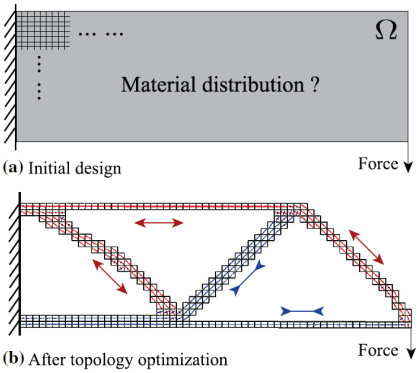


Figure 2 a) Initial design aimed for topology optimization, b) topology optimized design with stress presentation [8]

The traditional approach to topology optimization is to discretize the domain into a network of finite elements, which represents an isotropic solid microstructure. The distribution of material density within the project domain, variable $\rho(x)$, is discrete, and for each element is assigned a binary value [9]:

- $\rho = 1$, where material is required,
- $\rho = 0$, where the material is removed,

or describes whether the material exists at point $x \in \Omega$, as seen in Fig. 3.

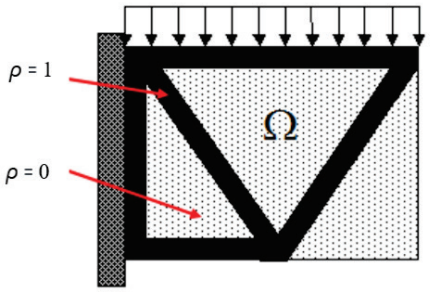


Figure 3 Distribution of material density - variable ρ(x) [9]

The most popular mathematical method for topology optimization is the Solid Isotropic Material with Penalization method (SIMP). The SIMP method predicts an optimal material distribution within a given design space, for given load cases, boundary conditions, manufacturing constraints, and performance requirements. The SIMP method is also used in Altair Inspire software, used in this paper. Topology optimization in general can be divided into 8 steps [10]:

1) Define design space
2) Define non-design space
3) Define boundary conditions
4) Define constraints and objectives
5) Define optimization settings
6) Solve
7) Interpret the results
8) Validate.

3 TOPOLOGY OPTIMIZATION OF TRAIN CABLE CARRIER

The primary function of the high-voltage cable of an electric train is to fasten the cable from undesirable shear when acting G forces and to direct the cable at a certain angle to the end devices. Its position in the CAD environment within the partial design of the train roof can be seen in Fig. 4.

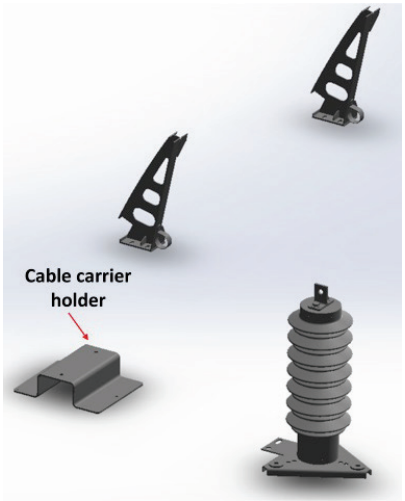


Figure 4 Cable carrier holder position in CAD environment [11]

3.1 Simulation Model Preparation

Before the execution of the cable carrier topology optimization, it is necessary to create the initial appearance of the structure. In order to exploit the full potential of the software and the topology optimization, the design will be quite primitively, solely for the purpose of performing the primary function of attaching and routing the cable as well as the possibility of attaching to the intended holder. The initial design of the cable carrier is shown in Fig. 5.

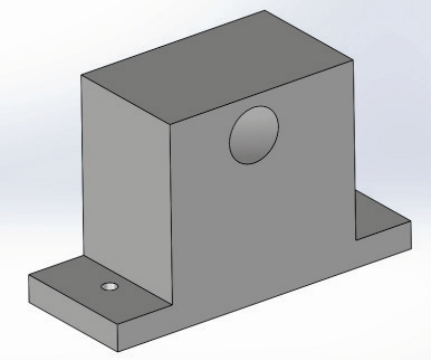


Figure 5 CAD model of initial cable carrier design [11]

The central bore, through which the cable passes, is made using a 3D sketch within assembly display, and using the material removal option per path shown in Fig. 6.

Topology optimization of the cable carrier design is performed in two ways. One way was optimization of the structure in such a way, that its mass is minimized, and to keep its mechanical resistance to the loads the unchanged.

Another way was to increase the stiffness of the structure with a certain limitation of the utilization of the volume space of the design domain according to multiple simulations. In this chapter, optimal design shapes for the two types of topology optimization will be presented.

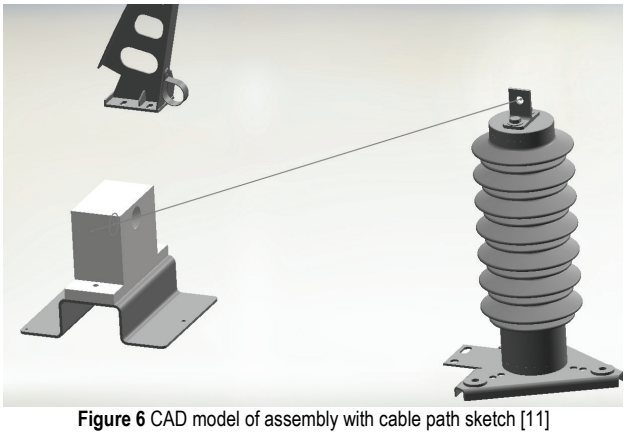


Figure 6 CAD model of assembly with cable path sketch [11]

When implementing topology optimization, it is necessary to determine the design and non-design space, and this is determined as part of the Altair Inspire program by isolating parts of the structure that do not enter the domain. First of all, it is necessary to make a "separation" bore from the volume limit of topology optimization, i.e. to make cylindrical shapes in holes that have a certain thickness and after topology optimization their shape will not change (Fig. 7).

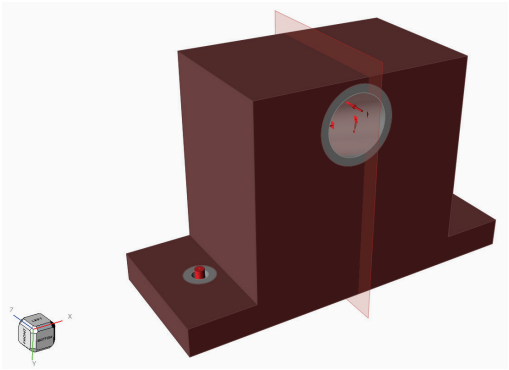


Figure 7 Prepared carrier CAD model for simulation (symmetric option) [11]

Before staring the design analysis, cable carrier material should be selected. As the final design should be 3D printed, as a referent material PLA (Polylactic acid) is selected.

3.2 Initial Design Static Analysis

Since the initial design is quite massive, the results shown in the Figs. 8 and 9 were expected. The displacements and peak stresses of the structure are almost negligible, while the minimum safety factor at the places of the supports is 3.5, which means that the initial design will not fails during the load (Figs. 8 and 9).

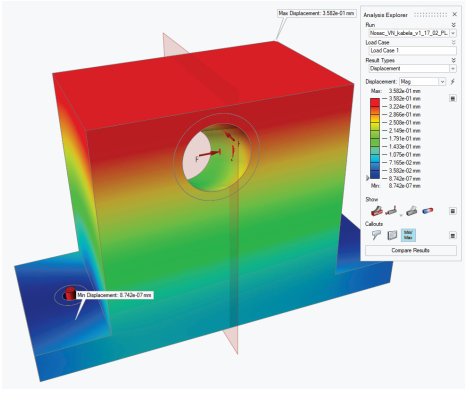


Figure 8 Initial design - displacement peak values [11]

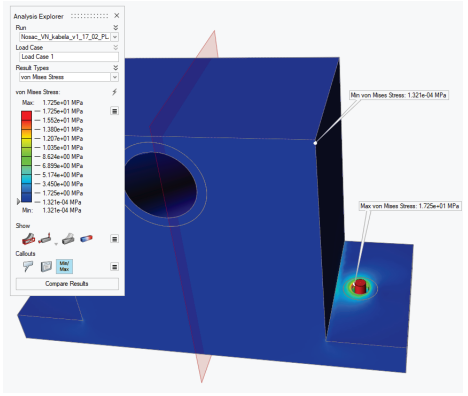


Figure 9 Initial design – stress peak values [11]

Here it can be observed that sharp transitions on the structure will be stress concentrators. Thus, it can be assumed that after topology optimization, it is the edge connecting the lower surface and the central part of the structure that will have the highest levels of stress.

Fig. 10 presents obtained distribution of safety factor over the initial cable carrier design.

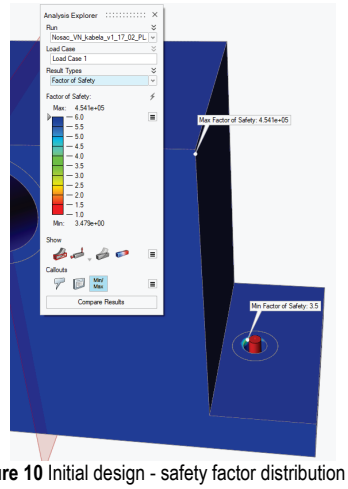


Figure 10 Initial design - safety factor distribution [11]

At this stage it is necessary to read the mass of the initial structure as a reference data for the end of the topology optimization. For given material (PLA), initial mass is 2.58 kg.

3.3 Initial Design Topology Optimization – Minimized Mass

The topology optimization of the existing initial design to minimize mass is performed with selection of safety factor 2 (Fig. 11). Selecting of this factor is the subjective estimate for making the first prototype. By conducting multiple simulations, with safety factors 1.2 and 1.5 it was concluded that, when minimizing mass, the program defines sharp transitions in structure with a lack of material. Selecting safety factor 2 gives a slightly more massive transition to the edge and will present itself as the optimal solution within this work.

Figure 11 Topology optimized design – minimized mass [11]

Obtained design must primarily be statically analysed as it was done for the initial design and evaluated the results of the analysis.

3.4 Topology Optimized Design Static Analysis – Minimized Mass

As a part of the analysis of the optimization results, the same parameters will be observed as with the initial design. When analysing peak safety factor results, it was found that there was present design error, where the right support safety factor was 0.9, which is a clear sign that the design would not withstand the load. A way to correct this is to return to the result of topology optimization and increase the wall thickness of the weak structure, while the shape of the structure itself has not changed significantly. The results of analysis for such a thicker structure, are given in following figures.

If the maximum displacement of the topology optimized structure is compared with the initial design (Fig. 8), here the values are almost 10 times the initial value, which is not negligible, but also not critical in this specific application.

By analysing the peak stress values according to Von Mises' strength theory, it is evident that it is a magnification of the maximum stress of almost 2 times (Fig. 13). Since in the area of maximum stress the safety factor is still greater than 1, this will also not be considered as a critical value. Also, it is important to emphasize that a support that appears to be under more stress is one in the direction of which the inertial force acts when braking the train. Analysis of the peak values of safety factors shows that uniformity was not achieved throughout the entire optimized model (Fig. 14).

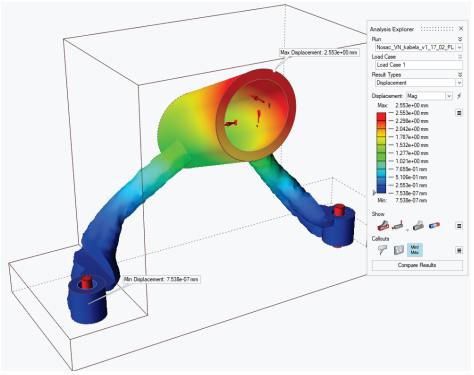


Figure 12 Topology optimized design – displacement distribution (min. mass) [11]

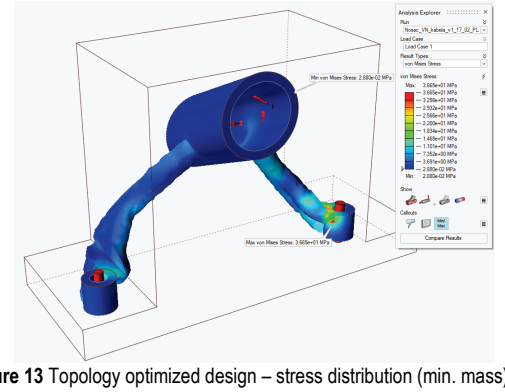


Figure 13 Topology optimized design – stress distribution (min. mass) [11]

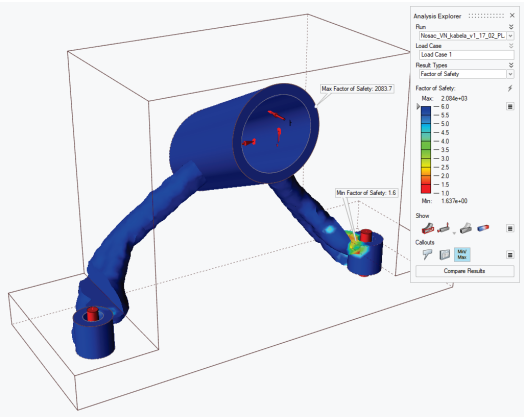


Figure 14 Topology optimized design – safety factor distribution (min. mass) [11]

Uniformity would occur if each point of a topology optimized design had a safety factor value lager than 2, as this was the basic condition given when defining optimization criteria. The value of the minimum safety factor is 1.6, and for the purpose of this paper it can be considered acceptable, so such design will be post-processed.

3.5 Topology Optimized Design Post-Processing – Minimized Mass

The post-processing of the topology optimized design is done for the aesthetic reason and to reduce the impact of

sharp edges, for example by implementing rounding. As a part of Altair Inspire software, this is done in such a way that either automatically or manually "withdraws" material within the domain specified by topology optimization (Fig. 15).

Figure 15 Post-processed topology optimized design (minimized mass) [11]

The most important aspect of this topology optimization was to minimize the mass of the cable carrier – total mass of the optimized design was 0.10 kg. Thus, the optimized carrier has only ~3.9 % of the initial structure mass, which means that the topology optimization of this carrier saved over 96 % of the cable carrier mass. Consequently, this result can be used as an indicative input parameter for the next topology optimization step, which is optimization with the aim of increasing structure stiffness.

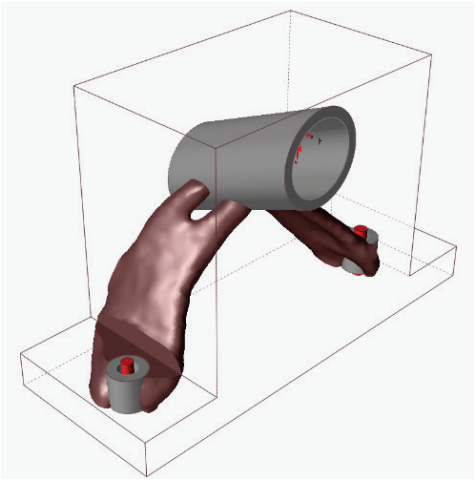


Figure 16 Topology optimized design – maximized stiffness [11]

3.6 Topology Optimized Design – Maximized Stiffness

Topology optimization with the aim of maximizing stiffness is the reverse of that with the aim of minimizing mass. Thus, it is logical to expect some conflicting results of the two analyses of the optimized design. The optimization process is performed in the same way as before, but instead of selecting the *Minimize mass* target function, *Maximize stiffness* is selected from the Altair Inspire menu. The result of this topology optimization step is shown in Fig. 16.

3.7 Topology Optimized Design Static Analysis – Maximized Stiffness

As it was the case for the previous type of topology optimization, static analysis will also be performed for the design obtained with optimization criterion – maximized stiffness. Fig. 17 presents results of displacement distribution and peak values.

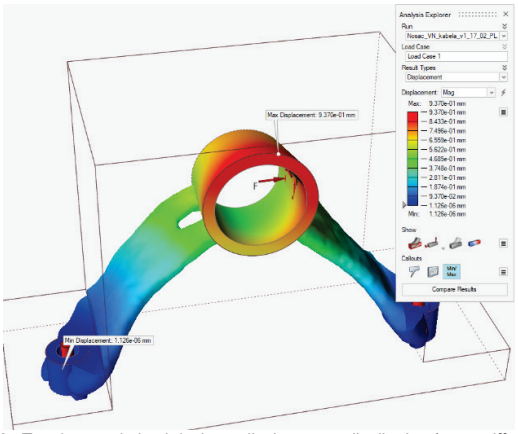


Figure 17 Topology optimized design – displacement distribution (max. stiffness) [11]

Analysing the results and comparing with previous one, it can be concluded that the value of the maximum displacement is almost 2.5 times less than the maximum displacement of the previously topology optimized structure. Consequently, it is concluded that the value of the maximum displacement is by no means critical for the optimized design.

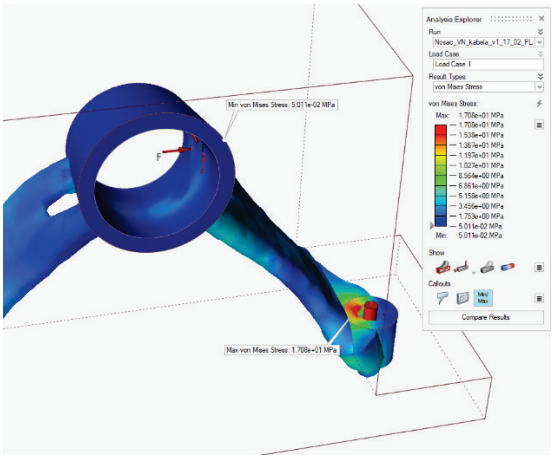


Figure 18 Topology optimized design – stress distribution (max. stiffness) [11]

According to the results of the analysis of the stress distribution and peak values of the topology optimized design (Fig. 18), it can be concluded that the maximum stress is also in the right support area, but here the stress value is almost the same as in the initial structure. Thus, it can be further concluded that such an optimized design will have similar peak values of stress as the initial design but will occupy a

much smaller volume and consequently have a significantly smaller total mass.

From the Fig. 19, it can be concluded that the minimum value of the safety factor of optimized design is the same as of the initial structure.

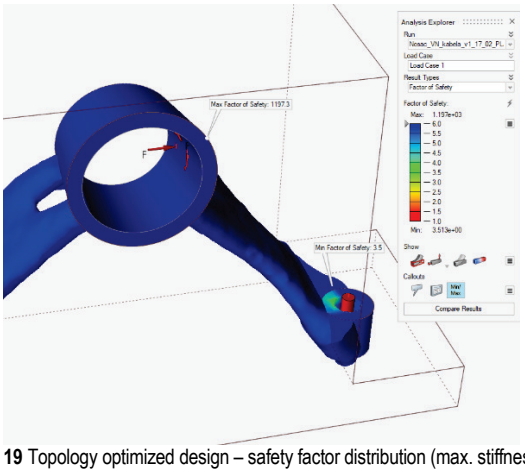


Figure 19 Topology optimized design – safety factor distribution (max. stiffness) [11]

Fig. 20 shows post-processed optimized design for maximized stiffness. The final mass of such optimized design is only 0.22 kg.

Figure 20 Post-processed topology optimized design (max. stiffness) [11]

Thus, the optimized carrier has only ~8.8 % of the mass of the initial structure, which means that the topology optimization of this carrier saved over 91 % of the mass of the predicted volume domain. Tab. 1 shows a summary and comparison of main parameters peak values of all three observed designs.

Table 1 Summary of topology optimization results [11]

Design	Max. stress (MPa)	Min. safety factor	Mass (kg)
Initial	17.35	3.5	2.58
TO – min. mass	36.65	1.6	0.10
TO – max. stiffness	17.08	3.5	0.22

4 CONCLUSION

This paper presents the process of topology optimization of the cable carrier of an electric train using the Altair Inspire software. The paper is structured as a case study, where two types of topology optimization on the same design are displayed and the results of their analysis are compared.

Structure topology optimization significantly reduces its mass, in both cases. Both topology optimized structures have a mass of less than 10 % of the initial structure, which means that a mass reduction of 90 % and more has been achieved. The first topology optimized design is slightly weaker than the second, mainly due to the topology optimization criterion. By minimizing mass, the software searches for the optimum in such a way as to sacrifice the stiffness of the structure, using the safety factor criterion assigned to it. The second topology optimized design is heavier than the first approximately 2 times but achieves significantly better stiffness properties and a safety factor the same as that of the initial structure. It can be concluded that using topology optimization on a given cable carrier drastically reduces its mass while retaining acceptable mechanical properties.

Acknowledgements

This paper is part of the research included in project Increasing Excellence on Advanced Additive Manufacturing. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 810708. The authors would like to thank the EU for the financing of the project. The authors also would like to thank the Altair Engineering Inc for providing licence of Altair Inspire software.

5 REFERENCES

[1] Zawadzki, P. & Żywicki, K. (2016). Smart Product Design and Production Control for Effective Mass Customization in the Industry 4.0 Concept. *Management and Production Engineering Review*, *7*(3), 105-112. https://doi: 10.1515/mper-2016-0030

[2] Sorić, J. (2021). *Metoda konačnih elemenata – Linearna i nelinearna analiza konstrukcija*. Golden Marketing – Tehnička knjiga. (in Croatian)

[3] Diegel, O., Nordin, A. & Motte, D. (2020). *A practical guide to design for Additive Manufacturing*. Springer. https://doi.org/10.1007/978-981-13-8281-9

[4] Bendsøe, M. P. & Sigmund, O. (2004). *Topology optimization – Theory, methods and applications*. Springer. https://doi.org/10.1007/978-3-662-05086-6

[5] Liu, J., Gaynor, A. T., Chen, S. et al. (2018). Current and future trends in topology optimization for additive manufacturing. *Structural and Multidisciplinary Optimization*. Springer. https://doi.org/10.1007/s00158-018-1994-3

[6] Christensen, P. W. & Klarbring, A. (2009). *An Introduction to Structural Optimization*. Springer.

[7] Saadlaoui, Y., Milan, J. L., Rossi, J. M. & Chabrand, P. (2017). Topology optimization and additive manufacturing: Comparison of conception methods using industrial codes. *Journal of Manufacturing Systems, 43*(1), 178-186. https://doi.org/10.1016/j.jmsy.2017.03.006

[8] Meng, L., Zhang, W., Quan, D. et al. (2020). From Topology Optimization Design to Additive Manufacturing: Today's Success and Tomorrow's Roadmap. *Archives of Computational Methods in Engineering*, *27*, 805-830. https://doi.org/10.1007/s11831-019-09331-1
[9] Dassault Systemes. SIMP Method for Topology Optimization. SOLIDWORKS Help. https://help.solidworks.com/2019/English/SolidWorks/cworks/c_simp_method_topology.htm.
[10] Godec, D., Gonzalez-Gutierrez, J., Nordin, A., Pei, E., & Alcázar, J. U. (2022). *A Guide to Additive Manufacturing*. Springer. https://doi.org/10.1007/978-3-031-05863-9
[11] Brozović, M. (2022). *Topološko optimiranje nosača visokonaponskog kabela elektromotornog vlaka*. Završni rad, Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje. (in Croatian)

Authors' contacts:

Damir Godec, PhD, Full Prof.
(Corresponding author)
University of Zagreb,
Faculty of Mechanical Engineering and Naval Architecture,
Ivana Lučića 5, HR-10000 Zagreb, Croatia
+38516168192, damir.godec@fsb.hr

Mario Brozović, student
University of Zagreb,
Faculty of Mechanical Engineering and Naval Architecture,
Ivana Lučića 5, HR-10000 Zagreb, Croatia
mb494197@stud.fsb.hr

Tomislav Breški, mag. ing. mech.
KONČAR - Električna vozila d.d.
Ante Babaje 1, HR-10090 Zagreb, Croatia
tomislav.breski@koncar-kev.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221220200605

Preliminary communication
Received: 2022-12-20, Accepted: 2022-12-22

OpenPose based Smoking Gesture Recognition System using Artificial Neural Network

Tae-Yeong Jeong, Il-Kyu Ha*

Abstract: Smoking is an extremely important health problem in modern society. This study focuses on a method for preventing smoking in non-smoking areas, such as public places, as well as the development of an artificial neural network based smoking motion recognition system for more accurately recognizing smokers in such areas. In particular, we attempted to increase the rate of recognition of smoking behaviors using an OpenPose based algorithm and the accuracy of such recognition by additionally applying a hardware device for recognizing cigarette smoke. In addition, a preprocessing method for inputting a dataset into the proposed system is proposed. To improve the recognition performance, four types of dataset models were created, and the most suitable dataset model was selected experimentally. Based on this dataset model, test data were created and input into the proposed neural network based smoking behavior recognition system. In addition, the nearest neighbor interpolation method was selected experimentally as an image interpolation approach and applied to the image preprocessing. When applying experimental data based on learned data, the developed system showed a recognition rate of 70-75%, and the smoking recognition accuracy was increased through the addition of the hardware device.

Keywords: artificial neural network; OpenPose algorithm; smoking gesture recognition system; smoking recognition sensor

1 INTRODUCTION

Smoking is an important issue in modern society and is directly related to health and the environment. In the past, people smoked everywhere; however, as the awareness of secondhand smoke has increased, the perception of smoking and its health effects has changed, and negative views on smoking have increased, resulting in an increase non-smoking areas. However, smoking in public places, such as stations, schools, and downtown areas where many people gather, is still common despite laws prohibiting such activity, causing various problems including environmental pollution and second-hand smoke [1]. In the Republic of Korea, according to Article 9 of the National Health Promotion Act, a non-smoking area within 10 m of the outer wall of a public building is designated as a non-smoking area [2]. However, indiscriminate smoking frequently occurs near buildings that are not designated as smoking areas. In addition, although smoking areas are used in many locations to prevent smoking in non-smoking areas, smokers often do not use these areas for various reasons. For example, in many cases, the smoking areas are not fully functional [3-5]. Therefore, this study focused on a method for preventing smoking in non-smoking areas. We developed an artificial neural network-based smoking behavior recognition system. In particular, we used the OpenPose-based algorithm to design our recognition model. Several characteristic tasks, such as preprocessing of training data, selection of efficient training datasets, and interpolation of test data, were performed to improve the performance of the recognition module. A hardware device was developed for judging smoking behavior by recognizing cigarette smoke, to guarantee the recognition of smoking behavior and improve its accuracy.

2 RELATED WORK

2.1 Analysis of Existing Studies

Tab. 1 list the characteristics of existing studies on human motion recognition and major studies on the OpenPose based skeleton model. In [6], a skeleton model is used with the part affinity fields (PAF) method, which connects points placed on the joints of the body, and captures the posture of an individual with a relatively high accuracy. In [7], although a partitioning method is proposed for recognizing multiple people in an image using a deepcut structure; it has a limitation in that it takes a long time to recognize a single image. In [8], an AI-based recognition technique is proposed for recognizing human hands as objects; however, it has difficulty estimating various human postures. The approach in [9] has the advantage of directly learning the inference process, unlike the general object recognition methods from previous studies, but has difficulty detecting complex or rare postures and obtaining datasets for rare postures for artificial intelligence learning. In [10], the spatio-temporal affinity fields (STAF) method is used to strengthen the existing PAF skeleton model method. However, when a large number of objects appear, the computational speed increases linearly, and the objects are not properly recognized when the camera framerate is high. In [11], a regional multi-person pose estimation (RMPE) framework method is proposed for estimating multiple human objects. Although RMPE more accurately identifies the existence of multiple objects, it has difficulty estimating detailed postures. In [12], a method is proposed for generating a 3D human object skeleton to estimate the motion of human objects in sporting events. The method focuses on image processing using pipeline structures, rather than machine learning.

In [13-17], major studies on OpenPose based skeleton models are described. In [13], an OpenPose based skeleton model is used to determine the key point, i.e., the center of movement in the human body, and the characteristics of the falling behavior are investigated. This is a study on the detection of unusual human behaviors. In [14], the authors pointed out the problem of using markers as a method for estimating human motion and proposed a method for estimating the positions and postures in 3D without markers when applying OpenPose. In [15], also using OpenPose, a method is proposed for estimating a falling posture by detecting the movement of the center of the joint of a person in 2D in real time. The motion posture of a person is

determined based on 25 joint points. In [16], an OpenPose skeleton based basketball free-throw posture analysis model is proposed and the accuracy of free-throw posture prediction is analyzed. In [17], a method for predicting the 3D skeleton of a human hand using OpenPose is proposed, allowing the robot to recognize the hand movements of the human operator. As described above, there are various methods and models for estimating the posture of human objects. Based on the OpenPose based skeleton model [6], recently developed at Carnegie Mellon University, we focused on the development of a smoking behavior prediction model.

Table 1 Comparison of existing studies on motion recognition

Studies	Characteristic
[6]	Introduces filter learning that can encode not only the location of body parts but also the connection relationships between body parts, and encodes expressions (part affinity fields) including location and direction information into 2D vectors.
[7]	Devised a method to conduct calculations more efficiently by changing the structure of Deepcut; however, several to several hundred seconds may be required depending on the image.
[8]	A technique is presented for separating the hands by mixing color and depth images, and by learning the skin model in real time, the skin model is adaptively updated for the lighting environment, thereby resulting in a stronger recognition performance.
[9]	To solve the difficulty of taking complex poses with the existing graphic model, a method of directly learning the inference process is used over a graphical pose estimation method.
[10]	Based on the PAF(Part Affinity Fields) method, the TAF(Temporal Affinity Fields) method is applied in the existing online tracking approach. PAF shows only those parts constituting the skeleton while connecting key points, whereas TAF strengthens the connectivity by connecting the surrounding key points, similar to a graph.
[11]	A regional multi-person pose estimation (RMPE) framework is proposed as a method of estimating multiple objects of a person. Although it achieves a more accurate identification of multiple objects, there is a lack of detail in their posture.
[12]	A method for generating a 3D human object skeleton in estimating the motion of human objects in sports events is proposed. It focuses on image processing using pipelines rather than learning.
[13]	A study is conducted on determining the key points in the human body and investigating the characteristics of falling behavior using an OpenPose based skeleton model.
[16]	An OpenPose skeleton based basketball free throw posture analysis model is proposed and the accuracy of the free throw posture prediction is analyzed.

2.2 OpenPose based Skeleton Model

OpenPose [6] is a method of human pose estimation that predicts the body, face, and knuckles of an individual using only a single camera. Early research began with a method published by Carnegie Mellon University in 2017. Based on a deep learning convolutional neural network, it estimates the feature points of the body, hand, and face from only images and photographs. With a traditional method, it is necessary to attach related devices to the human body, which is expensive and limited in terms of space, and achieves varying results depending on the body type of the person. With the development of knowledge and computing performance, deep learning technology has been used in the field of computer vision [21]. By combining these technologies, it has become possible to predict human postures using only images. Fig. 1 show the human posture prediction process based on OpenPose.

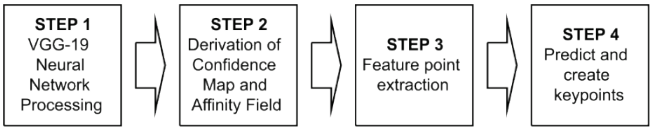


Figure 1 Prediction process of human posture based on OpenPose

The human posture prediction process based on OpenPose is as follows: The first stage is the VGG-19 [18] neural network processing stage. In this step, a batch-type input image is input into the VGG-19 neural network. After the input image passes through the neural network, it becomes image data with a feature size of 28 × 28 × 512. These data become the input data for step 2. In step 2, a confidence map and affinity field are derived. A confidence map is used to determine the joint position of a person in an image, and the preference field is used to determine who owns the joint extracted from the image. Step 3 involves extracting the feature points. At the beginning of step 3, although meaningless features are created according to the input image, these features are compared with the posture of the individual, optimization is performed, and the features gradually converge in the direction pointing toward the joint position of the person. Subsequently, they continuously pass through the next branch and predict a 2D preference field based on the similarity between each part of the human body. In Step 4, a keypoint is predicted. In other words, keypoints of individuals in the image are predicted and generated using the confidence map and preference field calculated in the previous step. When each joint from the reliability map is combined, the greedy algorithm (greedy relaxation) determines the owner of each joint if there are several people in the image. By repeating this process, it becomes possible to predict the human posture.

3 DESIGN OF PROPOSED SYSTEM

3.1 Image Data Processing Procedure Used by the Proposed System

The proposed system uses the OpenPose based skeleton algorithm to identify and analyze the smoking behavior of an object from motion images captured by a CCTV camera. Fig. 2 shows the image data processing procedure used by the proposed system.

The image processing procedure of the proposed system is as follows: First, the image data to be used for learning are preprocessed to facilitate data processing, and the preprocessed smoking image is input into the OpenPose skeleton model. The images processed by the skeleton model undergo a learning process. As a result, a smoking behavior recognition model is created. CCTV footage is input into the generated smoking behavior recognition model to determine whether smoking is taking place. To increase the accuracy of the smoking behavior recognition, the result of the smoking behavior recognition model is transmitted to the hardware device and combined with the smoking behavior result of the hardware device, allowing the smoking behavior to be

determined. The hardware device developed in this study can be used to increase the accuracy of the software when recognizing smoking behavior. The device might mistake the presence of residual gas as the occurrence of smoking and is therefore used only as an auxiliary tool.

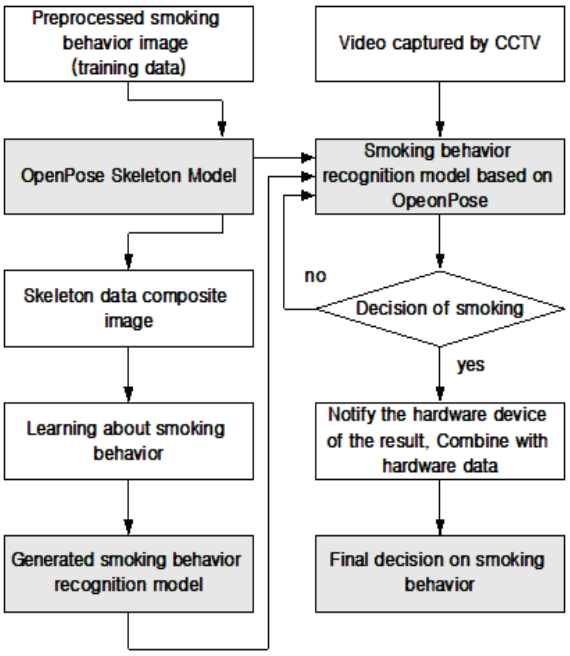


Figure 2 Image data processing procedure used by the proposed system

3.2 Image Data Preprocessing Procedure for Machine Learning

With the proposed system, preprocessing of the learning image data is required to create a smoking behavior recognition model. Fig. 3 shows the preprocessing of the training data. In the training image, the longer horizontal and vertical lengths are adjusted to a standard of 500 mm, and the missing part is filled in with black. The middle image in Fig. 3 illustrates this phenomenon. An OpenPose based skeleton model is applied to this image to create a dataset for learning the smoking behavior. The image on the right side of Fig. 3 shows this phenomenon. Using this image, an artificial intelligence neural network learns the smoking behavior for model creation. This model is combined with the skeleton model to input CCTV images in real time and detect whether a human object is smoking. If the individual is determined to be smoking, the result is sent to the hardware device.

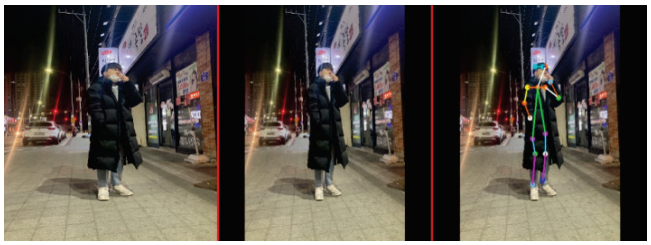


Figure 3 Preprocessing of the training data

3.3 Hardware Development to Increase the Accuracy of Smoking Behavior Recognition

A hardware device is used as an auxiliary tool to increase the accuracy of the smoking image recognition in the proposed system. Fig. 4 shows the systematic structure of the proposed hardware device. It consists of sensors that can detect a variety of smoke produced by cigarettes, a warning sound generation device, and a Raspberry Pi microcontroller. Fig. 5 shows the operational procedure of the hardware device. First, the CCTV footage is input into the above-mentioned smoking behavior recognition OpenPose based skeleton model to determine whether an individual is smoking. If it is determined that smoking behavior is taking place, the information is transmitted to a Raspberry Pi, which controls the hardware sensor. Second, when data are received from the Raspberry Pi, the sensors detecting gases and flames generated from smoking respond. Third, each sensor collects information on the smoking behavior and cigarette smoke. Fourth, when smoking occurs, the sensor values for collecting information on the presence of gases increase, and when the increased sensor values exceed a certain reference value, they are recognized as cigarette smoke. As a final step, when all sensors detect cigarette smoke, it is finally determined whether smoking is occurring, and a warning sound indicating a non-smoking area is output. By calculating an average value through repeated experiments, a standard value for cigarette smoke recognition is achieved.

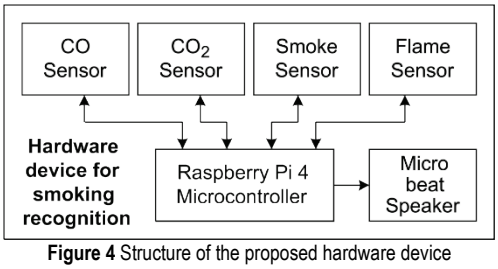


Figure 4 Structure of the proposed hardware device

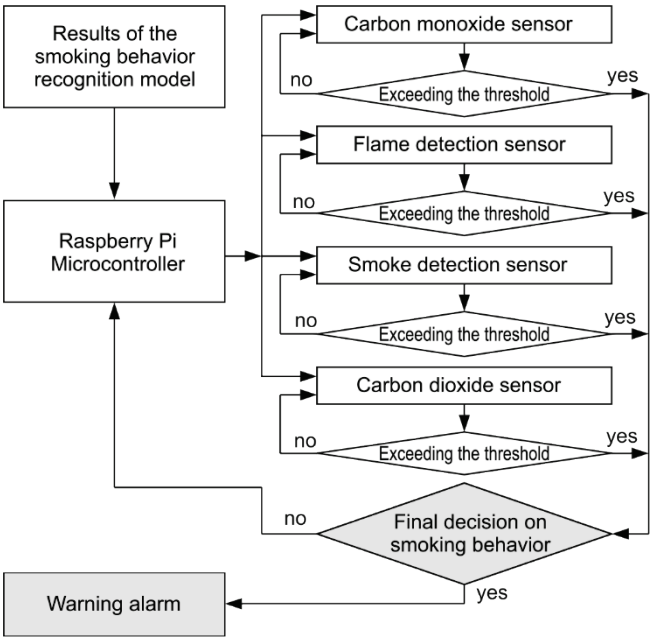


Figure 5 Operating procedure of hardware device

```
aug = ImageDataGenerator(
    rotation_range = 20,
    zoom_range = 0.5
    width_shift_range = 0.2,
    height_shit_Range = 0.2,
    shear_range = 0.15,
    horizontal_flip = True,
    fil_mode = "nearest")

baseModel = MobileNetVe(weight = "imagenet",
    include_top = false, input_tensor = Input(shape=(224, 225, 3)))

headModel = baseModel.output
headModel = AveragePooing2D(pool_size=(7, 7))(headModel)
headModel = Flatten(name = "flatten")(headModel)
headModel = Dense(128, activation = "relu")(headModel)
headModel = Dropout(0.5)(headModel)
headModel = Dense(2, activation = "softmax")(headModel)
model = Model(inputs = baseModel.input, outputs = headModel)

for layer in baseModel.layer:
    layer.trainable = False
opt = Adam(lr=INIT_LR, decay = INIT_LT / EPOCHS)
model.compile(loss="binary_crossentropy", optimizer=opt,
    metics=["accuracy"])

H=model.fit(aug.flow(trainX, trainY, batch_size = BS),
    steps_per_epoch = len(trainX) // BS,
    validation_data = (testX, testY),
    validation_steps = len(testX) // BS,
    epochs = EPOCHS)
```

Figure 6 Neural network algorithm for smoking behavior recognition

4 IMPLEMENTATION OF PROPOSED SYSTEM

4.1 Creation of Smoking Recognition Neural Network

In this paper, we describe the process of generating a smoking recognition neural network using the proposed system. Fig. 6 shows the main parts of the proposed smoking recognition neural network algorithm. The implemented smoking recognition artificial intelligence neural network uses the TensorFlow module (ImageDataGenerator) [19], by which the number of training data is increased through a slight transformation when learning the images. To create a large number of smoking data, an image is created and used by rotating, zooming, and moving based on the original image. The detailed parameters of this module are presented in Tab. 2. MobileNetV2 is used as the base model, and is a neural network light enough to be applied in a mobile environment. Because of the small number of parameters applied, the burden of manipulating a neural network can be reduced and the computational cost can be decreased.

Table 2 Image generator parameters

Parameters	Contents
Rotation_range	Random rotation angle within 20 degrees
Zoom_range	15% random zoom range
Width_shift_range	15% left and right movements
Height_shift_range	15% vertical movement
Shear_range	15% floor push strength
Horizontal_flip	Randomly flip horizontally
Fill_mode	When there is a blank space in the image from a reduction in rotation.

Average pooling is applied on the spatial data using the AveragePooling2D module to construct a neural network. With this module, the parameter (7, 7) indicates that the image is scaled down to 1/7 for a two-dimensional space, and thus an image size of 14 × 14 × 3 is created. The generated image is input into the Relu activation function, negative values are discarded, and positive values remain. Subsequently, it is normalized using the Dropout function to solve the overfitting problem, which is a phenomenon in which a large number of data are learned for only a specific model, and the accuracy of the analysis is significantly lowered for the untrained data. Finally, using the activation function Softmax, the input value is normalized to a value between zero and 1 as an output, and the sum of the output values is always 1. Softmax can be used for binary classification by allowing only the largest value of the result of this function to have a true value and the rest to have false values. Because a compilation is a binary classification of smoking and non-smoking, the model was constructed using the typical loss function NBinary_crossentropy.

The main software packages used for the smoking recognition model are listed in Tab. 3. Python 3.8 is used, and the latest versions of other artificial intelligence libraries such as TensorFlow and Keras were applied.

Table 3 Open software package versions

Libraries	Version	Libraries	Version
Python	3.8	Matplotlib	3.4.3
TensorFlow	2.5.0	TensorFlow-estimator	2.5.0
Keras	2.4.3	Keras-Preprocessing	1.1.2
scikit-learn	0.24.2	scicy	1.6.2
pandas	1.3.3	pip	21.2.2
OpenCv2	4.0.1	TensorBoard	2.5.0

4.2 Training Process of Dataset for Smoking Recognition Model

The training process of the dataset used for creating the smoking recognition model is as follows: First, a dataset was prepared. In this step, smoking behavior images were collected for data mining using Kaggle. Initially, 3,000 images, including 1,500 smoking and 1,500 non-smoking images, were prepared. Second, an image-resizing step was applied. The sizes of the 3,000 dataset images prepared were all different, and we therefore resized them to 500 × 500 × 3 for image learning. Resizing was conducted based on the longer horizontal and vertical lengths, and bilinear interpolation with high efficiency was used. The third step is image padding. When applying the second step, because resizing was applied based on the longer distance between the horizontal and vertical lengths, space occurred on one side. Empty spaces were filled with black. Resizing not only allows the images to be uniform, it also increases the recognition rate of the skeleton model. Fig. 7 shows an example of a resized image. When applying a skeleton to the original image, the arm part of the original data in the left image is not normally covered with the skeleton model; however, in the image on the right, it can be seen that the arm is normally covered with a skeleton model. Fourth, this stage involves the creation of a skeleton neural network. A skeleton neural network model was created by applying the skeleton model to the padded image in the previous step. It took approximately 15 h to process approximately 3,000 sheets.

The fifth stage is the data training stage. Machine learning was conducted based on the image data created in the previous step.

(a) (b)

Figure 7 Resized data image (a) (b)

The following four test dataset models were created based on the neural network created in the previous step.

For test model 1, 1500 smoking images and 1500 non-smoking images to which the skeleton model was not applied were trained.

For test model 2, 1500 smoking images and 1500 non-smoking images to which the skeleton model was applied were trained.

For test model 3, 1,000 smoking images and 1,000 non-smoking images to which the skeleton model was applied were added, for a total training of 5,000 images.

For test model 4, 1,100 smoking images and 1,400 non-smoking images, i.e., 2,500 images, were trained by deleting garbage data to which the skeleton model was not properly applied.

A final test model was selected through a performance analysis of the generated models, and the selected model was used as the final model for smoking behavior recognition.

4.3 Hardware Device Implementation

For the proposed system, a smoking behavior detection device for determining tobacco smoke was developed as an auxiliary tool to increase the accuracy of detection. The system was developed using the sensors shown in Figure 8, and is located at the top of the surveillance camera placed in a non-smoking area, as shown in Fig. 9. A Raspberry Pi 4 is applied as the hardware device controller for smoke gas recognition and data transmission, and MQ7 and MH-Z14A sensors are used to measure CO and CO_2 in the air. In addition, a GSAS61-P110 smoke sensor is used to detect cigarette smoke and an NS-FDSM flame sensor is used to detect the flame used to light the cigarettes as well as cigarette cinder.

A socket program that can be accessed from an external IP is used to transmit the resulting software value to the Raspberry Pi. A PC equipped with a smoking-aware skeleton model is used as the client, a Raspberry Pi is placed as the server, and data transmission is smoothly applied during socket programming. For the transmitted data, a value of 1 is used for smoking, and a value of 0 is applied for non-smoking. When a value of 1 is received, the Raspberry Pi device starts measuring the sensor value. Hardware sensors collect information on smoking, and when the collected sensor values exceed a set threshold value, smoking behavior is determined. Finally, the system recognizes that smoking has occurred and outputs a warning sound.

Figure 8 Sensors and hardware devices

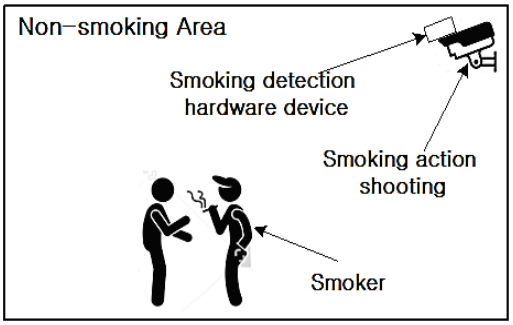


Figure 9 Smoking environment for experimentation

Table 4 Determination of sensor threshold values

Sensor	Threshold	Initial value	End value
CO	600	470	736
CO_2	1170	1182	1164
SMOKE	600	105	688
Flame	True	False	False

The threshold value of the sensor for cigarette smoke detection was determined through an environmental experiment on cigarette smoke using an acrylic box (40 × 40 × 40 cm). Tab. 4 shows the value of each gas measured for a period of approximately 60 s after the cigarette was lit (in order, the initial value when the sensor started measuring, the standard value indicating that smoking is occurring, and the value when the experiment was finished). The sensor threshold value is calculated by averaging the values collected from several environmental experiments. The CO, CO_2, SMOKE, and FIRE sensor threshold values were used

as the criteria for determining whether smoking is taking place.

4.4 Performance Analysis of Test Dataset Model

The most suitable dataset model for smoking recognition was selected through performance analysis experiments conducted on the four neural network test dataset models described in Section 4.2. The performance analysis proceeded as follows. First, a classification experiment was conducted for each model. This is an experiment on how well the classification (Classification_report) function was applied to the model (dataset). Tab. 5 shows the experimental results for models (dataset) 1 and 2, and Tab. 6 shows the experimental results for models 3 and 4.

Table 5 Performance analysis results for neural network models 1 and 2

#model 1					#model 2				
	Precision	Recall	F1-score	Support		Precision	Recall	F1-score	Support
Not-smoking	0.79	0.82	0.81	255	Not-smoking	0.82	0.83	0.83	255
Smoking	0.88	0.86	0.87	396	Smoking	0.89	0.89	0.89	396
Accuracy			0.85	651	Accuracy			0.86	651
Macro avg.	0.84	0.84	0.84	651	Macro avg.	0.86	0.86	0.86	651
Weighted avg.	0.85	0.85	0.85	651	Weighted avg.	0.87	0.86	0.86	651

Table 6 Performance analysis results for neural network models 1 and 2

#model 1					#model 2				
	Precision	Recall	F1-score	Support		Precision	Recall	F1-score	Support
Not-smoking	0.83	0.84	0.84	439	Not-smoking	0.82	0.85	0.83	227
Smoking	0.88	0.87	0.87	571	Smoking	0.87	0.85	0.86	281
Accuracy			0.86	1010	Accuracy			0.85	508
Macro avg.	0.85	0.85	0.85	1010	Macro avg.	0.85	0.85	0.85	508
Weighted avg.	0.86	0.86	0.86	1010	Weighted avg.	0.85	0.85	0.85	508

The experimental results are as follows: First, precision is the ratio of the number of samples that actually belong to the positive class to the number of samples that are output as belonging to this class. The higher the precision is, the better the judgment of the model. A precision of 1.0 means that there are zero false positives (FPs). Tab. 7 shows a classifier matrix of the precision.

Table 7 Classifier matrix

		Real result	
		True	736
Classification result	True	Classification result	True
SMOKE	False	105	False

The meaning of each item is as follows, and the precision can be obtained using Eq. (1):

- True positive (*TP*): The model accurately predicts that the sample is higher than the threshold value.
- True Negative (*TN*): The model accurately predicts that the sample is lower than the threshold value.
- False positive (*FP*): The model inaccurately predicts that the sample is higher than the threshold value.
- False negative (*FN*): The model inaccurately predicts that the sample is lower than the threshold value.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Second, recall is the ratio of the number of samples that are output as belonging to the positive class to the number of samples belonging to that class. The higher the number is, the better the model. A recall of 1.0 means that the *FN* is zero. The recall can be calculated using Eq. (2).

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

Third, the F-score is a weighted harmonic average of the precision and recall. In particular, beta is the weight assigned to the precision, and when the value is 1, it is called the F1 score. The higher the number is, the better the model. The F1 score can be calculated through Eq. (3).

$$PF_{\beta} = \frac{(1+\beta^2)(Precision \times Recall)}{(\beta^2)(Precision \times Recall)} \tag{3}$$

Fourth, accuracy refers to the ratio of the number of correctly predicted samples to all samples. The higher the number is, the better the model. In general, it is used as an optimization objective function in learning. The accuracy can be calculated through Eq. (4).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Fifth, the macro-average is the weight assigned to each class. The same weight is assigned to each class. In other words, an imbalance in the number of samples is not considered. Because it does not consider the imbalance in the number of samples, a larger penalty occurs when the performance of a minority class is low.

Finally, the weighted average is based on the number of samples belonging to each class. An imbalance in the number of samples is considered. Because the weighted average is applied, the influence of a class with a small number of samples is reduced.

The accuracy of the results for each dataset model were 0.85, 0.86, 0.86, and 0.85, respectively, indicating no significant differences. In addition, there were no significant differences in the precision, recall, and F1 scores. However, it can be determined that relatively uniform data can be output regardless of the input data because the difference in value between the non-smoking data and the smoking data for dataset model 4 is smaller than that of the other models. Therefore, we proceeded with the final smoking recognition experiment using dataset model 4.

To judge the smoking behavior, the image to which the skeleton model is applied must pass through a smoking recognition neural network. To input the image into the neural network, we need to resize it to 244 × 244 × 3, and in this case, interpolation is used. Interpolation is a method for estimating unknown values using known data. When changing the size of an image, if the ratio of the image is changed, new values must be assigned by mapping new pixel values to non-existent areas or by compressing the existing pixels. When an image is enlarged, an interpolation method for the pixels is applied, and when an image is reduced, a pixel merging method is used. There are five types of nearest-neighbor interpolation provided by OpenCV: neighbor, bilinear, bicubic, domain, and Lanczos interpolation [20].

Among these five interpolation methods, the nearest neighbor interpolation method was found to be the most suitable through experimentation. By applying each interpolation method to the test data consisting of 100 images, the proposed smoking recognition module is executed, and the interpolation method with the most correct answers is selected as the optimal interpolation approach. Finally, the selected interpolation method is applied to the image interpolation.

4.5 Smoking Recognition Experiment

Experiments were conducted to analyze the performance of the proposed neural network model. Dataset model 4, described in Section 4.4, was applied during the experiment, and the nearest neighbor interpolation method was used as the image interpolation method. As the experimental results. A total of 198 images of smokers and 198 images of non-smokers were applied during the experiment. The correct answer rate was 0.75 for the smoking images (148 correct answers) and 0.7 for the non-smoking images (139 correct answers).

Fig. 10 shows example results when applying the neural network smoking behavior recognition module proposed in this study. The object number appears on the head of the person who is smoking. When the smoking behavior recognition module identifies a human object, a skeleton model is applied, and for each object to which the skeleton model is used, it is determined whether a smoking action has occurred. As shown in the figure, when the system judges that both the left and right objects are smoking, it outputs the messages "AI has detected that the Nth object is smoking" and "Send a value of 1 (smoking behavior) to the manager" in the output window below. The recognition results are then transmitted to the hardware device. Tab. 8 shows the result of the recognition module being transmitted to the hardware device, where the final result is determined according to the sensor value. When a smoking action is determined, the message "The standard value has been exceed and output a warning sound" is displayed, as shown in the red box.

Figure 10 Results of smoking behavior recognition module

Table 8 Final results of the recognition module and hardware device

	Output Example
Algorithm execution and output example	CO: 580, CO_2: 1173, Smoke: 811, Fire: False CO: 610, CO_2: 1173, Smoke: 837, Fire: False The standard value has been exceed and output a warning sound. CO: 625, CO_2: 1173, Smoke: 840, Fire: False The standard value has been exceed and output a warning sound. CO: 637, CO_2: 1173, Smoke: 833, Fire: False The standard value has been exceed and output a warning sound. …

5 CONCLUSIONS

In this study, an artificial neural network based smoking behavior recognition model using an OpenPose based skeleton module for detecting smoking in a non-smoking area and providing a warning was proposed, along with a hardware device for improving the accuracy of smoking behavior recognition. To improve the recognition performance, the best dataset model among the four learning dataset models was selected experimentally, and the best performing method among the five interpolation methods was applied to generate the final recognition model. Based on the performance analysis of the proposed model, the smoking behavior recognition rate was 70% for *TP*s and 75% for *FN*s. The hardware device combines the transmitted results with the results recognized by the sensor and generates an accurate warning sound when the result of the smoking behavior recognition module is transmitted to the hardware device.

Compared with related studies [9], the image recognition rate of the proposed system is similar to the average recognition rate of 72% for each part of the body; however, this study is significant and differs in the following ways.

First, the subject of this study was the intricate posture of smoking, rather than easy-to-judge postures, such as walking and falling, addressed in other studies. This makes it difficult to judge only the numerical results of the motion recognition accuracy shown in each study. Second, several characteristic tasks were performed to increase the smoking behavior recognition rate. A preprocessing process was first performed for the convenience of processing learning data. Four dataset models were established, and the best training dataset was determined through experiments. Finally, various interpolation methods were tested, and the test data were processed using the nearest neighbor interpolation method to increase the recognition probability of the test data. These tasks are characteristic image recognition procedures of this study and are factors that can improve the performance of the proposed smoking behavior recognition module in the future. Third, a hardware device was developed to verify and supplement the accuracy of the smoking motion recognition software module. That is, a hardware device was developed and applied as an auxiliary tool to increase the accuracy of smoking behavior recognition. That is, the hardware device is notified of the result of the software recognition model, and thus the final smoking behavior is recognized and judged. In the final performance experiment fusing the software recognition module and hardware device, when the software result is transferred to the hardware device, the hardware device properly senses and reacts to the gas generated from smoking.

Despite the significance of this study, additional studies are required in the future. First, it is necessary to quantitatively expand the training image data for machine learning as well as select and train an image suitable for the OpenPose algorithm. In this study, 1500 smoking and 1500 non-smoking data points were used as training data. It is necessary to learn more data and select and apply various posture data to increase the recognition rate. Second, a better dataset model is needed to improve smoking recognition accuracy. In this study, the best-performing dataset model was selected based on precision and accuracy, but it is necessary to develop more diverse dataset models and use the best-performing dataset model. Finally, the hardware device that recognizes smoking motion needs to be improved. In other words, it is essential to add a sensor that can detect several gases produced by cigarette smoke, and it is equally important to select a sensor that can identify the same gas with excellent detection performance.

Acknowledgements

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (No. 2021R1I1A3044091).

6 REFERENCES

[1] Lee, G., Kim, J., & Jang, Y. (2020). A study on types of smoking location and space through analysis of smokers' behaviors. *Journal of the Korea Institute of the Spatial Design, 15*(8), 417-428. https://doi.org/10.35216/kisd.2020.15.8.417

[2] Kim, N. (2017). Change of University students' smoking behavior before and after non-smoking policy. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, 7*(6), 419-426. https://doi.org/10.35873/ajmahs.2017.7.6.038

[3] Noh, J., Lee, Y., Yoo, K., & Yoon, J. (2018). Overseas trend of non-smoking area guidelines and outdoor smoking room. *Korean Public Health Research, 44*(3), 53-64. https://doi.org/10.22900/kphr.2018.44.3.005

[4] Jee, Y. & Kim, Y. (2021). Study on the factor influencing smoking cessation among adolescent smokers. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 7*(8), 251-260. https://doi.org/10.47116/apjcri.2021.08.23

[5] Park, H. & Lee, C. (2021). Effects of anti-smoking advertising messages on smokers. *Asia-pacific Journal of Convergent Research Interchange, FuCoS, 7*(9), 131-140. https://doi.org/10.47116/apjcri.2021.09.12

[6] Cao, Z., Hidalgo, G., Simon, T., Wei, S., & Sheikh, Y. (2021). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172-186. https://doi.org/10.48550/arXiv.1812.08008

[7] Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P., & Schiele, B. (2016). DeepCut: Joint subset partition and labeling for multi person pose estimation. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1-9. https://doi.org/10.48550/arXiv.1511.06645

[8] Urooj, A. & Borji, A. (2018). Analysis of hand segmentation in the wild. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4710-4719. https://doi.org/10.48550/arXiv.1803.03317

[9] Ramakrishna, V., Munoz, D., Hebert, M., Bagnell, J., & Sheikh, Y. (2014). Pose machines: Articulated pose estimation via inference machines. *Proceeding of the European Conference on Computer Vision, Lecture Notes in Computer Science*, 8690, 33-47. https://doi.org/10.1007/978-3-319-10605-2_3

[10] Raaj, Y., Idrees, H., Hidalgo, G., & Sheikh, Y. (2019). Efficient online multi-person 2D pose tracking withrecurrent spatio-temporal affinity fields. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, 4620-4628. https://doi.org/10.48550/arXiv.1811.11975

[11] Fang, H., Xie, S., Tai, Y., & Lu, C. (2017). RMPE: Regional multi-person pose estimation. *Proceeding of the International Conference on Computer Vision ICCV 2017*, 2334-2343. https://doi.org/10.1109/ICCV.2017.256

[12] Bridgeman, L., Volino, M., Guillemaut, J., & Hilton, A. (2017). Multi-person 3D pose estimation and tracking in sports. *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1-10. https://doi.org/10.1109/CVPRW.2019.00304

[13] Chen, W., Jiang, Z., Guo, H., & Ni, X. (2020). Fall detection based on key points of human-skeleton using OpenPose. *Symmetry, 12*(5), 1-17. https://doi.org/10.3390/sym12050744

[14] Slembrouck, M., Luong, H., Gerlo, J., Schutte, K., Cauwelaert, D., Clerdq, D., Vanwanseele, B., Veelaert, P., & Philips, V. (2020). Multiview 3D markerless human pose estimation from OpenPose skeletonsm. Proceeding of the International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS) 2020, Lecture Notes in Computer Science, 12002, 1-12. https://doi.org/10.1007/978-3-030-40605-9_15

[15] Lin, C., Dong, Z., Kuan, W., & Huang, Y. (2021). A framework for fall detection based on OpenPose skeleton and LSTM/GRU models. *Applied Sciences, 11*(1), 1-20. https://doi.org/10.3390/app11010329

[16] Nakai, M., Tsunoda, Y., Hayashi, H., & Murakoshi, H. (2018). Prediction of basketball free throw shooting by OpenPose. *Proceeding of the International Symposium on Artificial Intelligence (JSAI-isAI) 2018: New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science*, 11717, 435-446. https://doi.org/10.1007/978-3-030-31605-1_31

[17] Mazhar, O., Ramdani, S., Navarro, B., Passama, R., & Cherubini, A. (2018). Towards real-time physical human-robot interaction using skeleton information and hand gestures. *Proceeding of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 1-7. https://doi.org/10.1109/IROS.2018.8594385

[18] Simonyan, K. & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. *Proceeding of the 3rd International Conference on Learning Representations, ICLR 2015*, 1-14. https://doi.org/10.48550/arXiv.1409.1556

[19] Urakawa, T., Tanaka, Y., Goto, S., Matsuzawa, H., Watanabe, K., & Endo, N. (2019). Detecting intertrochanteric hip fractures with orthopedist-level accuracy using a deep convolutional neural network. Skeletal Radiology, 48(1), 239-244. https://doi.org/10.1007/s00256-018-3016-3

[20] Jose, S., Miguel, C., Rafael, A., & Marta, S. (2020). OpenCV basics: A mobile application to support the teaching of computer vision concepts. *IEEE Transactions on Education, 63*(4), 328-335. https://doi.org/10.1109/TE.2020.2993013

[21] Nesreen Samer E.-J., & Samy S. A.-N. (2018), Diabetes Prediction Using Artificial Neural Network. *International Journal of Advanced Scienc+e and Technology, NADIA, 121*, 55-64. https://doi.org/10.14257/ijast.2018.121.05

Authors' contacts:

Tae-Yeong Jeong, Graduate
School of Computer Science,
Kyungil University,
50, Gamasil-gil, Hayang-eup, Gyeongsan-si,
Gyeongbuk, 38428, South Korea
taeyeong.jeong419@gmail.com

Il-Kyu Ha, Professor
(Corresponding author)
School of Computer Science,
Kyungil University,
50, Gamasil-gil, Hayang-eup, Gyeongsan-si,
Gyeongbuk, 38428, South Korea
ikha@kiu.kr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230417140747

Preliminary communication
Received: 2023-04-17, Accepted: 2023-04-21

Challenges in Flexible Manufacturing Technologies for the Final Assembly in the Commercial Vehicle Industry

Tarik Demiral, Jürgen Bock*, Pierre Johansson

Abstract: Increasing customer demands and product diversity as well as emerging technologies and market trends, such as the establishment of new driveline technologies, like e-mobility or hydrogen, present challenges for manufacturing companies in the commercial vehicle industry. Consequently, companies must strengthen their focus to the aspect of flexibility within their manufacturing processes. This paper contributes to the state of the art in flexible manufacturing technologies research, which enables manufacturing companies to deal with these increasing flexibility requirements. Focussing on the area of final assembly the paper takes a holistic perspective and characterizes the readiness of automotive companies to be able to implement flexible manufacturing technologies. The system ecosystem and the process organization of automotive companies are examined with respect to the requirements of flexible manufacturing. Finally, gaps that hinder the implementation of flexible manufacturing technologies are identified and described, and possible solution concepts for the identified gaps are proposed.

Keywords: business capabilities; final assembly; flexibility; gap analysis; manufacturing

1 INTRODUCTION

The commercial vehicle industry is facing new challenges that are derived from increasing customer demands leading to higher complexity due to the diversity of variants. On the other hand, challenges arise from emerging technologies, such as electromobility and digitalization [1]. Particularly, these developments have a major effect on manufacturing, including manufacturing preparation and process engineering. One way how companies are dealing with these challenges is the introduction of new concepts in the area of flexible manufacturing. This has a great impact on state-of-the-art manufacturing and influences the digital infrastructure and organization of companies.

The challenges an organization is facing can be studied along the capabilities that are required in order to meet the requirements that are imposed by the novel concepts of flexible manufacturing. These so-called business capabilities [2] can be identified for state-of-the-art manufacturing systems, as well as for novel flexible manufacturing systems. In addition to presenting the currently realized as well as the required system capabilities for future manufacturing, the main contribution of this paper is an analysis of the gaps that currently prevent the realization of required novel system capabilities to increase manufacturing flexibility.

The research for this paper was conducted following the scientific method of expert interviews on the one hand, and a case study on the other hand. Regarding the former, qualitative interviews with open-ended questions ensure openness and flexibility. The conducted type of interview is particularly suitable for uncovering unknown findings step by step. In total, seven experts were interviewed: three enterprise architects, two solution architects from two different final assembly plants, as well as two manufacturing engineers. By conducting the interviews with different roles and areas of expertise, it was possible to cover as broad an area of the research topic as possible and to reflect the issue with relevant results from different perspectives. Regarding the case study as a second scientific method, a realistic future scenario is described as a case that is derived from thorough research and study of the domain. It includes a sufficient level of information and detail that can be analysed an interpreted from different perspectives. A case study is a suitable method since the manufacturing system investigated in this paper is not technically implemented yet in any organization, and thus has no real-life environment that can be used for observations and thus for field research.

In the following, Sect. 2 describes the state of the art regarding the conventional truck assembly line. Moreover, the business capability map is introduced for manufacturing operations related business capabilities currently realized. Section 3 presents the elements of a future manufacturing system including its required system capabilities. Section 4 introduces the gaps that were identified and relates them to the required system capabilities. Section 5 evaluates the presented findings, before Sect. 6 concludes the paper with a summary and an outlook on future work. The paper is based on a more extensive scientific work conducted through the application of the mentioned scientific methods at a leading OEM of commercial vehicles [3].

2 STATE OF THE ART

The contribution of this paper is based on the state of the art in the final assembly in the commercial vehicle industry and the business capabilities currently characterizing such assembly systems.

2.1 Final Assembly in Commercial Vehicle Production

The final assembly in the European automotive industry is characterized by customized products and different vehicle-variations on one assembly line. High variation requires a high degree of flexibility, which cannot be provided by automated solutions, thus, manual or only partially mechanised assembly operations are applied,

despite high labour costs [4]. Heavy-duty truck manufacturing, as a sub-section of the automotive industry, has even lower volumes and higher customization levels [5].

Final assembly lines are dominated by linearly arranged workstations that are rigidly connected with each other using conveyor technology [6, 7]. A workstation can be described by the assembly operations conducted in it, the number of operators, the available equipment and the materials delivered to the station [7]. The process of allocating assembly operations to the manufacturing resources is called line balancing [8]. Each workstation is designed to conduct a number of specific assembly operations, according to which the required number of operators for each station will be derived. An assembly line can be further characterized by a tact time per station, where a station's total assembly time cannot exceed the tact time [6, 7]. The difference between the total assembly time per station and the tact time is unproductive "idle time". Assembly line balancing has the overall goal to evenly distribute all assembly activities throughout the workstations and minimize idle time [9]. Apart from the tact time, further technical and organisational restrictions, as well as equipment availability and the level of qualification of operators need to be considered [10].

Assembly operations are generally planned on standardized assembly sequences. As a result, operators will complete assembly tasks in a fixed sequence, which needs to be revised and updated by manufacturing engineers regularly due to product changes and changing manufacturing parameters. To this end, not all vehicle configurations require the same order of operations, e.g., a vehicle without a glass roof still needs to pass the corresponding station without any operations conducted here [6].

Regarding material handling, which involves the storage, control, and delivery of production parts for the final assembly, Just-In-Time (JIT) and Just-In-Sequence (JIS) are heavily utilized concepts in the automotive industry. The JIT approach ensures that only the required items in the right quantity are supplied into the production system when needed. The goal is to reduce inventories, and eliminate waste and inconsistencies while increasing productivity. The JIS approach goes further and aims to ensure that the material is delivered in the order in which it will be processed later. For JIT and JIS to be effective, the company must have a very smooth operational system for material handling in place, as any disruptions in the material supply chain can cause a significant impact on the entire system [10].

2.2 The Business Capability Map

Business capabilities refer to the abilities possessed by an organization, individual, or system to perform a fundamental business function. A business capability map provides a structured representation of all business capabilities within the organization to coordinate and align a variety of distinct business capabilities [2]. The business capability map provides a clear definition for each business capability, which is further broken down into IT systems that assist the organization in achieving these capabilities. This section will outline the current business capability map for manufacturing-related business capabilities (see Fig. 1).

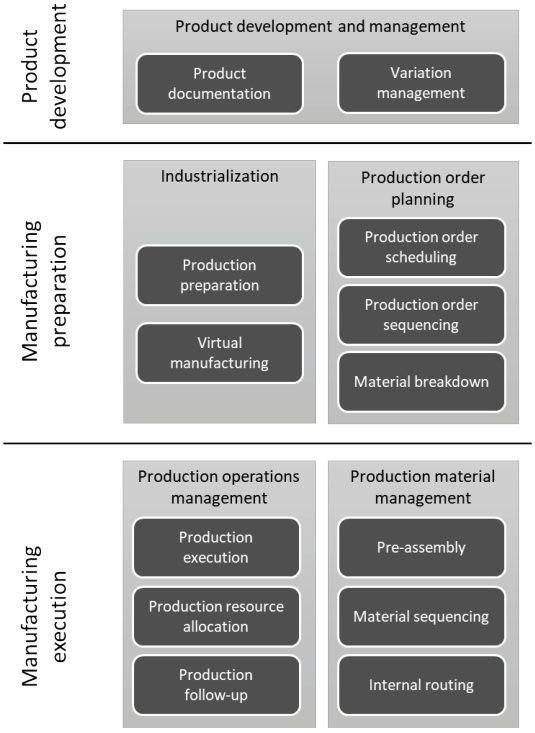


Figure 1 Current business capability map

Product development and management. The capability of developing new products and improvement is a crucial requirement for competitiveness in high-tech industries such as the automotive industry. Product development involves all activities related to the design of the product and its properties, including geometric design of individual components and the product as a whole. The main output of the product development process is the intellectual product, containing the product description along with all associated documentation, specifications, and digital models [11]. *Product development and management* heavily influences manufacturing processes.

Industrialization. After product development, the next phase is referred to as *industrialization* or production planning. Industrialization acts as the interface between product development and manufacturing. It involves planning production processes, providing necessary manufacturing resources, and conducting production tests to verify manufacturing feasibility and optimization [12]. The focus of *industrialization* is on the operational aspects of designing and implementing the manufacturing solution within the production environment. The business capability *industrialization* describes an organization's ability to develop, configure, prepare, verify, and implement the manufacturing solution for developed products.

Production order planning. Efficient processing of customer orders requires the conversion of orders into production orders and planning in terms of time and capacity with available resources. There are several core processes within *production order planning.* Production order

scheduling involves converting customer orders into production orders and assigning them to different plants while respecting production capacity constraints, sales forecasts, order prioritization and the model-mix within the plant [13]. Production order sequencing converts scheduled production orders into a detailed production program for a certain period of time. Material requirements planning is relevant for material planning aspects. Overall *production order planning* describes an organization's ability to plan and fulfil customer orders within their manufacturing system.

Production operations management. Also known as shopfloor management, *production operations management* involves coordinating and managing all functions directly within the operational manufacturing execution area. The primary goal is to control production processes efficiently and without any disruptions. However, being able to generate appropriate process deviations to keep the assembly flow running smoothly in case of any disruptions is an important part of *production operations management* area [14]. This business capability describes the organization's ability to efficiently steer manufacturing processes on the execution level as well as collect and use real-time data from the shopfloor for optimization and reporting purposes.

Production material management. The supply of materials to workstations is a core task of the final assembly. Through the implementation of JIT and JIS concepts, materials need to be delivered in the right sequence with respect to its further processing at their point of use [10]. In this context, intralogistics covers the planning, control, implementation, and optimization of the internal flow of goods and materials, including the required information flow. Immediate intralogistics processes within final assembly include pre-assembly and kitting operations, material sequencing, as well as internal routing of materials. This business capability describes the organization's ability to effectively supply the right parts and sub-assemblies in the right quantity, at the right time and the right point of use.

3 FUTURE FLEXIBLE MANUFACTURING SYSTEM

Flexibility in manufacturing refers to the system's ability to adapt to changing market demands and quantities, as well as the required effort for installations and process adjustments [15]. Because of shorter product life cycles, smaller lot sizes as well as increasing product variety and complexity, concepts, which increase manufacturing flexibility gain in importance, are presented in the following. A possible future flexible assembly line is presented, which is used to derive required system capabilities.

3.1 Future Manufacturing Concepts

Flexibility in future manufacturing systems is characterized in the following dimensions [16]. ***Operation flexibility*** denotes the ability of a part or a manufacturing process to be executed in different ways or sequences. ***Process flexibility*** describes the ability of a workstation to quickly produce a variety of products without any major reconfigurations. ***Routing flexibility*** denotes the ability to produce a part by using alternate routes through the system.

Flexible material delivery. Automotive assembly lines are characterized by standardized and rigid assembly sequences, resulting in components always being delivered to the same workstation for their processing [17]. By combining operational flexibility for parts and processes, process flexibility for workstations, and routing flexibility for the material handling system, the concept of flexible material delivery is introduced. It allows workstations to execute different assembly operations through the delivery of different parts. The assembly sequence is not standardized and may vary between production orders. The allocation of assembly operations determines the material delivery points for the required parts. Within JIS-manufacturing, a higher degree of synchronization between process execution and the material handling system is required. Flexible material delivery results in more decision freedom in manufacturing preparation, however through increased levels of complexity in the execution area and for the operators are a consequence.

Plug and Produce. Besides flexible material delivery, the provisioning of production resources such as tooling and equipment need to be considered. The Plug and Produce concept, refers to a technology in which manufacturing devices can be utilized without requiring any modifications to the system or additional registrations to the database [18]. For instance, fastening tools in manual assembly can be utilized in multiple workstation for a variety of assembly operations, without the prior need for readjustments.

Flexible production sequencing. In flexible manufacturing environments, the production of a single final product can be achieved using various sequences of operations [19]. Instead of the rigid assembly line, a production program determines which products need to be processed in which workstations. Hence, the classification of the capabilities of manufacturing assets, such as workstations, equipment and operators is required. Furthermore, a link between product parts based on individual manufacturing related product requirements is essential. The goal is to generate an optimal sequence for operations for each individual production order while avoiding inefficiencies in the production process. However, the standardized assembly sequence and and uniform tact time for the entire assembly line need to be abandoned, resulting in individual assembly sequences, production routes, and tact times for each production order.

Collaborative robotics. Although automation is not common in final assembly systems and over 90% of tasks are still performed by humans, collaborative robots (cobots) are becoming increasingly relevant for the assembly area [20]. Cobots can help to reduce stress and ergonomics issues for human operators, allowing operators to focus on tasks that are more complex. As adaptiveness and situation awareness in robotics continue to improve and cost-effective solutions are established, the importance of cobots for the assembly sector will increase further. Cobots can be utilized flexibly for a variety of assembly operations. Hence, an important prerequisite for their use is their integration into the process engineering methodology.

3.2 The Future Flexible Assembly System

A new final assembly system combining the four flexible manufacturing concepts from Sect. 3.1 is designed with the goal to increase efficiency by enhancing the flexibility of the system while considering new driveline technologies, like e-mobility or hydrogen, in the commercial vehicle industry.

The proposed new concept utilizes flexible assembly work cells, capable of assembling different products in different batch sizes without the need for time-consuming reconfigurations to the work cells. The system incorporates the three principles of a flexible manufacturing system described by Askin et al. [21], specifically, routing flexibility, part volume flexibility and part mix flexibility to allow the system to adapt to changes in volume and type of product. However, since conventional assembly lines are designed for efficiency and can be optimized for speed, the general design of the assembly system was chosen to follow a mix of cellular manufacturing and characteristics of the conventional assembly line.

The presented assembly system replaces workstations with assembly zones and utilizes a mixture of cellular assembly and line manufacturing. High-level product classes are introduced based on manufacturing requirements. Consequently, product-class specific assembly zones are dedicated to specific variants, like conventional, electrical or hydrogen vehicles. Flexible production sequencing is used to ensure that each production order only passes through required assembly zones. To enable routing flexibility, AGVs are used for transportation. Flexibility within the individual assembly zones themselves is ensured through the concepts of flexible material delivery, Plug and Produce and cobots. Both material and equipment as well as cobots are guided by AGVs from central pools or warehouses to their point of use in the assembly zones. Since not every equipment can be moved flexibly between the assembly zones, specific assembly zones are introduced containing stationary equipment for specific operations. Moreover, the concept of a standardized tact time is replaced by dynamic tact times per individual assembly zone derived from the individual workload of production order configurations.

Based on this proposed hypothetical future flexible assembly system, a case study was created. The case study incorporates a realistic manufacturing scenario with different product variations, e.g., driveline technologies, assembly operations and multiple cycles. The goal was to create a feasible and realistic manufacturing scenario for the system, which would highlight the system's abilities in terms of flexibility and efficiency.

3.3 Required System Capabilities

Based on the four flexible manufacturing concepts and the introduced future flexible assembly system four required system capabilities were derived.

Autonomous decision-making. The dynamic and flexible allocation of assembly processes, resources, and production materials is too complex, costly and time consuming to be performed manually. For that reason, autonomous decision-making, with systems making independent manufacturing decisions is required.

Real-time data processing. Based on the dynamic process design, in which the system can adapt to short-term changes, it is fundamental to include real-time data capturing and processing. To enable real-time data processing, an efficient data infrastructure that allows for immediate access to the necessary data and information is necessary. Additionally, relying on manual processes can be too time-consuming, which is why it is necessary to implement a combination with autonomous decision-making.

Software management and maintenance. To implement a flexible assembly system, the development and implementation of new software systems is required. In this context, the essential aspect of the system ecosystem is scalability. It is essential to plan individual systems and the ecosystem with respect to system maintenance, expansion of functionalities and use cases in advance, in order to ensure that the system ecosystem remains flexible and adaptable.

Process methodology. The flexible assembly system requires the adaptation of the organization's way of working. This requires an adaptation of the planning methodology with regard to manufacturing preparation as well as for assembly processes in the execution area. It is important to train responsible individuals in the data-supported way of working of a new system architecture. Manufacturing-related product documentation, simulation-based verifications, and optimization must be integrated into the processes. Overall, the successful implementation of a flexible assembly system requires changes to both process and organization. A key factor is that the prevailing assembly structure is no longer a determining factor.

4 GAPS

Derived from the expert interviews and the case study seven major gaps were identified that represent considerable obstacles for companies in their transformation process towards flexible manufacturing systems. Fig. 2 illustrates the relation of the identified gaps to the required system capabilities as described in the following paragraphs.

Autonomous decision-making	Real-time data processing	Software management and maintenance	Process methodology
Data availability / accessibility	Data availability / accessibility	Data availability / accessibility	Data availability / accessibility
		Monolithic applications	
Manual input	Manual input		Manual input
System arrangement	System arrangement	System arrangement	
			Assembly structure
Use of CAD data			Use of CAD data
Variation and complexity			Variation and complexity

Figure 2 Identified gaps and their relation to required system capabilities

4.1 Data Availability and Accessibility

Data availability refers to the process of data acquisition and storage. It includes organizational decisions of which and how much data to collect. In addition, there are technical factors involved, e.g., data types and storage systems. Data accessibility refers to the provisioning of the relevant data for the respective stakeholders (people, systems, processes) at the time it is required. A major factor that influences data accessibility is the heterogeneity of data due to different sources, including technical issues such as the need for sufficient interfaces between different systems.

In the interviews, it was unveiled, that problems concerning data availability and accessibility become visible, e.g., in the assembly environment. Heterogeneous data sources providing production-relevant product data, and data and information about equipment are an issue. It should be pointed out that even if the data for the operating equipment is created, the information content was found to be insufficient for the decision-making process, e.g., regarding availability of certain tools and an analysis of the degree of utilization.

For ***autonomous decision-making*** the availability and accessibility of all data and information is essential. ***Real-time data processing*** requires direct accessibility of the the relevant data and information, which relies on an effective and efficient data infrastructure. Data accessibility has an important influence on ***software management and maintenance***, as well as ***process methodology***, since the responsible persons must be trained with respect to the data-supported way of working in a new system architecture.

4.2 Monolithic Applications

Monolithic applications are characterized by the fact that they are built as a single proprietary system that combines several business processes. Even though such monoliths can be internally organized in several layers, these layers are typically intertwined and no further decoupling or modularization is done, impeding other application to from fine-grained access to relevant data and functions. The agglomeration of functionality increases dependencies of other applications from different areas. Thus, monoliths have high responsibility for a smooth functioning of the overall system, since problems may affect the entirety.

Often monolithic applications are systems that are internally developed within an organization, and have been growing over many years continuously adding new functionality. This expansion increases complexity, dependencies, and complicate maintainability. Moreover, complexity and lack of transparency hamper further developments, particularly regarding dynamic changes required in flexible manufacturing systems. Criticality not only comes from the required stability of the system due to the risks imposed to depending applications in case of a system failure, but also from the high maintenance costs.

Regarding the required system capabilities, monolithic applications are a main obstacle for ***software management and maintenance***, particularly regarding the extension of the current system ecosystem.

4.3 Manual Input

The interviews revealed, that systems in the manufacturing area, particularly preparation systems, are largely based on manual input by production engineers. Since manual input is often error-prone, incorrect information propagates to other systems causing problems or inefficiencies. For instance, many process planning activities need to be done manually, which includes creation of assembly operations, estimation of assembly times, and determination of use points for material handling. Incorrect or imprecise estimation of assembly times, for example, leads to considerable inefficiencies within assembly line balancing.

Another reason why there is the need for many manual-input-reliant systems is insufficient interfaces and data connections between systems. If a process planning system, for instance, is not connected to a product lifecycle management (PLM) system the data must be provided manually. Additionally, missing data capturing systems, such as sensors that measure assembly operation times on the shopfloor, require manual theoretical estimations.

In future flexible manufacturing systems, the required capability ***autonomous decision-making*** is dependent on structured, high-quality data, which can hardly be provided using manual input processes. In addition, ***real-time data processing*** relies on fast data provisioning that cannot be managed if manual processes are involved. Regarding ***process methodology***, it is important to raise the awareness for the criticality of data created by manual inputs.

4.4 System Arrangement and Data Management Infrastructure

It can be observed that systems are typically arranged in a sequential ordering. This so-called "waterfall structure" bears the disadvantage, that flawed data in an upstream process has an impact on various downstream systems. A reason for this sequential arrangement is the lack of backward connections in the systems. Hence, there is no significant data flow from the shopfloor back to the manufacturing preparation systems. An example for an effect of this “waterfall structure” is if use pointes are specified incorrectly by the logistics engineer, which leads to material being delivered to the wrong zone on the assembly line.

Another gap regarding system arrangement is the data management infrastructure based on periodic batch updates. This means that data is entered into the systems at specific intervals, which prevents real-time use of the data [22].

Regarding the required system capabilities for flexible manufacturing, ***autonomous decision-making*** and ***real-time data processing*** suffer from unavailable real-time data due to data batches and missing backward data channels. The sequential arrangement of systems is also important in terms of ***software management and maintenance***.

4.5 Assembly Structure

The interviews showed that the physical assembly structure is the basis for both manufacturing preparation as well as for execution. The assignment of assembly operations, the determination of the tact time and the allocation of equipment to the stations are examples for important processes, which are based on the assembly structure. As a consequence, IT systems are based upon this structure as well, manifesting the rigid assembly processes within the underlying data infrastructure. For example, intralogistics operates on the basis of predefined material delivery points assigned to workstations, making it difficult to flexibly reorganize assembly processes on the assembly line. Moreover, the production program follows the rigid structure, since each production order is already pre-determined to pass through each station of the assembly line.

The assembly structure is a major obstacle for the implementation of flexible manufacturing technologies, as it dominates both system and process design. The removal of this structure is the initial consequence of the realization of flexible manufacturing technologies. But since, it is currently the standard way of working in the automotive industry and the IT system structure has evolved to fit this way of working, this is incredibly difficult both process- and systemwise.

The rigid assembly structure affects all business capabilities, however the main effect is on the required system capability ***process methodology***, as it is the current basis for all manufacturing preparation and execution activities. The use of flexible manufacturing concepts and the future flexible assembly system renders the current way of working based on a rigid assembly structure infeasible.

4.6 Utilization of CAD-Data

CAD product-data is currently not utilized for manufacturing purposes in either manufacturing preparation or execution. The interviews highlighted that visual simulations and CAD data are only used for certain verifications and validations for specific projects, but not implemented in the operational way of working in processes or systems. This can be classified as unused potential and an obstacle for flexible manufacturing, as visual data is essential for any simulations as well as for operator assistance.

The high product variation and the resulting complex product database is optimized for product development in the product documentation process. This makes the retrieval of correct visual data for the respective production process very difficult and time consuming affecting the company's ability to create and use virtual simulations. The necessary connections between the manufacturing preparation systems and the PLM system containing the CAD-data are not efficient. There is also a lack of clearly defined processes for the operational use of visual data within manufacturing.

In terms of the required system capabilities, the use of CAD-data has a significant impact on ***autonomous decision making***, since visual data is required in the data basis as well. Furthermore, the use of CAD-data needs to be considered within ***process methodology***.

4.7 Increased Variation and Complexity

The increasing product variation and complexity in the commercial vehicle industry represents a challenge for both the system ecosystem and process landscape. The process differences between different driveline technologies are extensive, resulting in increased complexity for both preparation and execution processes. This complexity is difficult to handle with manual-input based systems, and will lead to high inefficiencies on the conventional assembly line. Additionally, cognitive challenges related to handling the complexity of different processes for operators will increase heavily, requiring appropriate adaptations on the shopfloor, such as an effective system for operator assistance.

The increasing product variation and complexity mainly affect the required system capabilities of ***autonomous decision-making*** and ***process methodology.***

5 EVALUATION

The analysis conducted in this paper is evaluated by proposing a concept for a flexible assembly system and reviewing the required system capabilities in this context.

5.1 Solution Concept

5.1.1 Data Standard

One of the basic proposed adaptations concerns data standards, as unstructured data has a high proportion within the manufacturing domain. Due to the high proportion of unstructured data within the manufacturing domain, a suggested solution is to introduce clear data standards that require data in a structured format. For this purpose, process adaptations such as standardized assembly instructions can be implemented to structure the database. However, some data sets cannot be generated in a structured format and as a consequence solid ETL (Extract, Transform, and Load) processes and tools are required. ETL processes involve extracting data from various sources, transforming it into a structured format and integrate it into the target database.

5.1.2 Process Methodology

As the rigid assembly structure does not work with the flexible assembly system, a new planning and process-engineering methodology is required for manufacturing preparation as illustrated in Fig. 3. The proposal is to introduce a generic planning methodology that initially disregards physical factors, such as the number of workstations or the assembly sequence, and focuses on creating assembly operations and instructions only in relation to the components instead. These operations must be well described, with requirements like the required assembly operation time, equipment, and qualifications. Assembly processes are then assigned to assembly zones using an autonomous software solution. Logistics processes are generated automatically based on that allocation.

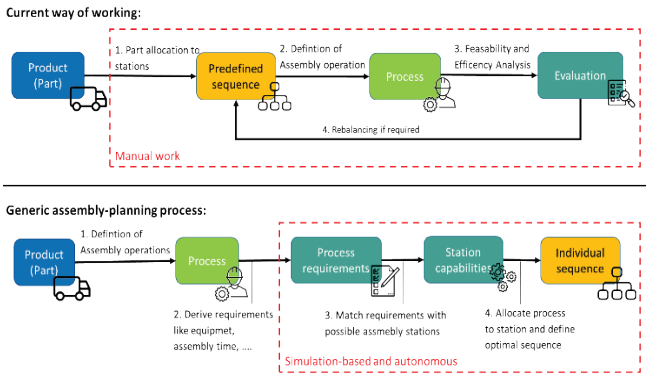


Figure 3 Generic process engineering

With a large number of possible assembly sequences for formulated assembly operations, a constraint-based model can be generated based on characteristics and technical capabilities of individual assembly zones. This model can be used to match the process requirements of assembly operations with assembly zones capabilities, described in the model. Simulations and AI preselect the solution with the best KPIs. The process engineering workload is then limited to generate accurate assembly operations, create and update the assembly system model, and verify simulated solutions. The principle of a generic approach to assembly planning was also discussed by Küber et al. [17].

5.1.3 System-Architecture Adaptation

A conceptual proposal for an adaptation of the system architecture is the transition from monoliths to service-oriented architectures (SOA), which offers easier maintenance of systems and a faster and more effective way to develop and deploy functionalities [23] thus increasing scalability. A specific well-defined implementation of the SOA is the Microservice Architecture (MOA). Microservices can encapsulate business capabilities and allow for a flexible, modular and evolvable architecture [24].

A concrete proposal for a system-architecture adaptation is a variation of the "Advanced Manufacturing Analytics Platform" by Groger et al. [25]. The foundation for the architecture is a model of the manufacturing system containing information about the manufacturing process and serving as the basis for analysis and simulation activities. The architecture consists of three layers:

The ***data integration layer*** is in charge of integrating all relevant data from different sources, and is built upon a manufacturing data warehouse, which integrates manufacturing-relevant execution data, e.g. originating from MES-systems, and operational data from PLM-, process engineering- or ERP-systems. The ***process analysis layer*** is responsible for identifying and managing process insights, which are generated using simulation-based techniques. Data provided by the manufacturing data warehouse is used in a manufacturing mining, manufacturing graph analysis and a manufacturing metrics management component. The ***process optimization layer*** takes care of the actual process optimization. Based on the process insights identified in the process analysis layer, the two techniques of indication-based manufacturing optimization and pattern-based manufacturing optimization are applied to support the process optimization.

5.2 Evaluation

For the evaluation of the proposed concept for the flexible assembly system, the required system capabilities are discussed in the context of the proposed concept adaptations.

Autonomous decision-making. Through the introduction of a data standard and concepts like the Manufacturing Analytics Platform autonomous decision making as well as an efficient database are introduced. The proposed concept is simulation-based and autonomous, reducing the amount of manual work required, and creates the data basis for autonomous systems, ultimately enabling required autonomous technologies in manufacturing preparation.

Real-time data processing. Concepts like the Manufacturing Analytics Platform aim to integrate relevant data from both manufacturing-execution and operational sources to detect deviations and optimize assembly configurations. This architecture concept avoids the sequential arrangement of systems and can be combined with the concept of SOA for faster calculations, creating the basis for a dynamic and fast responsiveness of the flexible assembly system. If the appropriate data capturing technologies are implemented as well, the basis for real-time data processing is given.

Software-management and maintenance. Monolithic architectures are a main obstacle for IT-scalability. Through the introduction of SOA, this problem is solved by providing a modular system architecture that reduces dependencies between individual applications and enables the faster development of new applications. The introduction of data standards has further positive effects on IT-scalability.

Process methodology. The proposed generic assembly-planning methodology enables the implementation of the flexible assembly system within a company's process landscape by allowing assembly planning to be conducted on the basis of the product rather than rigid assembly structures. This leads to more efficient production planning and avoids manual allocation of assembly operations to assembly zones.

6 CONCLUSION

The future of manufacturing is characterized by flexible manufacturing technologies that bring enormous challenges for the commercial vehicle industry. Based on classical business capabilities that constitute the state of the art in manufacturing-related automotive organizations, a future flexible manufacturing system was outlined comprising the concepts of flexible material delivery, Plug and Produce, flexible production sequencing and collaborative robotics. Based on these concepts, a case study for a future flexible assembly system was defined and required system capabilities were identified, namely autonomous decision-making, real-time data processing, software management and maintenance and a new process methodology. Interviews with seven domain experts in the field were conducted to identify the main gaps that need to be overcome by the commercial vehicle industry to realize a future flexible

manufacturing environment. In this paper, seven gaps were described and their relations to the required system capabilities were pointed out. A concept for a flexible assembly system was presented as a basis for evaluating the gap analysis.

Despite the inherently limited scope of a case study and expert interviews, both methods were designed in a way that the findings and conclusions are representative for the branch of commercial vehicle industry, and to a large extent also for the automotive industry in general. Future work will be a more detailed conceptualization and finally implementation of the solution concept, as well as studies in related branches.

7 REFERENCES

[1] Pietras, F. (2018, August 8). Trends in the truck & trailer market. *Roland Berger*. https://www.rolandberger.com/en/Insights/Publications/Trends-in-the-truck-trailer-market.html

[2] Aleatrati Khosroshahi, P. (2018). Using Business Capability Maps for Application Portfolio Complexity Management. *Dissertation*. Technische Universität München.

[3] Demiral, T. (2022). Future architecture of smart manufacturing systems - A holistic study of flexible manufacturing technologies for the final assembly in the commercial vehicle industry. *Master's thesis*. Technische Hochschule Ingolstadt.

[4] Henke, J. (2015). Eine Methodik zur Steigerung der Wertschöpfung in der manuellen Montage komplexer Systeme. *Dissertation*. Universität Stuttgart. Stuttgarter Beiträge zur Produktionsforschung: Band 47.

[5] Johansson, P., Eriksson, G., Johansson, P., Malmsköld, L., Fast-Berglund, Å., & Moestam, L. (2018). Assessment based information needs in manual assembly. *DEStech Transactions on Engineering and Technology Research.* https://doi.org/10.12783/dtetr/icpr2017/17637

[6] Küber, C. (2018). *Methode zur Planung modularer, produktflexibler Montagekonfigurationen in der variantenreichen Serienmontage - am Beispiel der Automobilindustrie.* Stuttgarter Beiträge zur Produktionsforschung: Band 69. Fraunhofer Verlag

[7] Groover, M. P. (2016). *Automation, production systems, and computer-integrated manufacturing*. 4th edition. Pearson.

[8] Sikora, C. (2021). Assembly-line balancing under demand uncertainty. *Dissertation*. Universität Hamburg.

[9] Boysen, N., Fliedner, M., & Scholl, A. (2007). A classification of assembly line balancing problems. *European Journal of Operational Research, 183*(2), 674-693. https://doi.org/10.1016/j.ejor.2006.10.010

[10] Hofmann, E. & Rüsch, M. (2017). Industry 4.0 and the current status as well as future prospects on logistics. *Computers in Industry, 89*, 23-34. https://doi.org/10.1016/j.compind.2017.04.002

[11] Mario, H., Dietrich, W., Gfrerrer, A., & Lang, J. (2011). *Integrated Computer-Aided Design in Automotive Development: Development Processes, Geometric Fundamentals, Methods of CAD, Knowledge-Based Engineering Data Management.* Springer.

[12] Schoner, P. (2008). Operative Produktionsplanung in der verfahrenstechnischen Industrie. *Dissertation*. Universität Kassel, Kassel University Press.

[13] Holweg, M. (2000). *The Order Fulfilment Process in the Automotive Industry - Conclusions of the Current State Analysis.* Cardiff Business School.

[14] Linda, M., Hinrichsen, S., Adrian, B., & Schulz, A. (2019). How to Improve Shop Floor Management. *Proc. of the 9th Int. Conference Production Engineering and Management.* https://doi.org/10.2307/41165745

[16] Van Ginste, L. de, Goos, J., Schamp, M., Claeys, A., Hoedt, S., Bauters, K., Biondi, A., Aghezzaf, El-Houssaine, & Cottyn, J. (2019). Defining Flexibility of Assembly Workstations Through the Underlying Dimensions and Impacting Drivers. *Procedia Manufacturing, 39*, 974-982. https://doi.org/10.1016/j.promfg.2020.01.391

[17] Küber, C., Westkämper, E., Keller, B., & Jacobi, H.-F. (2016). Planning Method for the Design of Flexible as Well as Economic Assembly and Logistics Processes in the Automotive Industry. *Procedia CIRP, 41*, 556-561. https://doi.org/10.1016/j.procir.2015.12.038

[18] Arai, T., Aiyama, Y., Maeda, Y., Sugi, M., & Ota, J. (2000). Agile Assembly System by "Plug and Produce". *CIRP Annals, 49*(1), 1-4. https://doi.org/10.1016/S0007-8506(07)62883-2

[19] Zhang, L. & Yue, X. (2011). Operations Sequencing in Flexible Production Lines With Bernoulli Machines. *IEEE Transactions on Automation Science and Engineering, 8*(3), 645-653. https://doi.org/10.1109/TASE.2011.2109061

[20] Fast-Berglund, Å., Palmkvist, F., Nyqvist, P., Ekered, S., & Åkerman, M. (2016). Evaluating Cobots for Final Assembly. *Procedia CIRP, 44,* 175-180. https://doi.org/10.1016/j.procir.2016.02.114

[21] Askin, R. G., Selim, H. M., & Vakharia, A. J. (1997). A methodology for designing flexible cellular manufacturing systems. *IIE Transactions, 29*(7), 599-610. https://doi.org/10.1080/07408179708966369

[22] Patel, K., Sakaria, Y. & Bhadane, C. (2015). Real Time Data Processing Framework. *Int. Journal of Data Mining & Knowledge Management Process, 5*(5), 49-63. https://doi.org/10.5121/ijdkp.2015.5504

[23] Ponce, F., Marquez, G. & Astudillo, H. (2019). Migrating from monolithic architecture to microservices: A Rapid Review. *Proc. of the 38th Int. Conference of the Chilean Computer Science Society (SCCC).* 1-7. https://doi.org/10.1109/SCCC49216.2019.8966423

[24] Lewis, J. & Fowler, M. (2014, March 25). Microservices. A definition of this new architectural term. *martinFowler.com*. https://martinfowler.com/articles/microservices.html

[25] Groger, C., Niedermann, F., Schwarz, H. & Mitschang, B. (2012). Supporting manufacturing design by analytics, continuous collaborative process improvement enabled by the advanced manufacturing analytics platform. *Proc. of the 2012 IEEE 16th Int. Conference on Computer Supported Cooperative Work in Design*, 793-799. https://doi.org/10.1109/CSCWD.2012.6221911

Authors' contacts:

Tarik Demiral
Technische Hochschule Ingolstadt,
Esplanade 10, D-85049 Ingolstadt, Germany

Jürgen Bock, Prof. Dr.
(Corresponding author)
Technische Hochschule Ingolstadt,
Esplanade 10, D-85049 Ingolstadt, Germany
+49 841 9348-3506
juergen.bock@thi.de

Pierre Johansson, PhD
Volvo Group,
Gropegårdsgatan 2, SE-417 15 Göteborg, Sweden
pierre.johansson@volvo.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230425154156

Preliminary communication
Received: 2023-04-25, Accepted: 2023-04-28

Leveling Maintenance Mechanism by Using the Fabry-Perot Interferometer with Machine Learning Technology

Syuan-Cheng Chang, Chung-Ping Chang*, Yung-Cheng Wang, Chi-Chieh Chu

Abstract: This study proposes a method for maintaining parallelism of the optical cavity of a laser interferometer using machine learning. The Fabry-Perot interferometer is utilized as an experimental optical structure in this research due to its advantage of having a brief optical structure. The supervised machine learning method is used to train algorithms to accurately classify and predict the tilt angle of the plane mirror using labeled interference images. Based on the predicted results, stepper motors are fixed on a plane mirror that can automatically adjust the pitch and yaw angles. According to the experimental results, the average correction error and standard deviation in 17-grid classification experiment are 32.38 and 11.21 arcseconds, respectively. In 25-grid classification experiment, the average correction error and standard deviation are 19.44 and 7.86 arcseconds, respectively. The results show that this parallelism maintenance technology has essential for the semiconductor industry and precision positioning technology.

Keywords: Fabry-Perot interferometer; interference image; leveling maintenance; machine learning; optical measurement

1 INTRODUCTION

The precision machinery and semiconductor industries are critical to high-precision positioning technology, such as semiconductor production, μLED mass transfer, and wafer positioning processing, which demand extremely high positioning accuracy [1, 2]. As technology advances and human needs require smaller and more efficient products, positioning accuracy has increased from sub-micron to nanometer scale. Therefore, positioning technology is currently one of the most important and critical technologies in these industries [3, 4].

However, traditional requirements for straightness and parallelism are insufficient to meet the demand for higher-precision mechanical components. A better active leveling maintenance system is required to meet industry demands for parallel positioning correction.

This research focuses on the development of a leveling maintenance mechanism (LMM) system that uses machine learning in conjunction with Fabry-Perot interferometer. The system is designed with considerations for optical structure, machine learning control, and feedback to enhance current industry technologies for precision parallelism correction. Based on the interference image, this study utilizes machine learning for training, effectively avoiding the accuracy and sensitivity issues caused by traditional methods.

2 THEORY AND PRINCIPLE

To construct an active parallelism maintenance system using non-contact optical interference methods, we introduce the theory of the optical structure and the machine learning methods of LMM as follows.

2.1 Fabry-Perot Interferometer

The Fabry-Perot interferometer (FPI) is an optical instrument consisting of two parallel reflecting mirrors that form a resonant cavity, as shown in Fig. 1. When light reflects inside the cavity, a series of interference fringes is formed, which are related to the parallelism of the resonant cavity [5]. When the parallelism of the two mirrors is better, the contrast and clarity of the interference fringes will be higher. Therefore, by observing the variation of interference fringes, we can determine the parallelism of the resonant cavity [6].

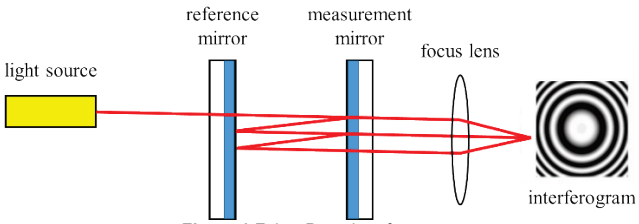


Figure 1 Fabry-Perot interferometer

The parallelism of the resonant cavity can be determined by calculating the angle of reflection mirror displacement from the number of interference fringes, as shown in Fig. 2. The relationship between fringes and angles is direct: the angle of displacement of the reflection mirror is proportional to the number of interference fringes. As the angle of displacement increases, so does the number of fringes, and vice versa. Eq. (1) illustrates the relationship between interference fringes and angles, based on optical principles.

$$2d \cdot \sin\theta = N \cdot \lambda \tag{1}$$

The letter d represents the distance between the resonant cavities, N is an integer representing the N^{th} dark fringe, and λ is the wavelength of the laser light, which is 632.8 nm. The distance between the resonant cavities is calculated as 6mm multiplied by the sine of the angle θ between the two mirrors. This formula can calculate that a displacement of 1 dark fringe corresponds to an angle θ of 0.003°, which is approximately 10.8 arcseconds. As the number of interference fringes is used as the basis for deviation, an error

of approximately 10.8 arcseconds is associated with a deviation of ±1 interference fringe.

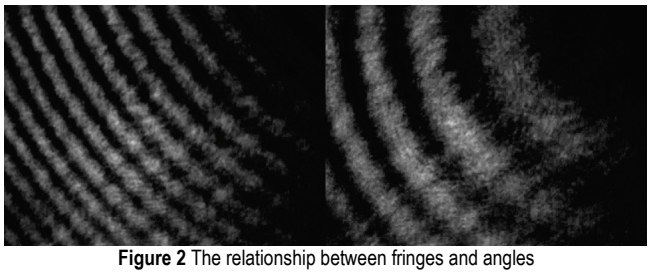


Figure 2 The relationship between fringes and angles

2.2 Machine Learning

Machine Learning is a process that uses algorithms to train data for prediction. Once a prediction model is obtained through Machine Learning, it can be used to predict new data based on the model [7, 8], as shown in Fig. 3. Machine Learning can be roughly categorized into three types: supervised learning, unsupervised learning, and reinforcement learning, as shown in Fig. 4. Supervised learning involves training a model on labeled data, where the correct output is provided for each input. Unsupervised learning involves training a model on unlabeled data, where the goal is to discover hidden patterns or structures in the data [9]. Reinforcement learning involves training a model to make decisions in an environment by receiving feedback in the form of rewards or punishments [10].

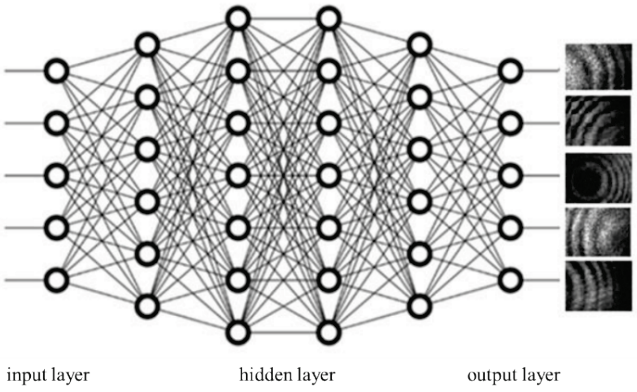


Figure 3 Machine learning algorithms

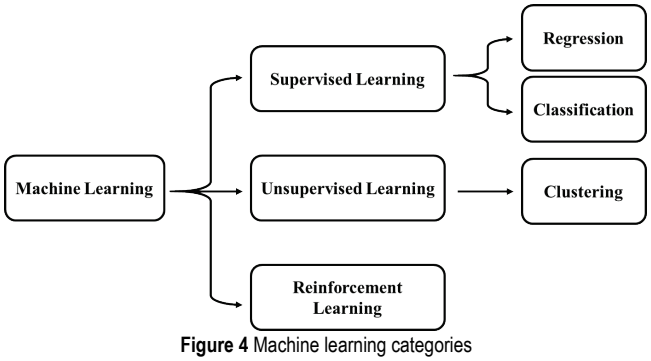


Figure 4 Machine learning categories

The experiment in this research used supervised learning, specifically classification. Supervised learning uses a training dataset to generate a model that can detect patterns and relationships between input data and output data. When new data is obtained, the model can produce accurate predictions results. In supervised learning, classification is a type of algorithm that accurately assigns test data to specific categories. It identifies specific data in a database and attempts to label or define that data to draw conclusions [11] [12]. Therefore, in this study, this method is applied to classify and train a predictive model for collected interference fringes. An interference fringe image database is established by categorizing the images into 17-grid and 25-grid categories, and machine learning is used to identify the current inclination position of the interference image, as depicted in Fig. 5 and Fig. 6.

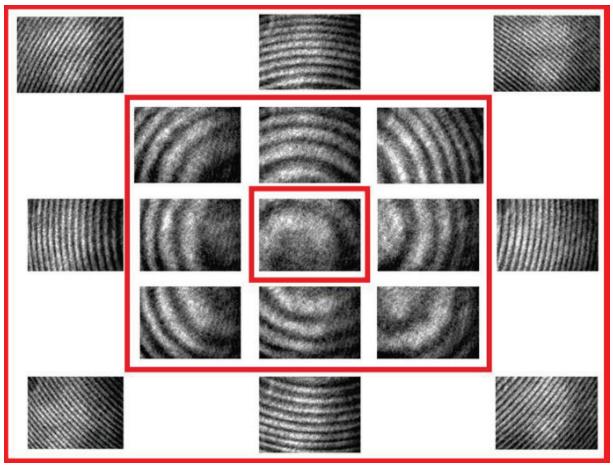


Figure 5 17-cell classification chart

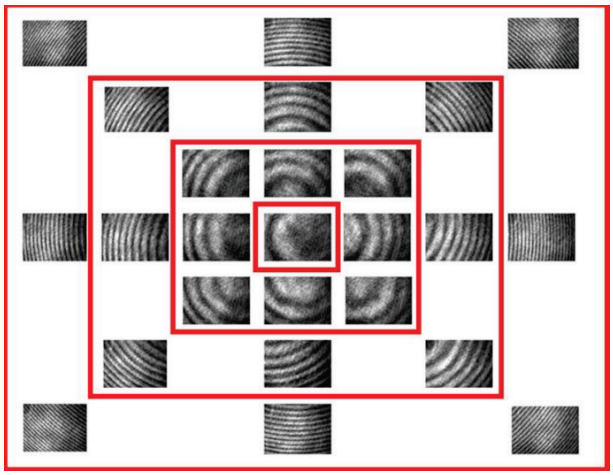


Figure 6 25-cell classification chart

The Inception v3 neural network architecture has been widely used in image recognition and computer vision tasks due to its exceptional performance in capturing complex patterns and features [13, 14]. In this study, we proposed a novel approach that utilizes Inception v3 for the automated detection of interference fringes in images captured by Fabry-Perot interferometers. Interference fringes are critical in various fields, including precision measurement and sensing, where the accurate detection of these patterns is essential for obtaining reliable data.

3 DESIGN OF PROPOSED LMM SYSTEM

The proposed LMM system is based on a Fabry-Perot interferometer, the optical structure of which is shown in Fig. 7. A laser light source that passes through a collimator, enters the beamsplitter, and then reaches the two main interferometer mirrors. The interference result is reflected back to the beamsplitter, and the interference result is obtained by the CCD.

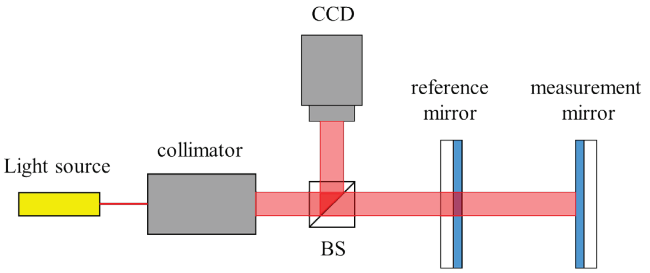


Figure 7 the structure of the LMM system

To automatically correct the parallelism of the resonant cavity, a stepper motor with adjustable angles was installed behind the measurement mirror in this study (as depicted in Fig. 8). Once the offset angle is determined, the motor can be controlled to adjust the pitch and yaw angles.

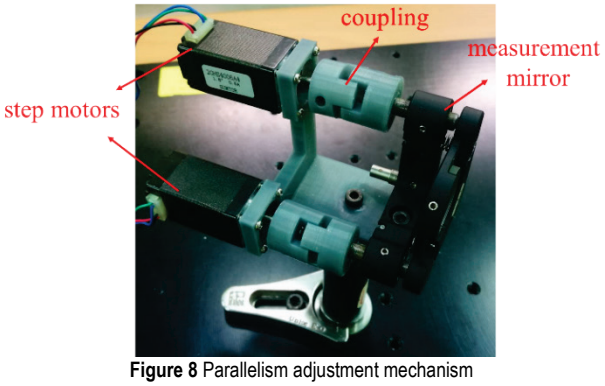


Figure 8 Parallelism adjustment mechanism

The experimental procedure of this study involves obtaining the current interference image and using a machine learning-based detection mechanism to identify whether the parallelism of the resonant cavity has shifted. After converting the recognition result into an angle, the angle can be adjusted by controlling the stepper motor. The interference image is then obtained again for confirmation. The experimental procedure is illustrated in Fig. 9.

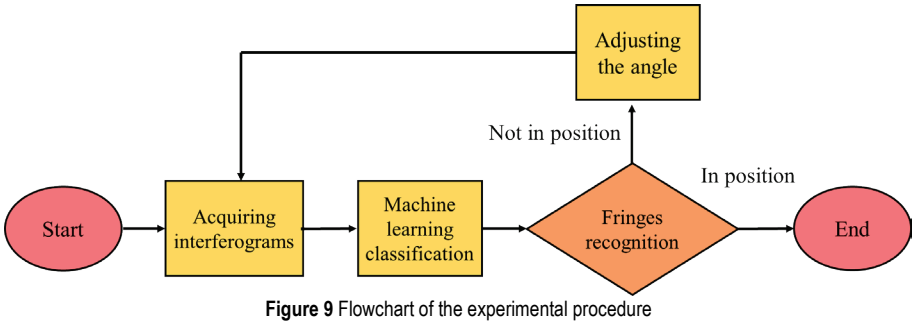


Figure 9 Flowchart of the experimental procedure

4 EXPERIMENT RESULT AND ANALYSIS

The experiments using 17-grid and 25-grid classifications were implemented in this research, as shown in Fig. 5 and Fig. 6. During the experiments, the resonance cavity angle was randomly adjusted, and the machine learning algorithm was used to predict the direction of the angle deviation. After the stepper motor was controlled to correct the angle deviation, the correction process was repeated until the predicted result was the center of the interference image.

The experiment results for the 17-grid classification are as follow. According to the experimental results, an average of three corrections could restore the parallelism of the resonator cavity, as shown in Fig. 10. The average parallelism error was ±32.38 arcseconds, as shown in Fig. 11, and the standard deviation of the error after correction was approximately ±11.21 arcseconds.

The experiment results for the 25-grid classification are shown as follow. On average, four corrections were sufficient to fix the parallelism of the resonator cavity based on the experimental results (Fig. 12). The average angle deviation was ±19.44 arcseconds (Fig. 13) and the standard deviation of the error after correction was around ±7.86 arcseconds.

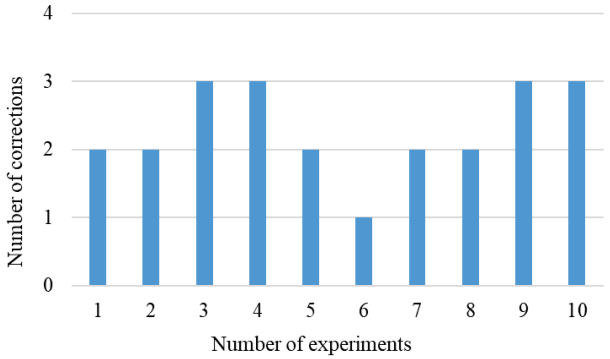


Figure 10 Number of corrections in 17-grid classification

Our proposed approach using Inception v3 and parallelism provides a powerful tool for accurate and efficient interference fringe detection in Fabry-Perot interferometers, with potential applications in precision measurement, sensing, and imaging. The combination of

Inception v3 and parallelism opens up new possibilities for enhancing the performance of interference fringe detection in various fields.

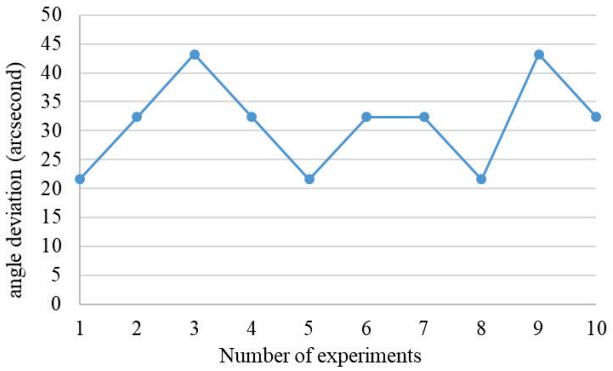


Figure 11 Angle deviation of 17-grid classification

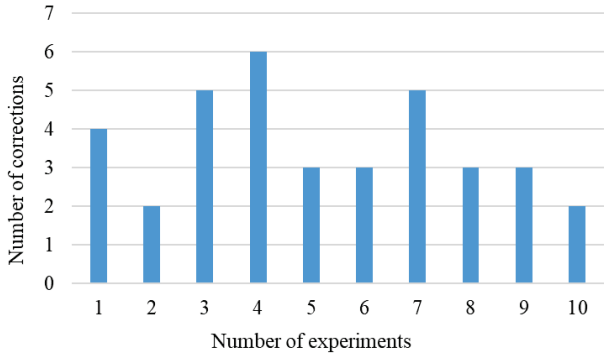


Figure 12 Number of corrections in 25-grid classification

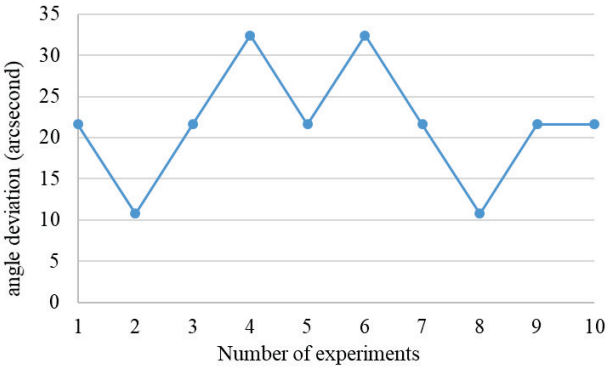


Figure 13 Angle deviation of 25-grid classification

5 CONCLUSION

This study applies machine learning to an interferometer system. According to the experimental results, the system has a resolution of 10.8 arcseconds. In the 17-grid classification experiment, the parallelism can be restored on average after three corrections, with an average correction error of 32.38 arcseconds and a standard deviation of approximately 11.21 arcseconds. In the 25-grid classification experiment, the parallelism can be restored on average after four corrections, with an average correction error of 19.44 arcseconds and a standard deviation of approximately 7.86 arcseconds.

Based on the latest research findings, it is evident that the proposed LMM system holds significant potential in enhancing precision machinery and semiconductor industries, effectively meeting their application requirements. Moving forward, further improvements will be made by integrating advanced imaging software and control systems, as well as expanding the data classifications to create a more comprehensive calibration system.

6 REFERENCES

[1] Wu, M. H. & Fang, Y. H. (2016). Picking-up and placing process for electronic devices and electronic module. *US patent*, US 20160268491 A1, Industrial Technology Research Institute.

[2] Mizuno, T., Tomoda, K., Oohata, T. (2010, October 14). Method of transferring device. US patent, US 20100258543 A1, Sony Corporation.

[3] Golda, D., Higginson, J. A. Bibl, A., Parks, P. A., Bathurst, S. P. (2016, June 12). Mass transfer tool manipulator assembly. US patent, US 9308649 B2, LuxVue Techonology Corporation, US 9308649 B2.

[4] Li, Y. C., Lai, Y. H., Lin, T. Y. (2017, February 28). Method for transferring light-emitting elements onto a package substrate. US patent, US 9583450 B2, PlayNitride Inc.

[5] Fabry, C., Perot, A. (1899). Theorie et applications d'une nouvelle methode de spectroscopie interferentielle. Ann. Chim. Phys., Vol. 16, No. 7.

[6] Gangopadhyay, T. K., mandal, S, Dasgupta, K., Basak, T. K. and Ghosh, S. K. (2005). Modeling and analysis of an extrinsic Fabry-Perot cavity. Applied Optics, 44(16), 3192-3196. https://doi.org/10.1364/AO.44.003192

[7] Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning*. MIT press.

[8] Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, 141, 61-67. https://doi.org/10.1016/j.patrec.2020.07.042

[9] Montavon, G., Kauffmann, J., Samek, W., & Müller, K. R. (2022). Explaining the predictions of unsupervised learning models. In xxAI-Beyond Explainable AI: *International Workshop*, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers, Cham: Springer International Publishing, 117-138. https://doi.org/10.1007/978-3-031-04083-2_7

[10] Brunke, L., Greeff, M., Hall, A. W., Yuan, Z., Zhou, S., Panerati, J., & Schoellig, A. P. (2022). Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, 5, 411-444. https://doi.org/10.1146/annurev-control-042920-020211

[11] Caruana, R. & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*, 161-168. https://doi.org/10.1145/1143844.1143865

[12] Niculescu-Mizil, A. & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning*, 625-632. https://doi.org/10.1145/1102351.1102430

[13] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2818-2826. https://doi.org/10.1109/CVPR.2016.308

[14] Li, Z., Zhang, L., Zhang, Z., Xu, R., & Zhang, D. (2022). Speckle classification of a multimode fiber based on Inception V3. *Applied Optics, 61*(29), 8850-8858. https://doi.org/10.1364/AO.463764

Authors' contacts:

Syuan-Cheng Chang, PhD student
National Yunlin University of Science and Technology,
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan
tso1147279@gmail.com

Chung-Ping Chang, Assist. Prof.
(Corresponding author)
National Chiayi University,
300 Syuefu Road, Chiayi 600355, Taiwan
cpchang@mail.ncyu.edu.tw

Yung-Cheng Wang, Prof.
National Yunlin University of Science and Technology,
123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan
wangyc@yuntech.edu.tw

Chi-Chieh Chu, MSc
National Chiayi University,
300 Syuefu Road, Chiayi 600355, Taiwan
Jacky01598@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230426181617

Preliminary communication
Received: 2023-04-26, Accepted: 2023-04-29

The Need for Digital Technologies in B2C Commerce from the Customer's Point of View: An Empirical Study with Focus on Sustainable Consumption

Daniela Ludin, Wanja Wellbrock*, Erika Müller, Paul Klußmann, Rebecca Schöttle

Abstract: Digitalization and digital technologies have risen sharply in commerce in form of online offers and advice, even in formerly less technology-based sectors. By using the example of flower shops, the aim of this study is to find out whether customers perceive digital offers to be generally useful or necessary, or if local service is still sufficient. To answer this question, a quantitative survey was conducted in selected flower shops in Germany. 82 customers took part. Although most customers have not yet resorted to an online offering when buying flowers, over 66% of respondents are generally in favor of digital offers in the floristry industry.

Keywords: B2C commerce; consumer behavior; digital transformation; digitalization; sustainable consumption

1 INTRODUCTION

Over the course of the 21st century, digitalization has become a necessary step toward competitiveness for many companies. For citizens, too, digital technologies are becoming an increasingly large part of their everyday live, for example through video conferencing, robots, online offers, various technologies or diverse digital infrastructures. Digitalization can help companies in various industries to open up new markets and expand their customer base. The term digitalization covers areas such as industry 4.0, innovations that help to improve processes and workflows, e.g. online offers or the digital networking of companies and customers. Digital offerings and innovations enable a company to set itself apart from other competitors and achieve potential success. Sustainability has also to be given attention to in the area of digitalization. After all, digitalization can also have a negative impact on the environment in the form of CO_2 emissions, for example, which are produced in the IT sector or in deliveries.

During the corona pandemic, many exclusively florist businesses had closed for months, whereas garden centers or supermarkets were allowed to continue selling flowers. Especially, small and medium-sized floristry businesses experienced lower sales due to fewer walk-in customers and the loss of event decorations, e.g. for weddings, etc. [1]. Therefore, for many floristry companies the only way to survive the pandemic and remain competitive was to develop or further expand online offerings, with delivery or pick-up service.

This raises the following research question: is digitalization necessary in the floristry industry and what is the level of customer acceptance?

The aim of the theoretical and empirical study is to find out whether customers generally find a digital offering in the floristry industry useful or even necessary, or whether a local offering is sufficient. For this purpose, QR codes were laid out for a survey in selected florist stores in the German city Schwäbisch Hall and thus an anonymized survey was conducted and subsequently evaluated.

This paper is divided into three thematic areas. The theoretical foundations, the methodology and the empirical results. In the first part of the paper, the theoretical foundations are explained and presented according to the superordinate scientific fields. The theory can be assigned to the scientific fields of behavioral economics, digitalization and sustainability management. The methodology is mostly based on a written survey, described in chapter 3. In chapter 4, the empirical results of the survey are presented and interpreted. Finally, a conclusion is drawn and an outlook on the need for further research is given.

2 THEORETICAL STATE OF THE ART

The paper covers the following three specific scientific fields: behavioral economics, digitalization and sustainability management.

Behavioral economics is a sub discipline of economics and thus belongs to the social sciences. It deals with human behavior within an economic situation, for example, how behavior can be influenced by economic parameters [2].

Digitalization is a complex consideration, which till now does not have a uniform conceptual definition. Digital infrastructures have already been known since the 1990s at the latest. Digitalization is connected to economic success and includes efficiency improvements, such as the creation of new digital ecosystems, the networking of companies and an expansion of diverse opportunities through cyber physical systems [3].

Sustainability management is a sub-discipline of economics. The term sustainability originally comes from forestry - one should only consume as much wood as the trees can grow back naturally. Today, sustainability refers to social, ecological and economic resources that should be consumed in such a way that future generations can also live well [4].

2.1 State of the Art – Behavioral Economics

Customers are looking for brands and companies with which they can identify themselves [5]. Environmental labels, for example, can positively influence the purchasing behavior of customers and raise consumer awareness. Wurster and Ladu believe that such certificates reveal

verified information on environmentally related properties, which would support consumers in their purchasing decisions and encourage them to buy sustainable products [6]. Psychologists had also found out that people primarily rely on their own experiences and would therefore sometimes be in conflict with new information. Thus, if people are not directly affected by a problem, such as climate change, they would find it difficult to comprehend the significance of this issue [7].

As a result, many would not act for environmentally friendly or sustainable reasons and often give little thought to the impact of their choices [8]. Customer loyalty has a high significance in retail. Not only towards a physical shopping store, but also with an online store, customers can build emotional bonds. Services, virtual representation and, for example, personalization options are very important for this. The creation of an emotional bond with an online store is of even greater importance right now, because after the covid pandemic, customers are happy to seek personal contact with retailers and other customers again [9]. During the covid pandemic, many retailers ventured towards digitalization and e-commerce. Even though some retailers adapted to hygiene regulations and made corresponding changes to their physical stores, customers preferred to use online offers, especially at the beginning of 2020 [10].

The buying behavior of consumers in the Internet also differs depending on the product. Products that do not represent a large investment for customers, i.e., whose purchase does not require a lot of thought, are mostly bought by feel. Decorations as well as bouquets of flowers are therefore ordered or purchased depending on the effect on the customer, without every last detail being checked in advance or compared with many suppliers [11]. When it comes to deliveries, the decision of many customers is shaped by the price and the delivery time. Although nowadays people often have the choice of offsetting deliveries and the resulting CO_2 emissions by donating money to charitable or environmentally friendly projects, many consumers still decide against this [12, 13].

2.2 State of the Art – Digitalization

The pandemic forced retailers to find new solutions in order to survive on the market. Therefore, many have integrated new processes and also (online) offers. Especially the delivery offer, online services and also interaction are very important in online retail. Changes and innovations are necessary for a company to increase its attractiveness and to be competitive in the long term [10]. Especially in recent years, the number of online orders has increased significantly.

For physical retailers, therefore, it is only advantageous to represent themselves online and thus create an expanded customer offering [14]. It is particularly important that the bond a customer feels when shopping is not lost, otherwise customer loyalty can suffer. The design and online environment with diverse functions and personalization options for consumers can have a major impact on customer satisfaction [9].

Innovations and developments such as industry 4.0 go hand in hand with sustainable development. They favor the integration of sustainability and thus, among other things, circular economy can also be promoted by industry 4.0 [15]. Circular economy seeks to close the loops of products and break the link between economic growth and exploitation [16]. Industry 4.0 technologies can promote models of circular economy, such as recycling or reuse of products. Appropriate technologies could be used to develop processes that make it as simple and efficient as possible to break down materials into their individual parts so that they can be reused for new products. Overall, linking technological and sustainable development can reduce resources such as energy and water, but also waste and emissions [17].

Soluk et al. [18] examine the role of dynamic capabilities as mediators in the relationship between family influence and digital business model innovation (BMI), as well as the moderating role of environmental dynamics. Based on unique survey data from 1,444 German companies with and without family influence, they show that knowledge utilization, risk management, and marketing capabilities mediate the positive relationship between family influence and digital BMI. Continuous renewal through innovation is essential to the long-term success of a business, including family-influenced businesses. Soluk et al. [18] report that previous studies on family businesses have already shown how product and process innovation activities essentially differ from non-family businesses. The background to this are the specific resources and often non-financial goals of family businesses.

The term "dynamic capabilities" describes the potential of companies to solve problems in a systematic and reliable manner as well as to seize opportunities through targeted further development and modification of their resource base. Among the various dynamic capabilities that exist, three specific dynamic capabilities may be particularly important in the context of digital BMI: the ability to utilize knowledge, the ability to actively manage risk as well as marketing capabilities. These capabilities are particularly noteworthy in the age of digital transformation, as they help companies to develop new customer-oriented business models and thus achieve competitive advantages [18].

Even though digitalization and digital technologies are not the core of the floristry business and service, research exists about combining the digital world with the flower industry, including supply chains, auctions and service offers [19, 20, 21].

2.3 State of the Art – Sustainability Management

Sustainability is becoming increasingly important in today's world, be it in food, packaging, automobiles or retail. Certified and sustainable products are used in order to achieve positive effects for the environment and for the common good of people and animals [22]. Often it is the poorer part of society that does not pay much attention to the issue of environmental protection or sustainability, as these people have other problems and concerns. Huckelba and van Lange emphasize the need for understandable sustainability

communication so that the importance of sustainability is comprehensible to everyone. People need to know how they can live more sustainably in their everyday live and what alternatives are available. Everyone has an obligation to think about their lifestyle and be aware of the long-term consequences for the environment [7].

Sustainability is also playing an increasingly important role in deliveries. In 2018, 9.3 billion packages were ordered and delivered in Europe alone. Since Corona, the figures have risen even more dramatically. Transport and deliveries, as well as the entire supply chain, generate huge amounts of CO_2 emissions [12].

The transport sector alone accounts for around 25% of global greenhouse gases [13]. According to various studies, the so-called "last mile" in particular generates between 30% and 50% of the CO_2 emissions generated by a product ordering process [12]. To make deliveries more sustainable, customers would have to settle for higher prices and longer delivery times. Thus, more efficient routes could be selected, drivers' working conditions and wages could be adjusted, and waste and emissions could be reduced [13]. Depending on the business sector or company size, sustainability plays a different role and is often seen as a costly challenge [23]. However, integrating sustainability can be seen as an investment and brings long-term financial benefits for the company, as the needs of the stakeholders are met and the company value can increase [24]. Young companies in particular often integrate sustainability into their corporate goals from the outset. This integration is important to make the jump to an established company and strengthen the chances of survival [23]. Retailers should also link their corporate strategy with a sustainability strategy in order to be successful and competitive [24].

Sustainability standards and labels provide more awareness and attention. For each of the three pillars of sustainability – economy, ecology and social – exist standards, some of which are represented in more than 170 countries. Sustainability standards can contain general requirements and improvements regarding sustainability in companies or, for example, serve the reporting of non-financial information. In agriculture, standards such as Fairtrade, Demeter or SA8000 can be found above all [25].

3 METHODOLOGICAL APPROACH

In a survey, random samples from a defined population are interviewed. The aim of a survey is to collect primary data on the responses of the respondents. The procedure of a survey contains the following steps: setting of a goal, defining the population, creating a schedule, considering the sample size, creating a questionnaire, conducting the survey, and finally analyzing it.

For the questionnaire, there are certain rules regarding the formulation of questions. Above all, questions should be formulated simply and briefly so that they are understandable for everyone. In addition, neutrality and a hypothetical formulation is important in order not to overwhelm the respondent. Furthermore, avoiding unnecessary personal data of the respondents as well as ensuring anonymity are very important.

Before starting the survey, a pre-test should take place to check whether the aforementioned regulations are met. A general distinction is made between open and closed questions. Open questions do not have any answer specifications and are mostly used if one expects a large diversity of answers. Closed questions, on the other hand, contain a certain number of answer options, which can also be provided with open question parts [26].

For the survey of this paper, in the period from April 7th to April 27th 2022, QR codes with a connection to the questionnaire were displayed in the florist stores "Blumen Bierbach GmbH" and "Straußbinderei Starz" in the German city Schwäbisch Hall. The stores represent small businesses as well as family businesses that only have local stores and do not belong to a big company chain like e.g. Blume 2000. They are located in the vicinity of the researchers' university and were therefore chosen due to practicality reasons. Besides that, small flower shops have been hit especially hard by the lockdowns [1] and might have limited capacities to adapt to digitalization processes. To map customers' wishes and positions on digital technologies might offer orientation for these companies to know what to focus on, when trying to change to a more digital service approach [27]. The costumers were surveyed mainly on the subject of digital technologies and offers in the floristry industry. The aim was to find out to what extent digitalization is generally desired in the floristry sector and whether a digital offering is already being used.

Straußbinderei Starz has a flower store in Schwäbisch Hall with a special focus on regionalism of their products. The company is represented online by a homepage and accepts orders by telephone. However, there is no online ordering procedure.

The family company Blumen Bierbach GmbH can be found both directly in Schwäbisch Hall and in Gelbingen. However, the QR codes were only displayed directly in Schwäbisch Hall. Blumen Bierbach GmbH offers a telephone order, delivery and pick-up service, and an online contact form can also be used to place orders.

In total, 82 customers took part in the survey and answered the standardized questionnaire with six closed questions on the scientific areas of behavioral economics, digitalization and sustainability management.

At the beginning, the questionnaire was tested in a pre-test by three experts to ensure the plausibility and comprehensibility of the survey.

The paper is based on an exploratory study to give an overview on the relevant research area. Therefore, in a first step only descriptive analyses are used. Further research should take more detailed analyses using more specific data and inductive methods as for example multivariate regression analysis.

4 EMPIRICAL RESULTS

4.1 Empirical Results – Behavioral Economics

Florists were closed for several months during the corona pandemic. Whether and where the customers bought flowers during this time is shown in Fig. 1. In total, the question was answered by 82 customers. The answer option "no" received 37 percent of the votes, 45 percent of the costumers bought

flowers in supermarkets, which were not closed during the pandemic, and only eight percent voted for the answer "online". For the category "others" the respondents were able to provide own examples of where flowers could be purchased during the Corona pandemic. Examples are gas stations and nurseries. Fig. 1 indicates that a large proportion of the customers has not purchased flowers during the corona pandemic, despite having several alternatives.

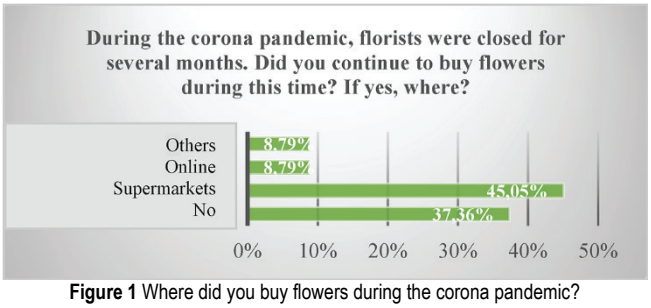


Figure 1 Where did you buy flowers during the corona pandemic?

The question whether customers have ever bought flowers online was neglected by a majority of 74% of the customers. Looking at the answer option "yes", "Fleurop" and "Blume 2000" are the most known online companies each with 13 percent of the votes. One person also indicated the online store of Floraprima. The opportunity to justify why they are using the online offer was only used by two of the respondents. The arguments given here were firstly the greater choice available in an online store and secondly the simplicity that comes with a digital ordering and delivery process. Although much more shopping was done online in general during the pandemic, this does not include the floristry industry. As indicated with figure 1, the majority of costumers continued to buy flowers in on-site alternatives during the lockdown phase. Most costumers still largely ignore digital offerings.

4.2 Empirical Results – Digitalization

The statement "an online offer weakens the bond with the local retailer" could be confirmed or declined by the costumers on a scale from 0 ("I fully disagree") to 10 ("I fully agree"). 27 percent of the 79 respondents voted for the category 10, whereas only one person voted for 0. Nine percent were neutral on this statement (see Fig. 2). Overall, more than half of the respondents agreed with the statement. Although, according to other research, an emotional bond is also possible with online opportunities, regarding the survey results, the situation in the floristry industry seems to be quite different. This result could also explain why the majority of respondents have never used an online florist offer.

As mentioned before, digitalization in the retail sector has increased in the last two years in particular, and the number of online shoppers has risen sharply. Although more than two-third of the respondents have never used the online services of florists, the clear majority is nevertheless not averse in principle to digital technologies and offers in the floristry industry. The question "does a digital floristry offering make sense in principle" was positively answered by 67 percent of the respondents.

The opportunity to make additional statements on this question in an open field was used by 23 respondents. Nineteen statements were in favor of digital technologies and offers, while three were against and one person was neutral. Reasons supporting a digital offer include simplicity due to independence of location, flexibility and a lower time requirement. Flowers can be bought outside regular opening hours or, for example, despite illness. In addition, by digital offers it is possible to send bouquets to family members and friends at a certain time or cause even if a spatial distance is present. Furthermore, the price and the selection range are called as arguments for an online offer.

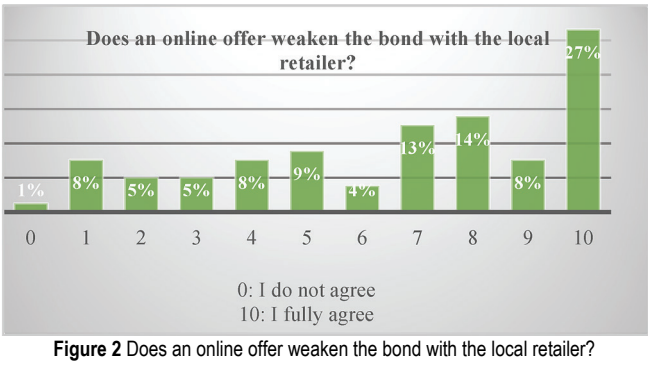


Figure 2 Does an online offer weaken the bond with the local retailer?

The following reasons speak against an online offer for the customers. Since flowers are a fresh product, the condition and quality of the flowers can best be assessed on site. In addition, the essential individual customer consultation is only possible on site. The arguments that there must be a basis of trust with the retailer and that it should be a local florist go in the same direction.

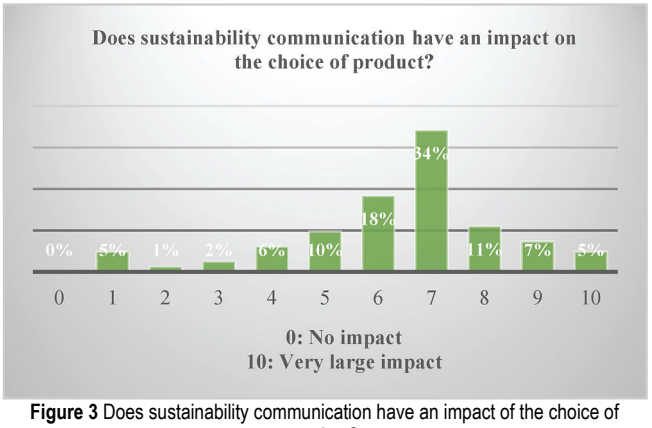


Figure 3 Does sustainability communication have an impact of the choice of product?

4.3 Empirical Results – Sustainability Management

In another question, scaled again from 0 to 10, the costumers were asked whether and to what extent a purchase decision depends on the sustainability communication of the product. The answer option 0 states that sustainability communication has no impact on the choice of product, while 10 stands for a large impact. For 5 percent of the customers, sustainability communication has a very large impact on the product choice. The answer options "0 to 4" cumulate 14 percent and the answer options "5 to 9" 81 percent of all

costumers (see Fig. 3). Therefore, the answers state clearly that sustainability communication has a major impact on the product choice of the 82 respondents.

Among other things, labels can be used to ensure that a product has been produced sustainably. The 82 respondents could indicate which labels they are looking for when buying flowers. A multiple selection was possible (see Fig. 4).

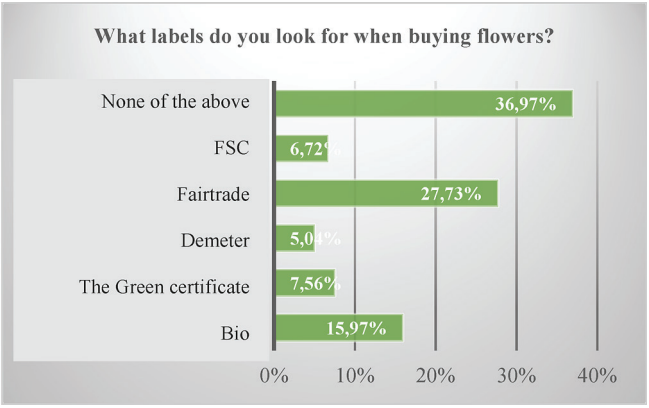


Figure 4 What labels do you look for when buying flowers?

The abbreviation "FSC" stands for the international non-profit organization "Forest Stewardship Council". The label issued by the organization guarantees that the wood used for a product comes from a sustainable and environmentally friendly production. This label can usually be found on wood and paper products. Flowers cannot be directly labeled with an FSC sign. However, it can be found on some wrapping paper for flowers. Only six percent of the costumers voted for this label. Therefore, it can be assumed that many customers either do not know what the FSC label stands for in detail, or that they extend their understanding of sustainability only to the origin of the flowers and not to the corresponding packaging material. The customers voted most frequently for the "BIO" and "Fairtrade" label, whereby the highest percentage answered "none of the above". This shows that although sustainability communication has a major influence on the purchase decision, labels are ignored by more than one third of all costumers when buying flowers.

5 CONCLUSION

Digital technologies are becoming increasingly important in today's world. Change and innovation are essential for companies to be competitive. Another essential component for a competitive company is the integration of the three pillars of sustainability: the economic, ecological and social dimension. It can also be seen in the floristry industry that digitalization is growing in importance. This is particularly evident in the increasing number of digital offerings. Following the results of this survey, most costumers are in favor of digital technologies and offers, even if the services of local retailers are still preferred. Nevertheless, the digitalization of the floristry industry should be driven forward in order to be competitive in the future.

The results of the online survey shed light on the purchasing behavior of consumers in the floristry industry. Although most customers have not yet made use of online offers, 67% of the respondents are generally in favor of digital technologies in this sector. However, it may still take some time before online floristry offers are completely accepted. Nevertheless, the findings from this survey are in line with findings from other surveys. According to a survey among German, French, Dutch and British customers by the market research agency Motivaction, 24% of the survey participants have bought flowers and 26% have bought plants online for the first time since the beginning of the corona pandemic. About 50% of the participants rated it as positive that flower stores and garden centres offer delivery services for flowers and plants [28].

This research has shown that digital technologies and digital offers are seen as useful for many industries and companies from the customer's point of view. However, since most of the respondents have never purchased flowers online, one of the questions that arises is what factors would make online floristry offers more attractive. Further research is also needed in the area of sustainable floristry. Although sustainability generally influences the purchasing behavior of consumers, many of the customers do not pay attention to sustainable labels when buying flowers. In this context, it could also be investigated why labels receive little attention in this industry and whether more transparency or even industry-own labels would ensure more awareness and greater attention to sustainability.

Additionally, it should be noted that this empirical research only looks at the consumer side and the supply side was not surveyed. This leads to further research into the acceptance and benefits of digitalization on the part of suppliers.

Given the fact that the survey sample only consists of customers of two flower shops in one German city, the results cannot be generalized, but should be seen as exemplary insights into this branch. Additionally, people that already made use of online offers to buy flowers online were not included. To be able to generalize the findings, it would be helpful to survey a higher number of flower shops and to take locality and region (rural or urban), customer habits and age as size, ownership structure and other characteristics of the flower shop into consideration. Besides that the survey was executed at a point in time, when pandemic restrictions were still in place more or less, which impacted not only customer behavior but also business behavior and their services. As a study among US florist shops showed that almost all flower shops offered no-contact delivery service due to the pandemic restrictions. Social media marketing was used by about one third of the stores as well as similar share offered no-contact shopping options [29]. It might also be interesting to research on possible changes in customer and shopping behavior after the pandemic.

6 REFERENCES

[1] Online Retailer (News 2021). https://www.onlinehaendler-news.de/online-handel/haendler/134953-krise-online-handel-umsatz-aufgefangen visited 24.04.23

[2] Fischer-Korp, C (2018). *Erfolgreiche Change-Prozesse im öffentlichen Bereich*. Wiesbaden. https://doi.org/10.1007/978-3-658-17331-9

[3] Roth, S. & Corsten, H. (2022). *Handbuch Digitalisierung*. München. https://doi.org/10.15358/9783800665631
[4] Kropp, A. (2019). *Grundlagen der nachhaltigen Entwicklung: Handlungsmöglichkeiten und Strategien zur Umsetzung*. Wiesbaden. https://doi.org/10.1007/978-3-658-23072-2
[5] Aybaly, R. (2017). Sustainability in the automotive world: the case of Tesla. *Procedia Computer Science*, 122. https://doi.org/10.1016/j.procs.2017.11.404
[6] Wurster, S. & Ladu, L. (2020). Bio-based products in the automotive industry: the need for ecolabels, standards and regulations. *Sustainability, 12*(4). https://doi.org/10.3390/su12041623
[7] Huckelba, A. L. & van Lange, P. A. M. (2020). The silent killer: consequences of climate change and how to survive past the year 2050. *Sustainability, 12*(9). https://doi.org/10.3390/su12093757
[8] Aliyev, F., Wagner, R., & Seuring, S. (2019). Common and contradictory motivations in buying intentions for green and luxury automobiles. *Sustainability, 11*(12). https://doi.org/10.3390/su11123268
[9] Horákocá, J., Uusitalo, O., Munnukka, J., & Jokinen, O. (2022). Does the digitalization of retailing disrupt consumers' attachment to retail places? *Journal of Retailing and Consumer Services*, 67. https://doi.org/10.1016/j.jretconser.2022.102958
[10] Pilawa, J., Witell, L., Valtakoski, A., & Kristensson, P. (2022). Service innovativeness in retailing: increasing the relative attractiveness during the COVID-19 pandemic. *Journal of Retailing and Consumer Services*, 67. https://doi.org/10.1016/j.jretconser.2022.102962
[11] Huang, S. & Lin, Y. (2020). Exploring consumer online purchase and search behavior: an FCB grid perspective. *Asia Pacific Management Review, 27*(4), 245-256. https://doi.org/10.1016/j.apmrv.2021.10.003
[12] Rai, H.B., Broekaert, C., Verlinde, S., & Macharis, C. (2021). Sharing is caring: how non-financial incentives drive sustainable e-commerce delivery. *Transportation Research Part D: Transport and Environment*, 93. https://doi.org/10.1016/j.trd.2021.102794
[13] Nogueira, G., Rangel, J., & Shimoda, E. (2021). Sustainable last-mile distribution in B2C e-commerce: do consumers really care? *Cleaner and Responsible Consumption*, 3. https://doi.org/10.1016/j.clrc.2021.100021
[14] Truong, D. (2022). How do customers change their purchasing behaviors during the COVID-19 pandemic? *Journal of Retailing and Consumer Services*, 67. https://doi.org/10.1016/j.jretconser.2022.102963
[15] Khan, S.A.R., Umar, M., Asadov, A., Tanveer, M., & Yu, Z. (2022). Technological revolution and circular economy practices: a mechanism of green economy. *Sustainability, 14*(8). https://doi.org/10.3390/su14084524
[16] Castro, C. G., Hofmann Trevisan, A., Pigosso, D. C. A., & Mascarenhas, J. (2022). The rebound effect of circular economy: definitions, mechanisms and a research agenda. *Journal of Cleaner Production*, 345. https://doi.org/10.1016/j.jclepro.2022.131136
[17] Laskurain-Iturbe, I., Arana-Landin, G., Landeta-Manzano, B., & Uriarte-Gallastegi, N. (2021). Exploring the influence of industry 4.0 technologies on the circular economy. *Journal of Cleaner Production*, 321. https://doi.org/10.1016/j.jclepro.2021.128944
[18] Soluk, J., Miroshnychenko, I., Kammerlander, N., & De Massis, A. (2021). Family influence and digital business model innovation: the enabling role of dynamic capabilities. *Entrepreneurship Theory and Practice, 45*(4), 867-905. https://doi.org/10.1177/1042258721998946
[19] Fernandes Barbosa, M. (2021). Digital Transformation in the Brazilian Flower Industry: Cultural and Technological Issues, Economic Benefits and Firms' Actual Adoption (*Doctoral dissertation*, Politecnico di Torino).
[20] van Heck, E. (2021). *Technology Meets Flowers: Unlocking the Circular and Digital Economy*. Springer Nature. https://doi.org/10.1007/978-3-030-69303-9
[21] Salvini, G., Hofstede, G. J., Verdouw, C. N., Rijswijk, K., & Klerkx, L. (2022). Enhancing digital transformation towards virtual supply chains: a simulation game for Dutch floriculture. *Production Planning & Control, 33*(13), 1252-1269. https://doi.org/10.1080/09537287.2020.1858361
[22] Wurster, S. & Ladu, L. (2020). Bio-based products in the automotive industry: the need for ecolabels, standards and regulations. *Sustainability, 12*(4). https://doi.org/10.3390/su12041623
[23] Bernal, P., Domínguez, B., & Montero, J. (2021). When are entrepreneurs more environmentally oriented? An analysis of stakeholders' pressures at different stages of evolution of the venture. *Business Strategy and the Environment, 31*(3), 828-844. https://doi.org/10.1002/bse.2920
[24] Buallay, A. (2022). Sustainability reporting and retail sector performance: worldwide evidence. *The International Review of Retail, Distribution and Consumer Research, 32*(3). https://doi.org/10.1080/09593969.2022.2048410
[25] Jellema, S. F., Werner, M. D., Rasche, A., & Cornelissen, J. (2022). Questioning impact: a cross-disciplinary review of certification standards for sustainability. *Business & Society, 61*(5), 1042-1082. https://doi.org/10.1177/00076503211056332
[26] Schnell, R. (2019). *Survey-Interviews: Methoden standardisierter Befragungen*. Wiesbaden. https://doi.org/10.1007/978-3-531-19901-6
[27] Priyono, A. et al. (2020). Identifying Digital Transformation Paths in the Business Model of SMEs during the COVID-19 Pandemic. *Journal of Open Innovation: Technology, Market, and Complexity, 6*(4), 104. https://doi.org/10.3390/joitmc6040104
[28] Blumenbüro Holland (2022). https://www.blumenbuero.de/f%C3%BCr-die-presse/auswirkungen-von-covid-19-auf-den-kauf-von-blumen-und-pflanzen (Accessed: 25.04.2023)
[29] Etheredge, C. L. & DelPrince, J. (2021). Retail Florists Use Novel Sales Approaches during the COVID-19 Pandemic. *HortTechnology, 31*(6), 786-792. https://doi.org/10.21273/HORTTECH04920-21

Authors' contacts:

Daniela Ludin, Prof. Dr.
Heilbronn University of Applied Sciences,
Bildungscampus, 74076 Heilbronn, Germany
daniela.ludin@hs-heilbronn.de

Wanja Wellbrock, Prof. Dr. (Corresponding author)
Heilbronn University of Applied Sciences,
Bildungscampus, 74076 Heilbronn, Germany
wanja.wellbrock@hs-heilbronn.de

Erika Müller
Heilbronn University of Applied Sciences,
Bildungscampus, 74076 Heilbronn, Germany
erika.mueller@hs-heilbronn.de

Paul Klußmann
Heilbronn University of Applied Sciences,
Ziegeleiweg 4, 74523 Schwäbisch Hall, Germany
paul.klussmann@outlook.de

Rebecca Schöttle
Heilbronn University of Applied Sciences,
Ziegeleiweg 4, 74523 Schwäbisch Hall, Germany
rebecca.schoettle@web.de

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230502171228

Preliminary communication
Received: 2023-05-02, Accepted: 2023-05-05

Identification of Inability States of Rotating Machinery Subsystems Using Industrial IoT and Convolutional Neural Network – Initial Research

Davor Kolar*, Dragutin Lisjak, Martin Curman, Juraj Benic

Abstract: Rotating parts can be found in almost all operational equipment in the industry and are of great importance for proper operation. However, reliability theory explains that every industrial system can change its state when failure happens. Predictive maintenance as one of the latest maintenance strategy emerged from the Maintenance 4.0 concept. Nowadays, this concept can include Industrial Internet of Things (IIoT) devices to connect industrial assets thus enable data collection and analysis that can help make better decisions about maintenance activity. Robust data acquisition system is a prerequisite for any modern predictive maintenance task as it provides necessary data for further analysis and health assessment of the industry asset. Fault diagnosis is an important task in the maintenance of industrial rotating subsystems, considering that early state change diagnosis and fault identification can prevent system failure. Vibration analysis in theory and practice is considered a correct technique for early detection of state changes and failure diagnostics of rotating subsystems. The identified technical state should be considered in a context of the ability and different inability states. Therefore, early different inability states identification is the next step in the rotary machinery diagnostics procedure. Most of the existing techniques for fault diagnosis of rotating subsystems that use vibrations involve the step of extracting features from the raw signal. Considering that the features that describe the behavior of the rotary subsystem can differ significantly depending on the type of equipment, such an approach usually requires an expert in the field of signal processing and rotary subsystems who can define the necessary features. Recently, the emergence of machine deep learning and its application in maintenance promises to provide highly efficient fault diagnostics while simultaneously reducing the need for expert knowledge and human labour. This paper presents authors aim to use self-developed IIoT system built as an IIoT accelerometer as the edge device, web API and database with convolutional neural network as deep learning-based data-driven fault diagnosis to detect and identify different inability states of rotating subsystems. Large dataset for two different rotational speed is collected using IIOT system and multiple convolutional neural network models are trained and tested to examine possibility of using IIOT for inability state prediction.

Keywords: accelerometer; automated data collection; CNN; fault diagnosis; Industrial Internet of Things

1 INTRODUCTION

Technology in today's world makes it possible to collect an ever-increasing amount of data. In such a situation, companies strive to optimize their processes by collecting and processing data and thus become better. According to the [1], companies that do not include data from the maintenance process in optimizing the efficiency of production processes are unable to fully utilize their resources and assets.

In order to do so, there are more and more examples in which predictive maintenance (PdM) shows clear potential and gain benefits when compared to other maintenance strategies. However, implementation of predictive maintenance usually includes additional costs in terms of hardware and software technology to support this strategy [2]. Nowadays, with the emergence of the Industry 4.0 concept, it can be said that one of the long-term technology an Industrial Internet of Things (IIoT) framework. Using of IIoT can help in ensuring robust and lightweight system for data collecting from the machinery and sending the data to the final destination. According to authors in [3], industrial internet of things seeks to connect various industry assets to identify, communicate, sense, process, operate and work together.

Assets characterized as a rotating system are often described as critical due to their working conditions and importance in the overall production process. Since rotary machines usually operate under severe operating conditions, this makes them more sensitive to various types of errors and increases the complexity of assessing their condition and identifying possible failures. Previous research and the author's experience lead to the conclusion that failures of such systems result not only in direct losses in the production process, but often have a long-term negative impact on the economic, environmental and security position of the company.

Vibration analysis is considered as powerful technique that can be used to assess rotary machinery health state since using vibration sensor can sense changes and enable fault diagnosis in early stage. Most commonly used sensors in vibration measurements are accelerometers. Using standard industrial IEPE accelerometers provides reliable yet expensive solution for data acquisition. The use of an IIoT system in combination with a MEMS accelerometer can be considered as an alternative due to the lower cost and the possibility of connecting to the Internet and transferring data over it. Due to this feature, data can be collected from different locations where internet connectivity is available.

With the ever-increasing amount of data collected, more and more resources are being invested in the development of techniques for assessing the condition of rotating equipment. Additionally, because of the continuous improvement of data acquisition ability [4], as well as the exponential growth of data volume [5], machine learning and artificial intelligence techniques for fault diagnosis have achieved great success and received widespread attention within PdM.

The main goal of this paper is to show the effort of the authors in both development of the IIOT data acquisition system and using sensor data in intelligent fault diagnosis performed by convolutional neural network model trained on data acquired from the developed system. The rest of the paper consists of the following: the related work in the field

of intelligent fault diagnosis using IIoT is presented in Section 2, Section 3 describes developed IIoT while performed experiment as well as CNN architecture description is shown in Section 4. Results of experiment are presented in Section 5 and conclusion is drawn in Section 6.

2 RELATED WORK

As the field of application of IIOT for fault diagnosis in PdM is increasing, the number of papers related to this topic is increasing daily. In this section, the papers are briefly described. In their paper [6], Tiboni et al. provides extensive review of vibration-based condition monitoring of rotating machinery, concluding that is very likely that innovative diagnostic methods based on machine learning will be developed in the near future.

Paper [7] gives an overview of current state of predictive maintenance and intelligent sensors in smart factories. The results show four different types of maintenance used in smart factories—Industry 4.0 for predictive maintenance, smart manufacturing for condition-based maintenance, fault diagnosis for maintenance and prognostics, and remaining useful life analysis. The importance of predictive maintenance is also growing due to the growing number of robots, digitisation, and artificial intelligence introduced into production lines to automate routine activities. Paper [8] presents experiences in setting-up two different remote vibration monitoring systems using low-cost MEMS accelerometers available on the market in two different industrial settings. The installed vibration monitoring systems have successfully detected faults on two different critical assets.

The paper [9] presents developed experimental setup that used low-cost MEMS accelerometers and simple edge computing module to acquire and process sensor data. Further on, in a related paper [10] Raspberry Pi edge module is used in combination with a MEMS accelerometer to perform continuous monitoring of the machine tools. Server side Phyton program has been implemented in order to process data and calculate Fast Fourier transform. Frequency spectrum is used for further assessment. Node RED technology for hardware device wiring was used in [11] to collect temperature and humidity data using a Raspberry Pi. Node RED is an open-source rapid embedded environment design for easier integration of IoT devices and related software. In their previous work [12, 13], authors discussed possibilities of developing and using the IIOT system for condition monitoring yet they did not implement intelligent fault diagnosis using convolutional neural network.

Currently, in the field of predictive maintenance, many works can be found that deal with the application of convolutional neural networks for intelligent fault diagnosis. These works can be classified in different ways. If the structure of the convolutional neural network is observed, they can be divided into 1-dimensional (1D) fault diagnosis models and 2-dimensional (2D) fault diagnosis models. Historically, convolutional neural networks were originally developed to classify images that are 2-dimensional, so in such systems the input signal is expected to be in two dimensions. The signals collected from the vibration sensors differ. They are usually in one-dimensional form. Therefore, for the purposes of processing such a signal, the convolutional neural network was adapted to accept 1D signals as well. In recent years, multiple 1D CNN techniques for induction motors [14, 15], pumps and rolling element bearings [16] are presented in various papers. Author of this paper in [17] their previous work presented multi-channels 1D CNN (MC-DCNN) for human activity classification modified for 3 axis vibration data input with input size of $6400 \times 1 \times 3$. This paper focuses on application of modified 1D CNN with input size of $1600 \times 1 \times 3$ that is capable of dealing with accelerometer data acquired from developed IIOT system. Data is acquired for healthy baseline as well as 3 different inability states. Later on, different CNN based models for inability state detection are trained and evaluated on the acquired data to provide better understanding in system behaviour under different inputs.

3 IIOT SYSTEM DESIGN

Developed IIOT system architecture capable of data acquisition and storage is presented in Fig. 1.

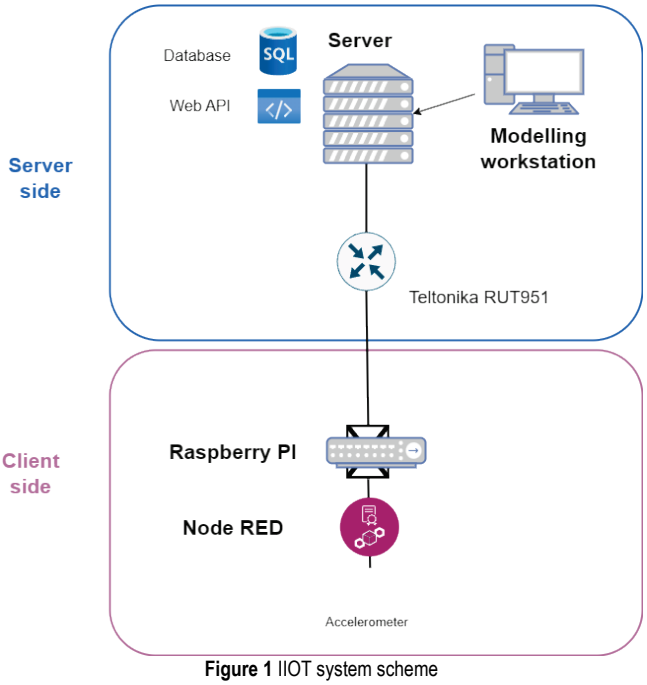


Figure 1 IIOT system scheme

It can be seen that the system consists of two separate subsystems named server and client side. The server side consists of the following components:

1) Teltonika RUT951 industrial IOT router – configured as access point for Raspberry PI and network holder.
2) Application/database server - hosts an IIS web server that runs the Web API. There is also an MSSQL server with a database on this server. The client side consists of the following components:
3) Raspberry PI 4 8GB – configured as edge module with Node-RED installed and HAT module used for sensor

connectivity. Installation-ready setup of the Raspberry Pi with HAT can be seen in Fig. 2. Raspberry PI is directly connected to local area network provided by Teltonika router.

4) IIOT sensor - for vibration data acquisition. Widely known KX122 MEMS sensor is used (Fig. 3).

The Raspberry Pi is configured as an edge node with an attached sensor. It runs the Node RED programming environment, which is designed for IoT environments. In the developed system, Node RED serves as a management tool which takes care of data collection and proper functioning of the client-side part of the IIoT system. Communication between the used sensors is performed with the help of the standards-based messaging MQTT protocol. MQTT is a fast and lightweight protocol that allows messaging between devices located on unstable networks and ensures secure, reliable and two-way messaging [18]. It is designed as a lightweight publish/subscribe messaging transport that is ideal for connecting remote devices with a small network bandwidth.

Figure 2 HAT module installed on the top of the Raspberry PI board

For this research, system is configured to acquire data from 1 accelerometer. Using described configuration, up to 8 sensors can be attached to one Raspberry PI edge node. This study use 3-axis MEMS accelerometer that can be configured to acquire data with sampling rate of up to 25,6 kHz. Accelerometer is shown in Fig. 4.

Figure 3 KX122 MEMS

Accelerometer specifications are listed in Tab. 1.

In this study, client-side is setup to acquire data with sampling rate of 1.6 kHz and ±8 g range with 16-bit resolution. It is possible to redefine sampling time for each sample that is initially set to 1 second.

Table 1 MEMS accelerometer properties

Property	Value
Output data rate	0.781 Hz - 25.6 kHz
Full-scale range	± 8 g
Sensitivity	4096 - 16384 counts/g
Offset	± 20 mg
Non-Linearity	0.6 %
Resolution	0.0001 g, 16-bit
Input voltage	1.71 – 3.6 V
Current consumption	145 mA
Output voltage	1.368 - 28.8 V

Data acquired is filtered using low-pass filter at 800 Hz to eliminate the possibility of aliasing. Using Node RED environment flows, each acquired sample containing 1600 samples is sent using the MQTT protocol to Web API and stored to SQL Server database. Authors previous research shown that, probably due to fact that MQTT protocol relies on publish/subscribe, it is not always possible to ensure data quality with higher sample rates (with sampling frequency <3 ms). To be able to collect with higher sampling rates, the system is developed in a way that it enable continuous call for streams of values that are sent from the accelerometer as series of values.

4 EXPERIMENTAL SETUP

4.1 Test Bench

This study use machine fault simulator for generating signal for both healthy baseline and inability technical states of the rotary machinery. A SpectraQuest variable speed Machinery Fault Simulator (MFS) installed in Laboratory for Maintenance of University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture is used. The simulation stand (Fig. 4) consists of main shaft loaded with main load and driven by 0.7 kW VFD powered motor. Main shaft is connected to motor with coupling. Additionally, there are two ER-12K rolling bearings that supports shaft assembly. Finally, simulator is equipped with a three-axis accelerometer and a tachometer that are connected to a developed IIOT system.

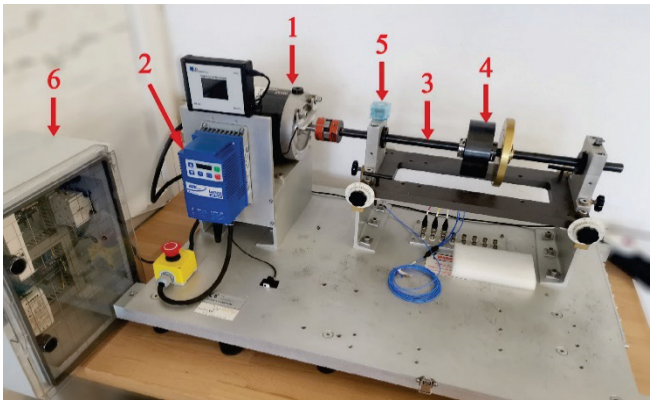


Figure 4 Experimental setup: 1 – Electric motor, 2 – VFD, 3 – main shaft, 4 – main load, 5 – three-axis accelerometer, 6 – Box with router and Raspberry PI edge node.

Table 2 Datasets collected for each rotating speed

Machine states	No. of samples	Training (80%)	Validation (5%)	Test (15%)
NS	1500	1200	75	225
ER	1500	1200	75	225
CR	1500	1200	75	225
DR	1500	1200	75	225
IRB	1500	1200	75	225
ORB	1500	1200	75	225
BB	1500	1200	75	225
CB	1500	1200	75	225
Total	12000	9600	600	1800

The triaxial vibration sensor is installed on the MFS left bearing housing as shown in Fig. 3. During the experiment, data was collected at a frequency of 1.6 kHz. The rotation speed was 150 revolutions per minute and 300 revolutions per minute, respectively. Accelerometer acquired data under both normal state (NS) and faulty conditions. There are three inability states related to rotor: due to eccentric rotor (ER), disbalanced rotor (DR), cocked rotor (CR). Additionally, four bearing related states were simulated as faulty conditions: outer ring damage (ORB), inner ring damage (IRB), ball bearing damage (BB) and combination of the ball bearing, inner ring and outer ring damage (CB), respectively. Bearings is intentionally damaged using electro erosion procedure. The operation in any of the inability state is considered as faulty condition, while running without any simulated faults is interpreted as healthy baseline, i.e. the operation of the rotary subsystem that remains in the healthy state (ability state). 1500 samples for each machine state for both 150 rpm and 300 rpm is collected. Composition of collected samples with quantities is presented in Tab. 2. All the samples are divided into training, test and validation sets required for modelling phase.

4.2 CNN Design

The artificial neural networks (ANN) used in this work as an algorithm for creating models are convolutional neural networks (CNN). It is a type of ANN that has also been used in other research to learn models based on inputs that are composed of multiple dimensions, and on the basis of which features can be learned. Although the initial application of the algorithm was envisioned in image processing, it was found that data from sensors in the maintenance area can also be interpreted using CNN-learned models [20-22].

In this paper, modified 1D MC-DCNN visualized in Fig. 5 is used to for model inability states modelling. Data from triaxial accelerometer is used as an input thus input layer of the 1D MC-DCNN is prepared to process exactly one sample of acquired data. As each sample consist of 1600 samples in each layer, input layer use 1600 × 1 × 3 structure. Convolutional layers of the CNN calculate the output of the neurons. Max pooling function is used to pass over sections (pools) of accelerometer data and extract maximum values of each section. Finally, fully connected layer is used as sample classifier using SoftMax activation function and classification output layer. The structure of the CNN used in this paper is drawn in Fig. 5. Although different CNN structures have been investigated in authors' previous research, it is concluded that usage of described structure is appropriate for learning of the model.

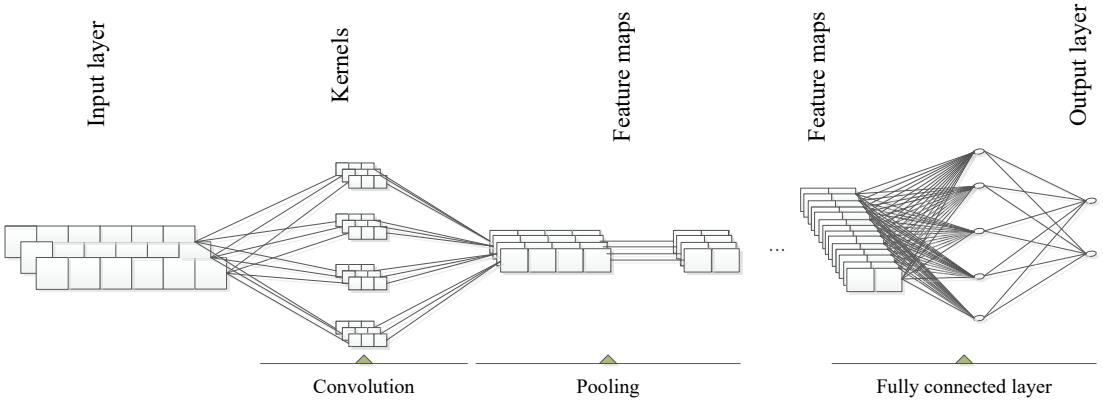


Figure 5 Structure of MC-DCNN

To ensure better results, authors tested grid of hyperparameters values by using variable k that connects multiple hyperparameters. For this research, grid of k = [2 4 8 16] is used. The CNN learning algorithm can be adjusted using a number of hyperparameters, and in this paper, the change in the number of kernels as well as the size of the kernels was tested. Hyperparameters that will be grid searched are number of convolutional layers, number of kernels, and kernel size, respectively. Details about each convolutional layer are given in Tab. 3. From the table, it can be concluded that every convolutional layer use k variable for the calculation of the number of kernels. In this paper, Matlab 2022b is used to design, train and test modified 1D MC-DCNN and all training is done on GPU (GeForce RTX 3080 graphics card).

The Tab. 3 shows that as k increases, so does the number of kernels in each convolutional layer, but also their size in the first two layers. The size of the cores does not change in the last 4 layers, which is marked italic. The paper examined neural networks with one to six convolutional layers. The procedure proceeded so that first a CNN with all 6 layers is learned, then the 6th layer is removed in order to learn a network with 5

layers. After that, the 5th layer is removed to train a 4-layer network. The last network that is learned is a network with only one convolutional layer. Each of those 6 neural networks is trained with 4 values of k (2, 4, 8, 16), leading to a total of 24 neural networks for one rotation speed. The number of 3 learning repetitions of each network is chosen, in order to obtain as few deviations as possible. Then the average accuracy of those 3 networks is calculated, which means that 72 convolutional neural networks were learned for one rotation speed. As the aim of the paper is to examine the results at two rotation speeds (150 rpm and 300 rpm), this leads to a final number of 144 learned convolutional neural networks. Other hyperparameters remain constant during the entire training period.

Table 3 Convolutional neural network layers information regarding *n*-factor

$k = 2$						
Convolutional layer	1st	2nd	3rd	4th	5th	6th
No. of kernels	k	$4k$	$4k$	$4k$	$4k$	$4k$
Kernel size	[2*n* 1 3]	[*n*/2 1 3]	[4 1 3]	[4 1 3]	[4 1 3]	[4 1 3]
$k = 4$						
Convolutional layer	1st	2nd	3rd	4th	5th	6th
No. of kernels	k	$4k$	$4k$	$4k$	$4k$	$4k$
Kernel size	[4 1 3]	[1 1 3]	[4 1 3]	[4 1 3]	[4 1 3]	[4 1 3]
$k = 8$						
Convolutional layer	1st	2nd	3rd	4th	5th	6th
No. of kernels	k	$4k$	$4k$	$4k$	$4k$	$4k$
Kernel size	[8 1 3]	[2 1 3]	[4 1 3]	[4 1 3]	[4 1 3]	[4 1 3]
$k = 8$						
Convolutional layer	1st	2nd	3rd	4th	5th	6th
No. of kernels	k	$4k$	$4k$	$4k$	$4k$	$4k$
Kernel size	[16 1 3]	[4 1 3]	[4 1 3]	[4 1 3]	[4 1 3]	[4 1 3]
$k = 16$						
Convolutional layer	1st	2nd	3rd	4th	5th	6th
No. of kernels	k	$4k$	$4k$	$4k$	$4k$	$4k$
Kernel size	[32 1 3]	[8 1 3]	[4 1 3]	[4 1 3]	[4 1 3]	[4 1 3]

5 RESULTS AND DISCUSSION

The learning process of a neural network can be described as a procedure with the aim of determining the weights and biases of neurons that is calculated in iterations. CNN adjusts learnable parameters by minimizing previously defined loss function:

$$E = -\sum_t \sum_k y_k^*(t) \log(y_k(t)) \quad (1)$$

where $y_k^*(t)$ and $y_k(t)$ are the target and predicted values of the t^{th} training example of the k^{th} class, respectively. To ensure learning capability, this work apply backpropagation algorithm that compute stochastic gradient descent in order to minimize error, consequently allowing network to update learnable parameters during training. In this research, learning rate of 0.009 with drop factor of 0.1 for each 10 iteration and momentum of 0.8 were used.

As previously stated, total of 24 neural network were trained. The results of models tested on test set using accuracy as metrics are shown in Tab. 4, Fig. 6 and Fig. 7, respectively.

Table 4 Convolutional neural networks models accuracy

150 rpm	k-factor / accuracy				300 rpm	k-factor / accuracy			
	2	4	8	16		2	4	8	16
Number of layers 1	0,537	0,755	0,696	0,689	Number of layers 1	0,606	0,784	0,805	0,677
2	0,845	0,862	0,934	0,986	2	0,732	0,847	0,822	0,892
3	0,907	0,975	0,987	0,991	3	0,928	0,944	0,959	0,986
4	0,988	0,995	0,999	0,999	4	0,991	0,991	0,995	1,000
5	0,998	0,999	0,999	0,999	5	1,000	1,000	1,000	1,000
6	0,999	1,000	0,999	1,000	6	1,000	1,000	1,000	1,000

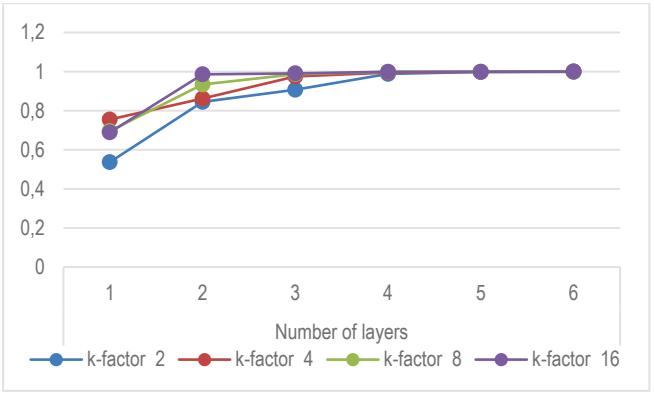


Figure 6 Neural network accuracy related to the number of convolutional layers and k-factor for 150 rpm

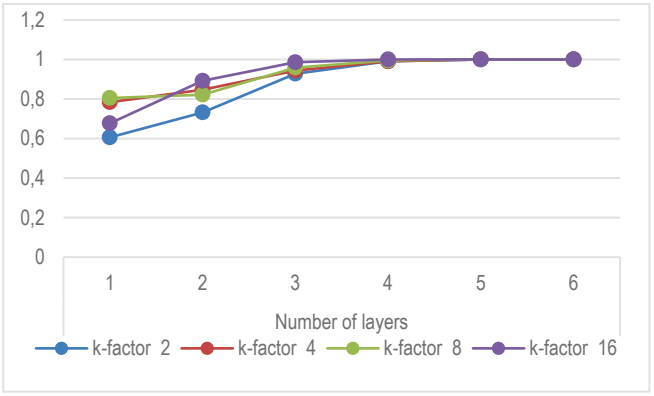


Figure 7 Neural network accuracy related to the number of convolutional layers and k-factor for 300 rpm

It can be seen that most networks with 3 or more convolutional layers would give extremely good results, while absolute accuracy on the test data set can be achieved with CNN with 5 or 6 layers, depending on the rotation speed. From the presented results, it can be concluded that CNN with smaller number of layers would not perform not as good as the network trained with higher number of layers. Other than that, factor k used as kernel size multiplicator plays important role during training of the network with less than 4 convolutional layers. This can be related to the fact that in the experiment this factor is used only in first two layers to confirm that in the shallow architecture higher values of kernel size and number of kernels produces models with better results. As kernels represents features learned during

the training period, it becomes clear that more features learned means better performance during the scoring phase. As data are collected for two different rotational speed, in the last phase of the research, the applicability of the network model learned at a lower rotation speed was tested on test data collected at a higher rotation speed. Namely, due to the large amount of rotary equipment that rotates at different speeds in the industry, the good performance of networks learned at one rotation speed on test data collected at another rotation speed would give reason to believe that one learned model can be applied for intelligent inability stete prediction of the same type of equipment, and which rotates at different speeds. Factor $k = 4$ is chosen for comparison and all networks regarding CNN number of layers factor is tested. Firstly, network learned on 150 rpm with best results for the $k = 4$ (named k4_150) for each CNN number of layers is tested on 300 rpm test data. Later on, network learned on 300 rpm with best results for the $k = 4$ (named k4_300) for each CNN number of layers is tested on 150 rpm test data. Results of such testing is shown in Tab. 5.

Table 5 Convolutional neural networks models accuracy

CNN layers	k4_150 on 150 rpm	k4_150 on 300 rpm	k4_300 on 300 rpm	k4_300 on 150 rpm
1	76,9 %	23,5 %	81,7 %	17,3 %
2	87,4 %	24,9 %	83,5 %	12,5 %
3	97,5%	20,3 %	98,1 %	15,8 %
4	99,5 %	23,5 %	99,8 %	12,5 %
5	100 %	17,7 %	100 %	12,5 %
6	100 %	12,5 %	100 %	12,5 %

According to Tab. 5, models learned for one rotational speed underperforms when tested on different rotational speed. Consequently, experiment show that using models trained on one set of data for prediction of machine state on different set of data is not recommended, although only rotational speed is changed as experiment input. Two another interesting rules are readable from Tab. 5. Model trained on lower speed performance are better on higher speed test then performance of the model trained on higher performance when tested on lower speed test data. Also, in this experiment models with fewer layers generally perform better which means they can learn more general features of the data they are learning on.

6 CONCLUSION AND FUTURE WORK

Finally, results of the paper can be summarized as follows: It is possible to acquire datasets containing raw accelerometer signal using the low-cost self-developed IIOT platform. It is shown that developed solution can acquire data in a laboratory environment.

Additionally, it is proven that IIOT system vibration data acquired during the experiment are useful for learning intelligent models for fault detection and diagnostics. That is shown using 1D-MDCNN, models with high accuracy on the test data can be trained. Multiple CNN-s are trained using the data acquired from the IIOT. The influence of different amounts of certain hyperparameters (number of kernels, kernel size, and number of layers) on the accuracy of the network was examined. Based on these results, it was concluded that increasing the number and size of kernels and the number of layers in the network contributes to increasing the accuracy of the model in the scoring phase.

Finally, application of CNN model learned at one rotation speed to predict classes of data collected at another rotation speed was also tested. The results obtained lead to the conclusion that such an application is not possible. Therefore, the question arises of the applicability of this technique of predicting results in production conditions in case of the variable rotational speeds.

In future research, authors plan to integrate different sensor fusion to enable more data, as well as develop specific purpose edge node with integrated prescriptive maintenance decision making model. In this way, it will be possible to process data directly on the edge, reducing the need of sending raw signal data to the server for further processing and prediction logic.

Acknowledgements

The Article Processing Charges (APCs) is funded by the European Regional Development Fund, grant number KK.01.1.1.07.0031.

7 REFERENCES

[1] Singh, S., Khamba J. S., & Singh, D. (2021). Analyzing the Role of Six Big Losses in OEE to Enhance the Performance: Literature Review and Directions. *Advances in Industrial and Production Engineering*, Phanden, R. K., Mathiyazhagan, K., Kumar, R., & Paulo Davim, J. Eds., in *Lecture Notes in Mechanical Engineering*. Singapore: Springer Singapore, 411-421. https://doi.org/10.1007/978-981-33-4320-7_37

[2] Ompusunggu, A. P., Eryılmaz, K., & Janssen, K. (2021). Condition monitoring of critical industrial assets using high performing low-cost MEMS accelerometers. *Procedia CIRP, 104*, 1389-1394. https://doi.org/10.1016/j.procir.2021.11.234

[3] A. Khademi, F. Raji, & M. Sadeghi, (2019). IoT Enabled Vibration Monitoring Toward Smart Maintenance. *The 3rd IEEE International Conference on Internet of Things and Applications (IoT)*, Isfahan, Iran, 1-6. https://doi.org/10.1109/IICITA.2019.8808837

[4] Xu, L. D., He, W., & Li, S. (2014). Internet of Things in Industries: A Survey. *IEEE Trans. Ind. Inform., 10*(4), 2233-2243. https://doi.org/10.1109/TII.2014.2300753

[5] Li, X., Li, D., Wan, J., Vasilakos, A. V., Lai, C.-F., & Wang, S. (2017). A review of industrial wireless networks in the context of Industry 4.0. *Wirel. Netw., 23*(1), 23-41. https://doi.org/10.1007/s11276-015-1133-7

[6] Tiboni, M., Remino, C., Bussola, R., & Amici, C. (2022). A Review on Vibration-Based Condition Monitoring of Rotating Machinery. *Appl. Sci., 12*(3), p. 972. https://doi.org/10.3390/app12030972

[7] Pech, M., Vrchota, J., & Bednář, J. (2021). Predictive Maintenance and Intelligent Sensors in Smart Factory: Review. *Sensors, 21*(4), 1470. https://doi.org/10.3390/s21041470

[8] Ompusunggu, A. P., Eryılmaz, K., & Janssen, K. (2021). Condition monitoring of critical industrial assets using high performing low-cost MEMS accelerometers. *Procedia CIRP, 104*, 1389-1394. https://doi.org/10.1016/j.procir.2021.11.234

[9] Magadán, L., Suárez, F. J., Granda, J. C., & García, D. F. (2020). Low-cost real-time monitoring of electric motors for the Industry 4.0. Procedia Manufacturing, 42, 393-398. https://doi.org/10.1016/j.promfg.2020.02.057
[10] Al-Naggar, Y. M., Jamil, N., Hassan, M. F., & Yusoff, A. R. (2021). Condition monitoring based on IoT for predictive maintenance of CNC machines. *Procedia CIRP, 102*, 314-318. https://doi.org/10.1016/j.procir.2021.09.054
[11] Lekic, M. & Gardasevic, G. (2018). IoT sensor integration to Node-RED platform. *The 17th IEEE International Symposium INFOTEH-JAHORINA (INFOTEH)*, East Sarajevo, 1-5. https://doi.org/10.1109/INFOTEH.2018.8345544
[12] Curman, M., Kolar, D., Lisjak, D., & Opetuk, T. (2021). Automated and Controlled Data Collection Using Industrial IoT System for Smart Maintenance. *Tehicki Glasnik, 15*(3), 401.409. https://doi.org/10.31803/tg-20210728122543
[13] Kolar, D., Lisjak, D., Curman, M., & Pająk, M. (2022) Condition Monitoring of Rotary Machinery Using Industrial IoT Framework: Step to Smart Maintenance. *Tehicki Glasnik, 16*(3), 343-352. https://doi.org/10.31803/tg-20220517173151
[14] Ince, T., Kiranyaz, S., Eren, L., Askar, M., & Gabbouj, M. (2016). Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks. *IEEE Trans. Ind. Electron., 63*(11), 7067-7075. https://doi.org/10.1109/TIE.2016.2582729
[15] Shao, S., Sun, W., Wang, P., Gao, R. X., & Yan, R. (2016). Learning features from vibration signals for induction motor fault diagnosis. *International Symposium on Flexible Automation (ISFA)*, Cleveland, OH, USA, 71-76. https://doi.org/10.1109/ISFA.2016.7790138
[16] Eren, L., Ince, T., & Kiranyaz, S. (2019). A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier. *J. Signal Process. Syst., 91*(2), 179-189. https://doi.org/10.1007/s11265-018-1378-3
[17] Kolar, D., Lisjak, D., Pająk, M., & Pavković, D. (2020). Fault Diagnosis of Rotary Machines Using Deep Convolutional Neural Network with Wide Three Axis Vibration Signal Input. *Sensors, 20*(14), p. 4017. https://doi.org/10.3390/s20144017
[18] Nemlaha, E., Střelec, P., Horák, T., Kováč, S., & Tanuška, P. (2023). Suitability of MQTT and REST Communication Protocols for AIoT or IIoT Devices Based on ESP32 S3. *Software Engineering Application in Systems Design*, Silhavy, R., Silhavy, P., & Prokopova, Z. Eds., in *Lecture Notes in Networks and Systems, 596*. Cham: Springer International Publishing, 225-233. https://doi.org/10.1007/978-3-031-21435-6_19
[19] https://www.tinkerforge.com/en/doc/Hardware/Bricks/HAT_Brick.html (Accessed Apr. 04, 2022).
[20] Shaheryar, A., Yin, X.-C., & Yousuf, W. (2017). Robust Feature Extraction on Vibration Data under Deep-Learning Framework: An Application for Fault Identification in Rotary Machines. *Int. J. Comput. Appl., 167*(4), 37-45. https://doi.org/10.5120/ijca2017914249
[21] Bagave, P., Linssen, J., Teeuw, W., Brinke, J. K., & Meratnia, N. (2019). Channel State Information (CSI) analysis for Predictive Maintenance using Convolutional Neural Network (CNN). *Proceedings of the 2nd Workshop on Data Acquisition to Analysis*, New York NY USA: ACM, 51-56. https://doi.org/10.1145/3359427.3361917
[22] Silva, W. & Capretz, M. (2019). Assets Predictive Maintenance Using Convolutional Neural Networks. *The 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Toyama, Japan, 59-66. https://doi.org/10.1109/SNPD.2019.8935752

Authors' contacts:

Davor Kolar, PhD
(Corresponding author)
Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Ivana Lučića 5, 10002 Zagreb, Croatia
davor.kolar@fsb.hr

Dragutin Lisjak, Prof. PhD
Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Ivana Lučića 5, 10002 Zagreb, Croatia
dragutin.lisjak@fsb.hr

Martin Curman, MEng
Klimaoprema d.d., Gradna 78A, 10430 Samobor, Croatia
martin.curman@klimaoprema.com

Juraj Benic, PhD
Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb,
Ivana Lučića 5, 10002 Zagreb, Croatia
juraj.benic@fsb.hr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20221206210933

Subject review
Received: 2022-12-06, Accepted: 2023-01-12

Comparative Study of Conventional and Microwave-Assisted Boriding of AISI 1040 and AISI 4140 Steels

Safiye İpek Ayvaz, Emre Özer*

Abstract: In this study, AISI 1040 and AISI 4140 steels were boriding using Ekabor-II commercial boriding powder with powder-pack boriding method using microwave and conventional heating methods. The samples were borided at 950 °C for 2 and 6 hours in an Ar atmosphere in a microwave oven of Enerzi-Mh2912-V8. Biphasic structure (FeB/Fe_2B) was formed in all borided AISI 4140 samples and AISI 1040 samples borided for 6 hours. A single-phase structure was observed in AISI 1040 steel borided for 2 hours. Compared to the conventional method, a 1.5-1.6 times thicker boride layer was obtained in AISI 4140 and AISI 1040 steels with microwave-assisted powder-pack boriding. The highest hardness was measured as 1561.8 $HV_{0.05}$ for boriding AISI 4140 steel and 1499.7 $HV_{0.05}$ for boriding AISI 1040 steel. The Vickers indentation fracture toughness of borided steels with microwave energy varied between 2.31 and 3.46 MPa·$m^{1/2}$. It was determined that in all samples borided by the microwave-assisted and conventional powder-pack boriding method, the adhesion strength between the boride layers and the substrate obtained was sufficient.

Keywords: Fe_2B; FeB; fracture toughness; microwave-assisted; powder-pack boriding; steel

1 INTRODUCTION

Boriding is a diffusion-controlled surface hardening process in which boride layers are obtained on the surface by creating a chemical reaction between boron atoms and metal substrate [1, 2]. It is achieved by diffusing boron atoms at high temperatures on the metal substrate in order to increase the wear resistance, surface hardness and corrosion resistance [3, 4]. Boriding, mainly applied to steel and iron alloys, can be used for various surface hardening and heat-treated steels, non-ferrous metal alloys such as titanium, cobalt, nickel, and ceramic materials [5]. With boriding, TiB and TiB_2 in Ti and its alloys, CoB and Co_2B in CoCr alloys and NiB, Ni_2B, and Ni_4B_3 borides are formed in Ni alloys [6-10].

In steel and alloys, the boriding process, generally applied in the temperature range of 800-1050 °C for 0.5-10 hours, is carried out in solid powder, liquid, paste and gas environments. As a result, FeB and Fe_2B hard boride layers are obtained [5, 11]. Apart from these methods, ion deposition, physical and chemical vapor deposition (PVD, CVD), plasma paste and laser methods are also used in order to increase the practicality of layer formation with the developing technology, as well as increase the mechanical properties [12-14]. Powder-pack boriding, one of the boriding varieties, is the most preferred method due to its advantages such as easy application, low cost and changeability of the boriding composition compared to other methods.

Diffusion kinetics of boron [15-17], mechanical [18-20], tribological properties [21-23] and corrosion resistance [24, 25] of borided steels have been investigated in the literature. Uslu et al. borided AISI 1040 steel using Ekabor 2 powder at temperatures of 800, 875 and 950 °C for 2, 4, 6 and 8 hours of soaking times. By the boriding process, they obtained boride layers consisting of FeB and Fe_2B phases, with a hardness of approximately 1500 HVN and a thickness of 10-180 μm [26]. Özerkan borided AISI 1040 steel using Ekabor 2 powder at temperatures of 850 and 950 °C for 2, 4 and 6 h of soaking times. In this study, he examined the microstructural and mechanical properties of the boride layers. It was reported that the obtained thicknesses of the boride layers were 123.4 and 156.3 μm, respectively, at 850 and 950 °C [27]. Ulutan et al. borided AISI 4140 steel using Ekabor 2 powder at temperatures of 900, 950, 1000 and 1050 °C for 2, 4 and 6 hours of soaking times. In the study, the microstructural and tribological properties of boride layers were investigated. They reported a 290 μm boride layer consisting of FeB, Fe_2B and CrB phases was obtained [28]. Keddam et al. borided AISI 4140 steel at 850 and 1000 °C for 2, 4, 6 and 8 hours. They examined the diffusion kinetics of the boron. They determined the boron activation energy as 189.24 kJ·mol^{-1} in AISI 41240 steel [29]. Dominguez et al. borided AISI 4140 steel at temperatures of 850, 900, 950 and 1000 °C for 2, 4, 6 and 8 h. In this study, they found the boron activation energy for the Fe_2B phase as 173 kJ·mol^{-1}. Experimentally, the maximum Fe_2B thickness obtained was 127.9 μm in the sample borided at 1000 °C for 8 h [30]. Çalık et al. borided AISI 316, AISI 1040, AISI 1045, and AISI 4140 steels at 937 °C for 4 hours. They compared the mechanical properties of these steels after boriding. They reported the boride layer thickness as 90 μm in AISI 1040 steel and 100 μm in AISI 4140 steel [31].

In this study, microwave energy was performed powder-pack boriding of AISI 1040 and AISI 4140 steels for the first time in the literature. After boriding and thickness measurements were carried out, some mechanical and microstructural properties were investigated.

2 MATERIAL AND METHOD

In this study, the mechanical and microstructural properties of AISI 1040 and AISI 4140 steels, which are pack-borided using conventional and microwave heating, were investigated. AISI 1040 and AISI 4140 samples were cut by a cut-off machine. Before boriding, the surfaces of the cut samples were cleaned in an ultrasonic bath in acetone

before. As the boriding pack, AISI 304L stainless steel boxes of 50 mm inner diameter and 50 mm height were used. Ekabor II boriding powders with 5% B_4C, 5% KBF_4 and 90% SiC were used as boriding powder. Microwave-assisted boriding process was carried out at 950 °C for 2 and 6 hours in an Ar atmosphere in an Enerzi-Mh2912-V8 brand microwave oven with a power of 2.9 kW and a frequency of 2.45 GHz (Fig. 1).

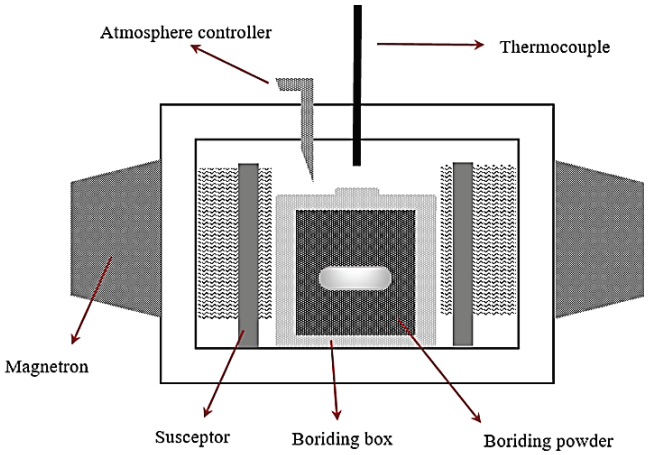


Figure 1 Schematic diagram of the microwave-assisted powder-pack boriding process

After conventional and microwave-assisted boriding, the samples were sanded using 180, 320, 600, 800, 1000 and 1200 grid abrasives and polished using 3 and 1 µm diamond polishing solutions by the Metkon Farcipol 1V brand grinder. After polishing, the samples were etched in a 5% solution of HNO_3 in ethanol (Nital etching reagent). Boride layer thicknesses of etched samples were measured with Nicon Eclipse LV150N brand optical microscope (OM) and Clemex image analysis system. Measurements were repeated eight times for each sample and the average of these measurements was given. Microhardness measurements of the borided samples were performed using the Future-Tech FM 700 brand microhardness measurement device from the cross-section of the sample. In the measurements, 50 gf was applied at an indenter approach speed of 10 µm/sec for 10 seconds. In this study, the equation used to calculate the fracture toughness is as follows:

$$K_{IC} = 0.016 \cdot \frac{P}{c^{1.5}} \cdot \left(\frac{E}{H}\right)^{0.5} \quad (1)$$

Here, E is the elastic modulus of the layer (kg/mm^2) whose fracture toughness is measured, H is the Vickers hardness (HV), P is the applied load (N) during indentation, and c (mm) is the crack length. SEM image of the mark and crack formed on the sample with the Vickers indentation technique is given in Fig. 2.

The adhesion strength between the substrate material and the boride layer in the borided samples was measured following VDI 3198 indentation test standards with a 120° conical indenter. The adhesion behavior was determined by examining cracks, delaminations and fractures in the craters. Microstructural investigations and determination of the adhesion strength of the boride layer were performed using the ZEISS GeminiSEM 500 model (ZEISS, Oberkochen, Germany) scanning electron microscopy (SEM).

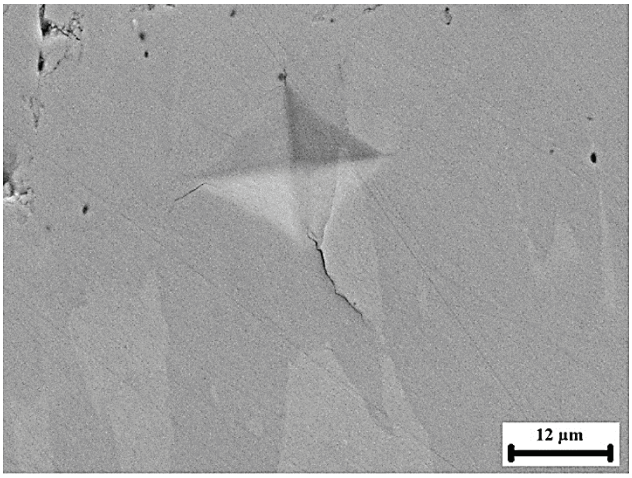


Figure 2 SEM image of the Vickers indentation mark and crack

3 RESULTS AND DISCUSSION

In Figs. 3 and 4, the SEM images of the boride layers of the conventional and microwave-assisted borided samples are given, respectively. While α-Fe has a body-centered cubic structure, neither FeB nor Fe_2B compounds have cubic crystal symmetry [32]. Therefore, B diffusion and boride grain growth exhibit an unnatural anisotropic structure. The preferred growth direction of Fe_2B and FeB compounds is along the 001 direction because B atoms diffuse faster in this direction. As a result, boride grains in the 001 direction perpendicular to the sample surface grow more quickly [20, 33, 34]. Therefore, a boride layer with needle/tooth-like morphology is formed. Saw-tooth morphology also increases the adhesion strength between the substrate and the boride layer [35, 36]. This formation depends on boriding processes such as temperature and time and material properties such as the ratio of the alloy elements [37]. As can be seen in Figs. 3 and 4, boride layers were formed in a saw-tooth form in all samples. As seen in Fig. 3, as the boriding time increased with conventional heating, pores were formed on the surface of the boride layer, especially in AISI 1040 steel. In microwave-assisted boriding, it was determined that both AISI 1040 and AISI 4140 borided steel samples had much less porosity and surface roughness. Therefore a much better surface quality could be obtained with microwave-assisted boriding (Fig. 4).

Due to the high brittleness of the FeB phase, only Fe_2B single-phase structure is preferred after boriding in the steels [38]. Boride layers consisting of the Fe_2B phase were obtained in both AISI 1040 and AISI 4140 steels in the boriding process performed with conventional heating (Figs. 3a-d). Dual-phase structure (dark-colored FeB compound, lighter-colored Fe_2B compound) was observed in microwave-assisted borided AISI 1040 steel for 6 hours (Fig.

4b). A single-phase (Fe_2B) structure was formed in the AISI 1040 steel sample, borided for 2 hours with microwave-assisted heating (Fig. 4a). In AISI 4140 steel, biphasic boride layers were formed as a result of microwave-assisted boriding for both 2 and 6 hours.

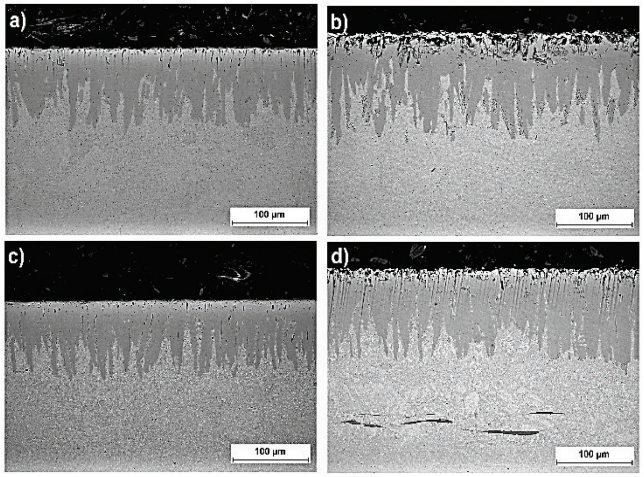


Figure 3 SEM images of the conventional borided samples: a) AISI 1040-2 h, b) AISI 1040-6 h, c) AISI 4140-2 h and d) AISI 4140-6 h

The elemental analysis results obtained from the boride layer and substrate formed in AISI 1040 steel, which was borided for 2 h with conventional heating, are given in Fig. 5. As determined from the EDS analysis, the boride layer was composed of the Fe_2B phase. The change in peak intensity obtained during the linear EDS scanning of B atoms in microwave-assisted borided AISI 4140 for 2 h is seen in Fig. 6. C atoms, like Si atoms, do not dissolve in the boride layer and are pushed toward the substrate during the diffusion of B atoms. In Fig. 6, while C atoms gave a very low peak intensity in the boride layer, it was seen that the peak intensity increased as the scanning went under the layer.

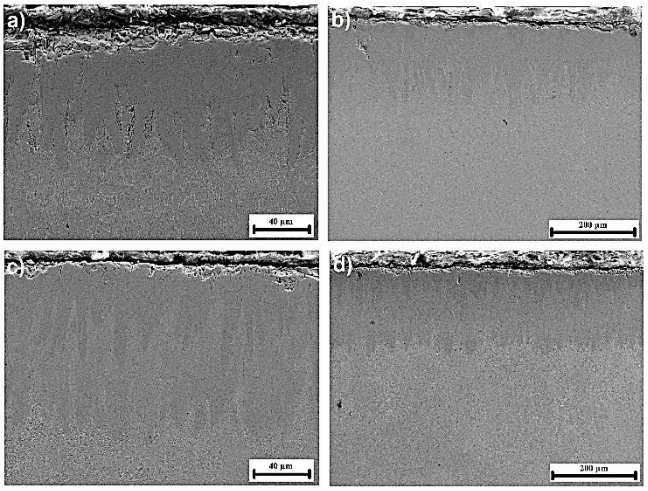


Figure 4 SEM images of the microwave-assisted borided samples: a) AISI 1040-2 h, b) AISI 1040-6 h, c) AISI 4140-2 h and d) AISI 4140-6 h

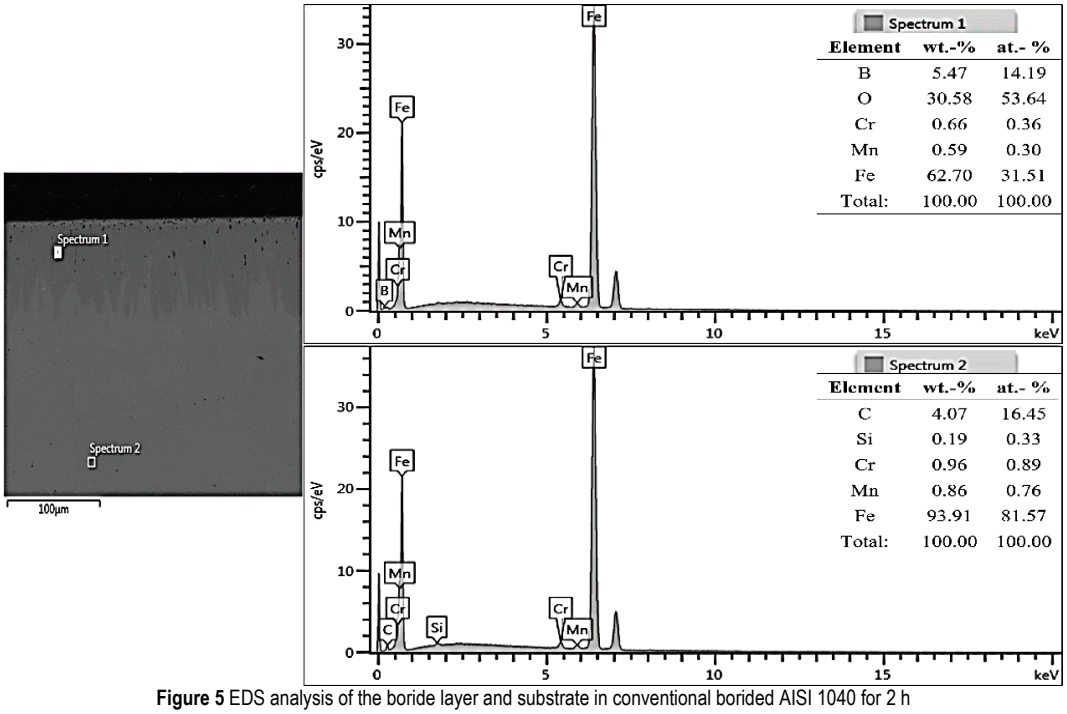


Figure 5 EDS analysis of the boride layer and substrate in conventional borided AISI 1040 for 2 h

Boride layer thicknesses measured in conventional and microwave-assisted powder-pack borided samples are graphically given in Fig. 7. Microwave heating is non-thermal heating. Here, the microwave energy absorbed by the material is transformed into thermal energy [39, 40]. With this energy conversion, rapid volumetric heating occurs. İpek Ayvaz and Aydın [5, 41] borided AISI 316L stainless steel with a microwave-assisted powder-pack boriding method. In this study, two times the thickness of boride was obtained according to the conventional method. In the diffusion kinetics analysis, it was determined that microwave heating increased the pre-exponential factor, and as a result, boron

diffusion increased. When the thickness values obtained in this study are compared with the literature summarized in Tab. 1, it can be seen that the thickness of the boride layer is significantly increased by microwave heating. In particular, the thicknesses obtained in microwave-assisted powder-pack borided AISI 4140 steel are approximately 1.5-2 times the thicknesses in the literature. In this study, it was determined that much thicker boride layers were obtained with microwave-assisted boriding in both AISI 1040 and AISI 4140 steels compared to conventional heating.

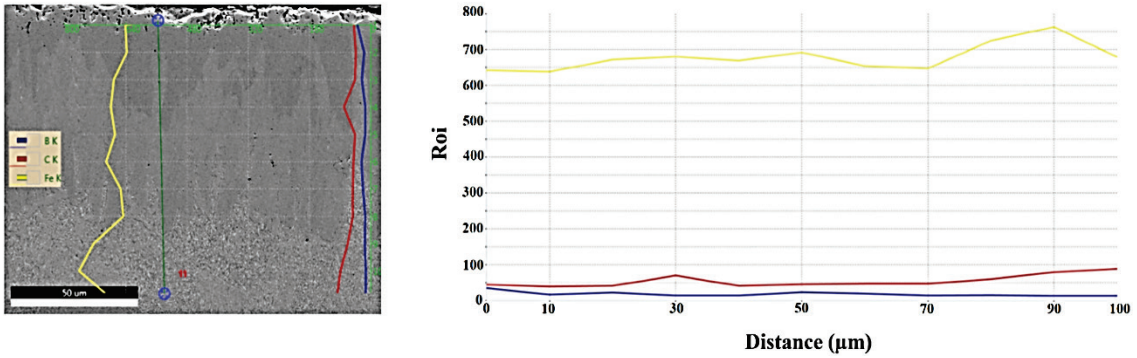


Figure 6 Results of EDS line scan analysis in microwave-assisted borided AISI 4140 for 2 h

Table 1 Boriding parameters applied to AISI 1040 and AISI 4140 steels with a conventional furnace and boriding thicknesses obtained in the literature

Material	Temperature (°C) / Time (h)	Thickness (µm)	References
1040	937 / 4	90-100	[18, 19]
4140	950 / 2	86	[28]
4140	950 / 6	167	[28]
4140	937 / 4	100	[31]
1040	937 / 4	90	[31]
4140	950 / 2	67	[42]
4140	950 / 3	92	[42]
4140	950 / 4	60	[43]
4140	950 / 6	80	[43]

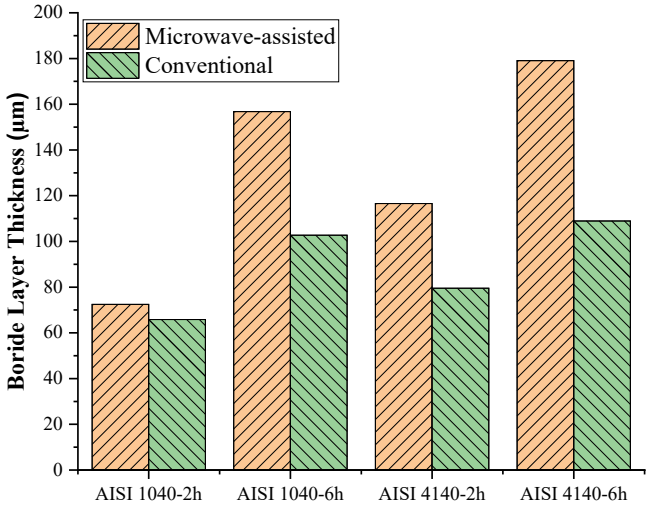


Figure 7 The thickness of boride layers of the conventional and microwave-assisted borided samples

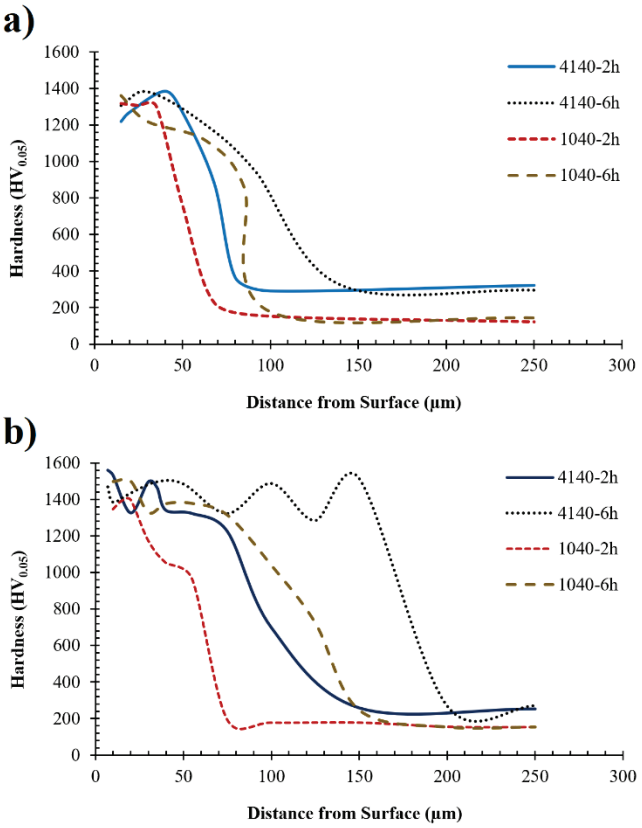


Figure 8 Variation in microhardness as a function of depth from the surface in borided samples a) conventional and b) microwave-assisted

Fig. 8 shows thickness-dependent hardness variations of microwave-assisted and conventional powder-pack borided AISI 4140 and AISI 1040 steels. After the boriding process was performed with the traditional method, the formation of a single-phase Fe_2B layer was detected. With the effect of single-phase structure, sequentially, the maximum microhardness of 1360 and 1384 $HV_{0.05}$ was obtained in AISI 1040 and AISI 4140 steels borided by conventional methods. Uslu et al. [26] obtained a hardness of 1200-1500 HV in borided AISI 1040 steel. Özerkan [27] measured the maximum hardness of 1539.6 HV in borided AISI 1040 steel. In this study, the highest hardness in AISI 1040 steel was measured as approximately 1499 $HV_{0.05}$ in the 10-20 µm depth range from the surface of the microwave-assisted borided sample for 6 hours. Ulutan et al. [28] measured the hardness of borided AISI 4140 steel in the range of 1309-1757 $HV_{0.05}$. Joshi and Hosmani achieved hardness between

1225-1367 $HV_{0.1}$ [42]. Arslan and Akgül Kayral reported the hardness of AISI 4140 steel, which they borided, in the range of approximately 1200-1400 $HV_{0.05}$ [43]. In this study, the highest hardness was measured as 1561 $HV_{0.05}$ in the microwave-borided AISI 4140 sample for 2 hours, and the maximum was 1497 $HV_{0.05}$ in the microwave borided for 6 hours.

Vickers indentation fracture toughness values of borided samples are shown in Fig. 9. When the results given in Fig. 9 are compared, although the fracture toughness of the boride layer obtained with microwave boriding appears to be lower, the crack formation was not observed in the Fe_2B phase in the microwave-borided samples in the tests performed. For these samples, cracks formed in the FeB phase were included in the calculation. However, cracks in the Fe_2B phase in conventionally borided samples were observed. Campos-Silva et al. calculated the fracture toughness of the Fe_2B phase in AISI 4140 steel as 2.9-4.0 $MPa \cdot m^{1/2}$ [20]. Uslu et al. determined the fracture toughness of the boride layers formed in P20 steel in the range of 2.79-4.79 $MPa \cdot m^{1/2}$ [26]. Taktak reported the fracture toughness of boride layers in AISI 304 stainless steel alloy as 2.45-4.08 $MPa \cdot m^{1/2}$ [44]. Campos-Silva et al. determined the fracture toughness of the FeB-Fe_2B interface as 3.56-4.45 $MPa \cdot m^{1/2}$ [45]. In this study, results compatible with the literature were obtained.

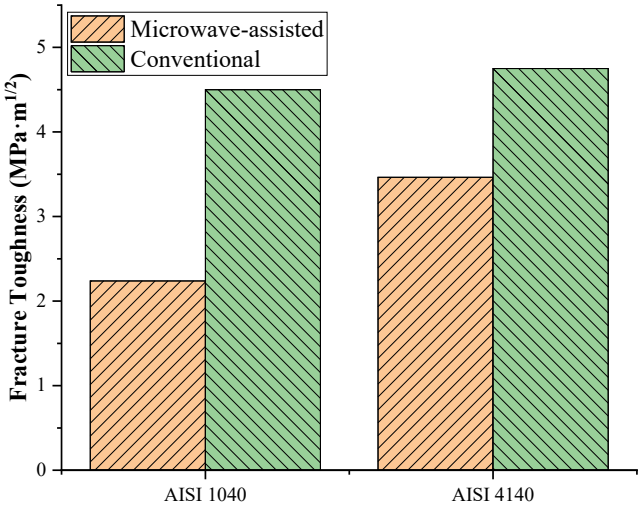


Figure 9 Vickers indentation fracture toughness of the conventional and microwave-assisted borided samples

Figs. 10 and 11 show the SEM images of the craters formed on the surface of the conventional and microwave-assisted borided samples after Daimler-Benz Rockwell-C adhesion tests. Adhesion can be defined as the resistance of a coating to separation or disintegration from the substrate. A standard Rockwell-C hardness tester was used in these tests. In these tests, the damage to the coating was compared with the HF1-HF6 quality map [44, 46, 47], and the adhesion strength between the boride layer and the substrate was evaluated. It is observed that the cratering formed in both microwave-supported and conventionally borided AISI 1040 steels is broader than that of AISI 4140 steel. This situation occurs due to the lower thickness and softer boride layer. This phenomenon can be seen more clearly in microwave-assisted borided samples (Fig. 11). This circumstance was caused by FeB phase formation in microwave-assisted borided AISI 4140 samples. In all samples, sufficient adhesion strength was found between the boride layers and the substrate, meeting the HF1 standard. It is predicted that the increase in the FeB phase may decrease the adhesion strength due to residual thermal stresses [44, 48]. However, it is known that this effect is minimized by hybrid heating that occurs with SiC susceptors in microwave-assisted powder-pack boriding [5, 41].

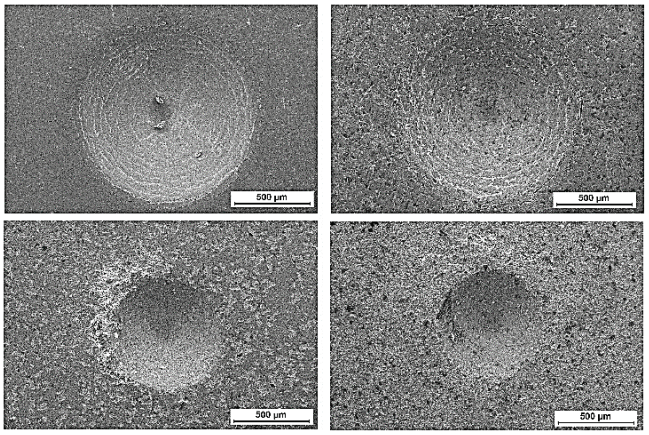


Figure 10 SEM images of craters of the Daimler-Benz adhesion tests on the conventional borided samples: a) AISI 1040-2 h, b) AISI 1040-6 h, c) AISI 4140-2 h and d) AISI 4140-6 h

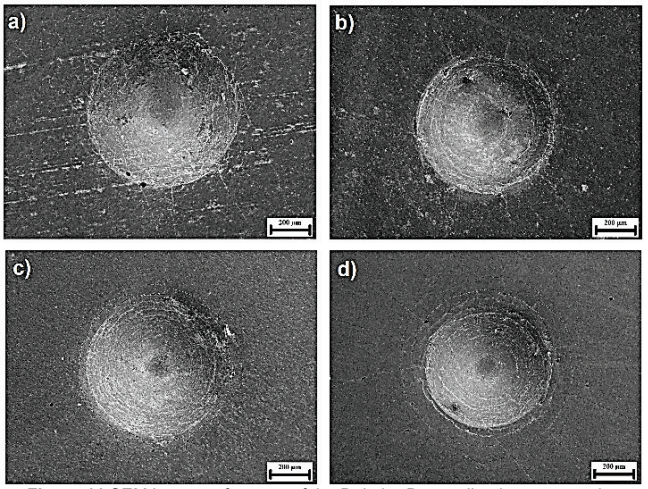


Figure 11 SEM images of craters of the Daimler-Benz adhesion tests on the microwave-assisted borided samples: a) AISI 1040-2 h, b) AISI 1040-6 h, c) AISI 4140-2 h and d) AISI 4140-6 h

4 CONCLUSION

In this study, conventional and microwave-assisted boriding in steel alloys was compared for the first time in the literature. AISI 4140 and AISI 1040 steels were borided using microwave-assisted and conventional powder pack boriding methods. The results obtained after the boriding process was applied at 950 °C for 2 and 6 hours are as follows:

- While a single-phase (Fe_2B) boride layer was observed in all conventional borided samples, a single-phase boride layer was obtained in AISI 1040 steel borided for only 2 hours from microwave-assisted borided samples. In other samples, FeB/Fe_2B biphasic boride layers were formed.
- Boride layer thicknesses were approximately 1.5-2 times thicker than the literature obtained in AISI 4140 steel due to microwave-assisted boriding. These thicknesses were 116.6 and 179.01 μm after boring for 2 and 6 hours, respectively. With microwave-assisted boriding, a ~60% increase in thickness was achieved compared to the conventional method.
- Boride layer thicknesses of 72.48 and 156.8 μm were obtained in AISI 1040 steel, respectively, after 2 and 6 hours of boriding. With microwave-assisted boriding, a ~50% increase in thickness was achieved compared to the conventional method.
- The highest hardnesses obtained in microwave-assisted borided AISI 4140 and AISI 1040 steels are 1561.8 $HV_{0.05}$ and 1499.7 $HV_{0.05}$, respectively.
- According to the results in the literature, fracture toughness was calculated between 2.31 and 3.46 $MPa \cdot m^{1/2}$ in microwave energy borided samples. The fracture toughness of the boride layers in the conventionally borided samples with a boride layer consisting of a single-phase Fe_2B compound is higher than in the microwave-assisted borided ones.
- It was determined that there was sufficient adhesion strength between the boride layers and the substrate in all samples.

5 REFERENCES

[1] Ozdemir, O., Omar, M. A., Usta, M., Zeytin, S., Bindal, C., & Ucisik, A. H. (2009). An investigation on boriding kinetics of AISI 316 stainless steel. *Vacuum*, *83*(1), 175-179. https://doi.org/10.1016/j.vacuum.2008.03.026

[2] Genel, K. (2006). Boriding kinetics of H13 steel. *Vacuum*, *80*(5), 451-457. https://doi.org/10.1016/j.vacuum.2005.07.013

[3] Abdellah, Z. N., Boumaali, B., & Keddam, M. (2021). Experimental evaluation and modelling of the boronizing kinetics of AISI H13 hot work tool steel. *Materials Testing*, *63*(12), 1136-1141. https://doi.org/10.1515/mt-2021-0056

[4] Yu, L. G., Chen, X. J., Khor, K. A., & Sundararajan, G. (2005). FeB/Fe_2B phase transformation during SPS pack-boriding: Boride layer growth kinetics. *Acta Materialia*, *53*(8), 2361-2368. https://doi.org/10.1016/j.actamat.2005.01.043

[5] Ayvaz, S. İ. & Aydın, İ. (2020). Effect of the microwave heating on diffusion kinetics and mechanical properties of borides in AISI 316L. *Transactions of the Indian Institute of Metals*, *73*(10), 2635-2644. https://doi.org/10.1007/s12666-020-02072-x

[6] Boumaali, B., Abdellah, Z. N., & Keddam, M. (2021). Computer simulation of boronizing kinetics for a TB_2 alloy. *Materials Testing*, *63*(12), 1130-1135. https://doi.org/10.1515/mt-2021-0051

[7] Deng, D., Wang, C., Liu, Q., & Niu, T. (2015). Effect of standard heat treatment on microstructure and properties of borided Inconel 718. *Transactions of Nonferrous Metals Society of China*, *25*(2), 437-443. https://doi.org/10.1016/S1003-6326(15)63621-4

[8] Makuch, N., Kulka, M., Keddam, M., Taktak, S., Ataibis, V., & Dziarski, P. (2017). Growth kinetics and some mechanical properties of two-phase boride layers produced on commercially pure titanium during plasma paste boriding. *Thin Solid Film*, *626*, 25-37. https://doi.org/10.1016/j.tsf.2017.02.033

[9] Tsipas, S. A., Vázquez-Alcázar, M. R., Navas, E. M. R., & Gordo, E. (2010). Boride coatings obtained by pack cementation deposited on powder metallurgy and wrought Ti and Ti–6Al–4V. *Surface and Coatings Technology*, *205*(7), 2340-2347. https://doi.org/10.1016/j.surfcoat.2010.09.026

[10] Campos-Silva, I., Bravo-Bárcenas, D., Cimenoglu, H., Figueroa-López, U., Flores-Jiménez, M., & Meydanoglu, O. (2014). The boriding process in CoCrMo alloy: Fracture toughness in cobalt boride coatings. *Surface and Coatings Technology*, *260*, 362-368. https://doi.org/10.1016/j.surfcoat.2014.07.092

[11] Sen, S., Sen, U., & Bindal, C. (2006). Tribological properties of oxidised boride coatings grown on AISI 4140 steel. *Materials Letters*, *60*(29–30), 3481-3486. https://doi.org/10.1016/j.matlet.2006.03.036

[12] Çelikkan, H., Öztürk, M. K., Aydin, H., & Aksu, M. L. (2007). Boriding titanium alloys at lower temperatures using electrochemical methods. *Thin Solid Films*, *515*(13), 5348-5352. https://doi.org/10.1016/j.tsf.2007.01.020

[13] Yoon, J. H., Jee, Y. K., & Lee, S. Y. (1999). Plasma paste boronizing treatment of the stainless steel AISI 304. *Surface and Coatings Technology*, *112*(1-3), 71-75. https://doi.org/10.1016/S0257-8972(98)00743-9

[14] Selvan, J. S., Subramanian, K., Nath, A. K., Kumar, H., Ramachandra, C., & Ravindranathan, S. P. (1999). Laser boronising of Ti–6Al–4V as a result of laser alloying with pre-placed BN. *Materials Science and Engineering: A*, *260*(1-2), 178-187. https://doi.org/10.1016/S0921-5093(98)00964-2

[15] Genel, K., Ozbek, I., & Bindal, C. (2003). Kinetics of boriding of AISI W1 steel. *Materials Science and Engineering: A*, *347*(1-2), 311-314. https://doi.org/10.1016/S0921-5093(02)00607-X

[16] Bektes, M., Calik, A., Ucar, N., & Keddam, M. (2010). Pack-boriding of Fe–Mn binary alloys: Characterization and kinetics of the boride layers. *Materials Characterization*, *61*(2), 233-239. https://doi.org/10.1016/j.matchar.2009.12.005

[17] Ucar, N., Yigit, M., & Calik, A. (2020). Metallurgical characterization and kinetics of borided 34CrNiMo6 steel. *Advances in Materials Science*, *20*(4), 38-48. https://doi.org/10.2478/adms-2020-0021

[18] Calik, A., Sahin, O., & Ucar, N. (2008). Specimen geometry effect on the mechanical properties of AISI 1040 steel. *Zeitschrift für Naturforschung A*, *63*(7-8), 448-452. https://doi.org/10.1515/zna-2008-7-811

[19] Calik, A., Taylan, F., Sahin, O., & Ucar, N. (2009). Comparison of mechanical properties of boronized and vanadium carbide coated AISI 1040 steels. *Indian Journal of Engineering and Materials Sciences*, *16*(5), 326-330.

[20] Campos-Silva, I., Ortiz-Domínguez, M., López-Perrusquia, N., Meneses-Amador, A., Escobar-Galindo, R., & Martinez-Trinidad, J. (2010). Characterization of AISI 4140 borided steels. *Applied Surface Science*, *256*(8), 2372-2379. https://doi.org/10.1016/j.apsusc.2009.10.070

[21] Fernández-Valdés, D., Meneses-Amador, A., López-Liévano, A., & Ocampo-Ramírez, A. (2021). Sliding wear analysis in borided AISI 316L steels. *Materials Letters*, *285*, 129138. https://doi.org/10.1016/j.matlet.2020.129138

[22] García-Léon, R. A., Martínez-Trinidad, J., Campos-Silva, I., Figueroa-López, U., & Guevara-Morales, A. (2021). Wear maps of borided AISI 316L steel under ball-on-flat dry sliding conditions. *Materials Letters*, *282*, 128842. https://doi.org/10.1016/j.matlet.2020.128842

[23] Ficici, F., Kapsiz, M., & Durat, M. (2011). Applications of taguchi design method to study wear behaviour of boronized AISI 1040 steel. *International Journal of the Physical Sciences*, *6*(2), 237-243. https://doi.org/10.5897/IJPS11.009

[24] da Costa Aichholz, S. A., Meruvia, M. S., Júnior, P. C. S., & Torres, R. D. (2018). Tribocorrosion behavior of boronized AISI 4140 steel. *Surface and Coatings Technology*, *352*, 265-272. https://doi.org/10.1016/j.surfcoat.2018.07.101

[25] D'Souza, B., Leong, A., Yang, Q., & Zhang, J. (2021). Corrosion behavior of boronized nickel-based alloys in the molten chloride Salt. *Corrosion Science*, *182*, 109285. https://doi.org/10.1016/j.corsci.2021.109285

[26] Uslu, I., Comert, H., Ipek, M., Celebi, F. G., Ozdemir, O., & Bindal, C. (2007). A comparison of borides formed on AISI 1040 and AISI P20 steels. *Materials and Design*, *28*(6), 1819-1826. https://doi.org/10.1016/j.matdes.2006.04.019

[27] Özerkan, H. B. (2019). Experimental fatigue life determination of thermo-diffusion surface boronized of AISI 1040 steel. *Journal of Mechanical Science and Technology*, *33*(10), 4957-4962. https://doi.org/10.1007/s12206-019-0935-4

[28] Ulutan, M., Yildirim, M. M., Çelik, O. N., & Buytoz, S. (2010). Tribological properties of borided AISI 4140 steel with the powder pack-boriding method. *Tribology Letters*, *38*(3), 231-239. https://doi.org/10.1007/s11249-010-9597-1

[29] Keddam, M., Ortiz-Domínguez, M., Gómez-Vargas, O. A., Arenas-Flores, A., Flores-Rentería, M. Á., Elias-Espinosa, M., & García-Barrientos, A. (2015). Kinetic study and characterization of borided AISI 4140 steel. *Materiali in Tehnologije - Materials and Technology*, *49*(5), 665-672. https://doi.org/10.17222/mit.2014.034

[30] Ortiz-Domínguez, M., Hernandez-Sanchez, E., Martinez-Trinidad, J., Keddam, M., & Campos-Silva, I. (2010). A kinetic model for analyzing the growth kinetics of Fe_2B layers in AISI 4140 steel. *Kovové Materiály - Metallic Materials*, *48*, 285-290. https://doi.org/10.4149/km_2010_5_285

[31] Calik, A., Sahin, O., & Ucar, N. (2009). Mechanical properties of boronized AISI 316, AISI 1040, AISI 1045 and AISI 4140 steels. *Acta Physica Polonica A*, *115*(3), 694-698. https://doi.org/10.12693/aphyspola.115.694

[32] Pomel'nikova, A. S., Shipko, M. N., & Stepovich, M. A. (2011). Features of structural changes due to the formation of the boride crystal structure in steels. *Journal of Surface Investigation: X-ray, Synchrotron and Neutron Techniques*, *5*(2), 298-304. https://doi.org/10.1134/S1027451011030165

[33] Ramdan, R. D., Takaki, T., Yashiro, K., & Tomita, Y. (2010). The effects of structure orientation on the growth of Fe_2B boride by multi-phase-field simulation. *Materials Transactions*, *51*(1), 62-67. https://doi.org/10.2320/matertrans.M2009227

[34] Brakman, C. M., Gommers, A. W. J., & Mittemeijer, E. J. (1989). Boriding of Fe and Fe–C, Fe–Cr, and Fe–Ni alloys; boride-layer growth kinetics. *Journal of Materials Research*, *4*(6), 1354-1370. https://doi.org/10.1557/JMR.1989.1354

[35] İpek, M., Celebi Efe, G., Ozbek, I., Zeytin, S., & Bindal, C. (2012). Investigation of boronizing kinetics of AISI 51100 steel. *Journal of Materials Engineering and Performance*, *21*(5), 733-738. https://doi.org/10.1007/s11665-012-0192-5

[36] Garcia-Bustos, E., Figueroa-Guadarrama, M. A., Rodriguez-Castro, G. A., Gómez-Vargas, O. A., Gallardo-Hernandez, E. A., & Campos-Silva, I. (2013). The wear resistance of boride layers measured by the four-ball test. *Surface and Coatings Technology*, *215*, 241-246. https://doi.org/10.1016/j.surfcoat.2012.08.090

[37] Ozbek, I. & Bindal, C. (2002). Mechanical properties of boronized AISI W4 steel. *Surface and Coatings Technology*, *154*(1), 14-20. https://doi.org/10.1016/S0257-8972(01)01409-8

[38] Ozdemir, O., Usta, M., Bindal, C., & Ucisik, A. H. (2006). Hard iron boride (Fe_2B) on 99.97 wt% pure iron. *Vacuum*, *80*(11-12), 1391-1395. https://doi.org/10.1016/j.vacuum.2006.01.022

[39] Ayvaz, M. (2022). Characterization and tribological properties of novel AlCu4.5SiMg alloy–(B_4C/TiO_2/nGr) quaternary hybrid composites sintered via microwave. *Metals and Materials International*, *28*, 710-721. https://doi.org/10.1007/s12540-020-00894-4

[40] Ayvaz, M. (2021). Microstructure and dry sliding wear behaviors of microwave-sintered Al-4.4Cu-0.7Mg-0.6Si-B_4C/nGr hybrid composites. *Transactions of the Indian Institute of Metals*, *74*(6), 1397-1408. https://doi.org/10.1007/s12666-021-02232-7

[41] Ayvaz, S. İ. & Aydin, İ. (2022). Tribological and adhesion properties of microwave-assisted borided AISI 316L steel. *Materials Testing*, *64*(2), 249-259. https://doi.org/10.1515/mt-2021-2031

[42] Joshi, A. A. & Hosmani, S. S. (2014). Pack-boronizing of AISI 4140 steel: Boronizing mechanism and the role of container design. *Materials and Manufacturing Processes*, *29*(9), 1062-1072. https://doi.org/10.1080/10426914.2014.921705

[43] Arslan, D. & Akgün, S. (2021). Mechanical characterization of pack-boronized AISI 4140 and AISI H13 steels. *International Advanced Researches and Engineering Journal*, *5*(1), 61-71. https://doi.org/10.35860/iarej.817274

[44] Taktak, S. (2007). Some mechanical properties of borided AISI H13 and 304 steels. *Materials and Design*, *28*(6), 1836-1843. https://doi.org/10.1016/j.matdes.2006.04.017

[45] Campos-Silva, I., Martinez-Trinidad, J., Doñu-Ruíz, M. A., Rodríguez-Castro, G., Hernandez-Sanchez, E., & Bravo-Bárcenas, O. (2011). Interfacial indentation test of FeB/Fe_2B coatings. *Surface and Coatings Technology*, *206*(7), 1809-1815. https://doi.org/10.1016/j.surfcoat.2011.08.017

[46] Vera Cárdenas, E. E., Lewis, R., Martínez Pérez, A. I., Bernal Ponce, J. L., Pérez Pinal, F. J., Domínguez, M. O., & Rivera Arreola, E. D. (2016). Characterization and wear performance of boride phases over tool steel substrates. *Advances in Mechanical Engineering*, *8*(2), 1687814016630257. https://doi.org/10.1177/1687814016630257

[47] Vidakis, N., Antoniadis, A., & Bilalis, N. (2003). The VDI 3198 indentation test evaluation of a reliable qualitative control for layered compounds. *Journal of Materials Processing Technology*, *143-144*, 481-485. https://doi.org/10.1016/S0924-0136(03)00300-5

[48] Krelling, A. P., Da Costa, C. E., Milan, J. C. G., & Almeida, E. A. S. (2017). Micro-abrasive wear mechanisms of borided AISI 1020 steel. *Tribology International*, *111*, 234-242. https://doi.org/10.1016/j.triboint.2017.03.017

Authors' contacts:

Safiye İpek Ayvaz, Lect. Dr.
Celal Bayar University,
Turgutlu Vocational School,
Department of Machinery and Metal Technologies,
45140 Manisa, Turkey
s.ipekayvaz@gmail.com

Emre Özer, Asst. Prof. Dr.
(Corresponding author)
Osmaniye Korkut Ata University,
Engineering Faculty,
Department of Industrial Engineering,
80000 Osmaniye, Turkey
mech.eng.emreozer@gmail.com

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230109111604

Subject review
Received: 2023-01-09, Accepted: 2023-01-11

Analysis of the Behavior of a Penetrator Advancing Through a Guide Surface

Jin Bong Kim

Abstract: The study concerns the transverse deformation behavior of a penetrator surrounded by sabot in a deformed gun barrel. In the gun barrel, transverse deformation occurs in the penetrator due to problems such as deflection by gravity, or geometric tolerance caused by the manufacturing process. This deformation causes structural instability problems and affects out-of-gun barrel movement. In addition, the deformation and structural safety of the penetrator is affected by the sabot supporting the penetrator. The finite element method was used to evaluate the effect of the sabot. Deformation and stress analysis were performed for the penetrator moving in the gun barrel, and the effect of the elastic modulus of the sabot on the deformation of the penetrator was studied.

Keywords: elastic modulus; numerical analysis; obturator; penetrator; sabot; von Mises stress

1 INTRODUCTION

Kinetic energy ammunition is being used as a countermeasure against the development of protection systems, such as tanks or light armored vehicles. Kinetic energy ammunition is being developed in the form of wing-stabilized armor-piercing (AP) ammunition using high slenderness ratio (L/D) penetrators to maximize penetration performance [1].

The AP consists of an obturator, fin, and sabot. The sabot is an important part and must withstand hundreds of MPa, and it transmits this pressure to the projectile. Fig. 1 [2] shows a schematic of such a kinetic energy projectile.

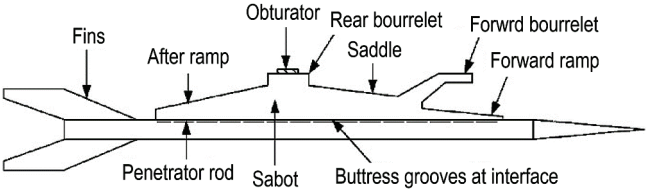


Figure 1 Components of a kinetic energy projectile

The sabot and the projectile both have interface buttress grooves that are used to transmit momentum. The projectile is pushed out of the barrel by high-pressure gas that is produced as the propellant burns on the sabot's rear ramp. A sealing device (seal) is installed on the rear barrel to prevent leakage of hot combustion gases. The projectile is supported throughout moving inside the barrel by a forward bourrelet, and the forward ramp is also utilized during the initial phases of separation to avoid separation failure brought on by obstruction behavior. Along the axial plane, various components make up the sabot. When the assembly of the projectile and the sabot exits the barrel, the sabot is no longer needed, and falls away.

Factors that affect the dispersion of AP [3] are uneven muzzle velocity, unevenness in warhead manufacturing, weather conditions, and the movement posture of flying bullets. In the case of wing-stabilized AP ammunition, the sabot is designed to be separated from the penetrator at the moment the warhead leaves the muzzle, so the function in the separation phase [2] will affect the dispersion of the bullet.

In AP, when the assembly of the projectile and the sabot exits the barrel, the sabot is not anymore required. Therefore, the separation of components, which must be done as quickly as possible, usually occurs in a barrel exit detonation. The most important consideration when designing a sabot that separates quickly is the balance between the aerodynamic force and the dispersion of the projectile [4] needed to separate the sabot quickly without compromising its ability to transmit force. The amount of destruction operating on the projectile during launch and free flight should be kept to a minimum for increased dispersion accuracy. To improve penetration, it is necessary to minimize aerodynamic drag [5], and move at a faster speed.

The existing sabot uses two bore riders [6, 7], a front part and a rear part, as shown in Fig. 1, so that the movement of the projectile is efficient with a high L/D ratio inside the bore [8]. The front bore rider of these two bore riders acts as an aerodynamic lifting surface to help the sabot disengage. Rotation of the sabot causes mechanical contact between the projectile and the sabot because the majority of the lift force is centered at the forward end of the sabot [9]. In general, asymmetric mechanical contact can make high yaw/pitch motion in the projectile [10], resulting in a loss of firing propulsion accuracy. Therefore, conventional sabots are known to have intrinsic problems of automated communication.

To overcome the shortcomings of the existing sabot, a solution to prevent non-uniform detachment due to pressure fluctuations of the sabot front bore rider has been proposed [11]. This design substitutes a web member for the front-bore rider, which also serves as a support to lessen turbulence in the entering airflow and lessen projectile impact.

If the projectile that is fired and progresses within the barrel vibrates [12] in the transverse direction, the firing condition of the projectile becomes unstable. The dynamic behavior of the projectile also becomes unstable, thereby reducing the accuracy of the projectile performance. The effectiveness of the projectile [13] is determined by the accuracy and consistency of hitting the target. The transverse

behavior depends on the difference between the centerline of the barrel and the projectile. The symmetry of the propulsion pressure applied to the projectile also affects the transverse behavior.

Studies have been made of the effect of the barrel deformation on the projectile at the barrel exit [14, 15], and the effect of the bore-rider stiffness of the sabot on the propulsion behavior of the warhead [16, 17]. Such changes in the barrel centerline include vibration, non-uniformity of cooling, manufacturing defects, and non-uniformity of barrel wall thickness. By analyzing these factors and minimizing the occurrence factors, an optimal weapon system can be developed. Analysis of the transverse motion of the projectile is important in the development of the projectile [18].

Sabot prediction methodologies for projectile systems have been studied for several decades [19, 20]. The design of the sabot relies heavily on various types of propulsion systems, and it is performed in four forms: supporting the projectile during launch acceleration, guiding the projectile along the center of the barrel, sealing the encapsulation in high-pressure propellant gas, and smooth separation after disengagement from the barrel. The length of the sabot supporting the kinetic energy projectile [21] is an essential factor affecting structural integrity

As it is usually undesirable to modify the gun system to improve barrel exit speed and strike accuracy, it is preferable to modify the configuration of the sabot during the design of the gun system/projectile, and to select a sabot medium with a proper strength-to-weight relation. Regarding the formation changes, the structural integrity of the sabot cannot be cooperated, and designing a sabot with less weight and better structural performance becomes a difficult task. Meanwhile, within the barrel, the sabot affects the mechanical behavior of the warhead. According to the structure and material of the sabot, it is expected that the deformation and stress of the penetrator will be affected.

Vibration occurs when a transverse load is applied due to a variation of contact area between the sabot and the barrel. When the penetrator proceeds in the barrel, the deformation occurs in the penetrator according to the contact form between the sabot and the barrel, resulting in a change in the stress (σ_v) of the penetrator. As changes in the pressure within the barrel and the deformation of the projectile affect the accuracy of the strike and the structural stability of the projectile, analysis of such behavior is required. However, research on this is very limited. Since these weapon systems have various characteristics for each product, general standards cannot be applied, and individual research must be carried out according to individual target barrels and projectiles. In this study, we intend to conduct related research on a weapon system currently being developed by using the limited data collected.

The study aims to develop an analysis technique for the deformation of the penetrator that proceeds along the changing guide surface. Using this result, we intend to derive essential data to developing an optimized projectile by changing the projectile and sabot material.

2 ANALYSIS METHOD

The behavior of the penetrator moving in the gun barrel was analyzed using FEM. Fig. 2 shows that 8-node hexahedral elements were used. The analysis method is as follows: The barrel deflects due to gravity, and flexure occurs within the barrel under the influence of the manufacturing process, etc. The case where the barrel drooped due to gravity alone, and the case where bending was due to gravity and geometric tolerance overlap, were analyzed. The solid line indicates that the barrel is bent due to gravity alone (type G), and the dotted line shows the centerline shape of the barrel when the deflection due to gravity, and the deformation due to machining defects, are overlapped (type G+O) in Fig. 3.

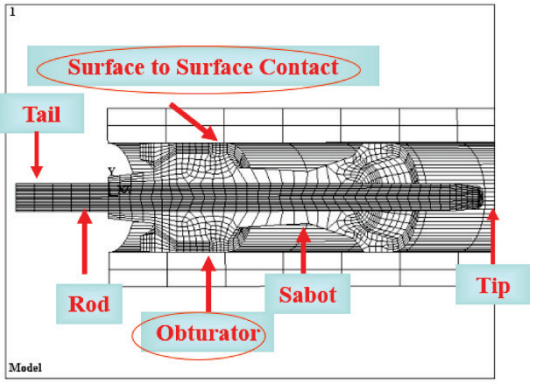


Figure 2 Analysis model

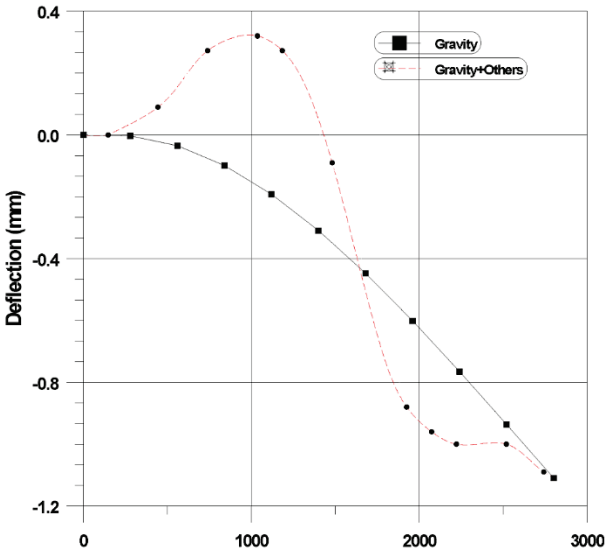


Figure 3 Gun tube profile for simulations

Table 1 Table title aligned centre

Type	AL-1	A	B	C	AL-2	D	E	F
Elastic Modulus (E) of Sabot (GPa)	72	49.2	47	48.5	72	49.2	47	48.5
Tube Centerline	Type G				Type G+O			

Tab. 1 shows the elastic modulus of the sabot and the deformation shape of the barrel. The elastic modulus of the sabot is (72, 49.2, 48.5, and 47) GPa, and the cases of behavior in 2 types of barrel are analyzed. Fig. 3 shows the

propulsion force applied to the projectile, while Fig. 4 shows the pressure change curve with time that is used.

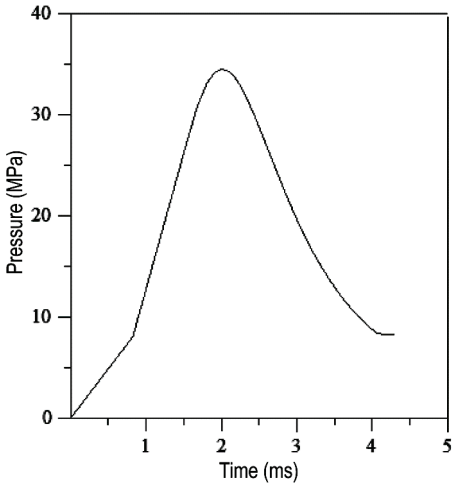


Figure 4 Breech pressure versus time

3 RESULTS AND DISCUSSION

3.1 Deformation of Penetrator

Fig. 5 and 6 show the deformation of the centerline of the projectile at (1 or 2) ms after firing. In the figure, the position represents the total length of the projectile, (0−40) mm represents the tail (fin tail) of the projectile, (40−160) mm is the part surrounded by the sabot, and (160−200) mm represents the front part of the projectile (warhead). When the sabot's modulus of elasticity is 47 GPa (△, ▲), which is the smallest at 1 ms, the deflection of the projectile was very small over the entire length of the projectile.

The difference in the amount of warhead deflection between the G case (●, ■, ▲, ◆) and the G+O case (○, □, △, ◇) when the projectile travels 2 ms is about 10^{-4} m. In addition, a sudden deflection occurs near the warhead

When the projectile movement time is 3 ms as shown in Fig. 7, this is a position where the primary inflection occurs in the barrel, in which the central trajectory of the barrel is inflected in two places. When the tube is deflected by its weight (●, ■, ▲, ◆), the transverse deflection of the entire length of the projectile is generally constant, and when E of the sabot is 72 GPa, the deflection of the warhead becomes larger than the other cases.

If there are two bends in the barrel (○, □, △, ◇), it can be seen that deformation of the entire length becomes larger than when the tube is drooping due to its weight only at the 3 ms point where the bending of the barrel is maximum. In this case, as E of the sabot decreases from 72 GPa (AL-2: ○) to 47 GPa (E: △), the difference between vertical deflection increases. From this result, it can be seen that the deformation of the warhead is affected by the bending of the barrel. If there is a bend in the barrel, the smaller the Young's modulus (E) of the sabot, the greater the difference of deflection between the warhead and the fin, as shown in Tab. 2.

Table 2 Deflection difference between tail and tip of projectile

Elastic Modulus (GPa)	$\Delta\delta$ only by Gravity (10^{-4}m)	$\Delta\delta$ by Superposition (10^{-4} m)
73	3	1.2
49.2	1	4.3
48.5	1.5	4.5
47	1.3	5.7

Tab. 2 shows the deformation difference between the tail and the warhead under each condition. In the case of superposing barrel deformation, when the sabot modulus of elasticity is 73 MPa, the difference in the amount of deformation is the smallest at 1.2×10^{-4} m. When it is 47 MPa, the difference in the amount of deformation is 5.7×10^{-4} m, which is the largest. It can be seen that the greater the elastic modulus of the sabot, the smaller the deformation difference between the warhead and the tail.

If the E of the sabot is 47 GPa in a projectile that travels when the barrel deflects with gravity, the deflection of the projectile according to the propagation time of the projectile is generally constant, and the difference between the deflection of the warhead and the tail is small. When the elastic modulus of the sabot is large as 73 GPa, the difference in deflection between the warhead and the fin is very large at the beginning of the launch. As time goes on, this difference in deflection gradually decreases; and just before the projectile leaves the barrel, this difference decreases considerably, and the deflection maintains a minimal shape.

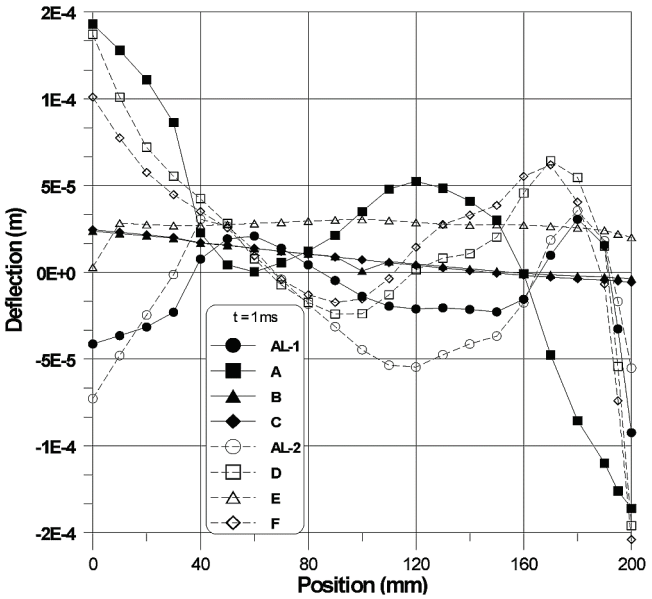


Figure 5 Deformation of penetrator centerline (t = 1 ms).

In the case where the barrel is type G+O, the overall fin is + deflection and the warhead is − deflection at the initial 1 ms of movement of the barrel, and the difference in deflection is large. At 2 ms, when the firing progress becomes stable, the warhead is directed in the + direction, and there is almost no bending deformation. The projectile in the barrel shows severe bending over its entire length. After some time has elapsed, the deflection of the total length decreases, and the stability of the deflection becomes stable.

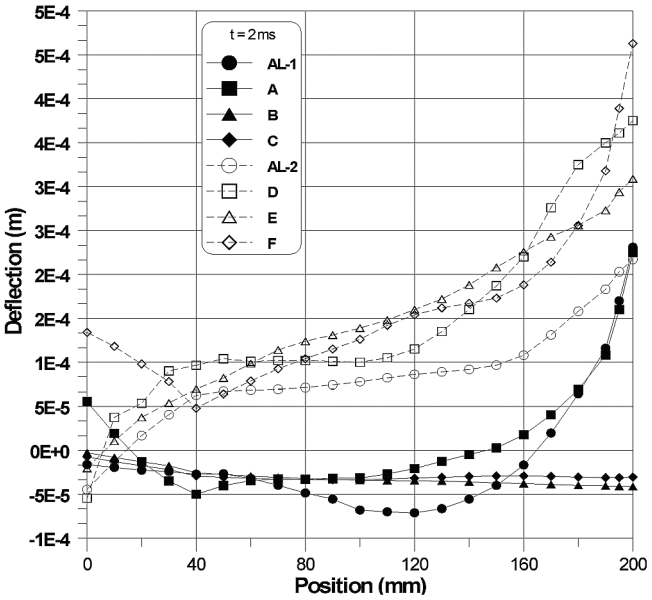


Figure 6 Deformation of penetrator centerline (t = 2 ms).

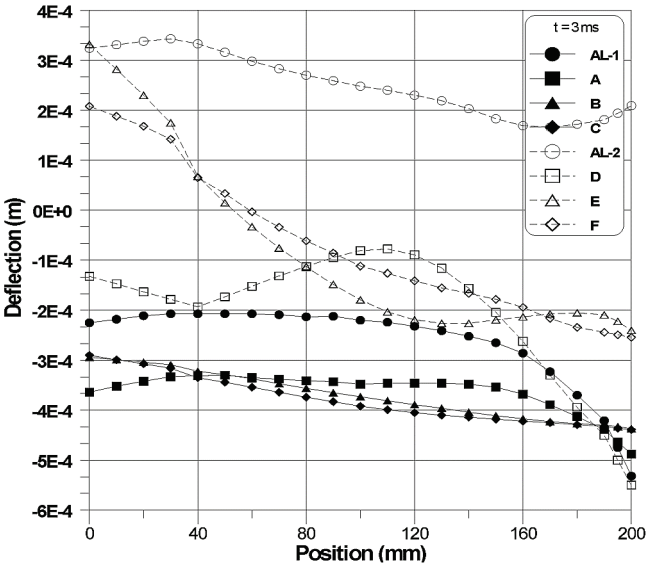
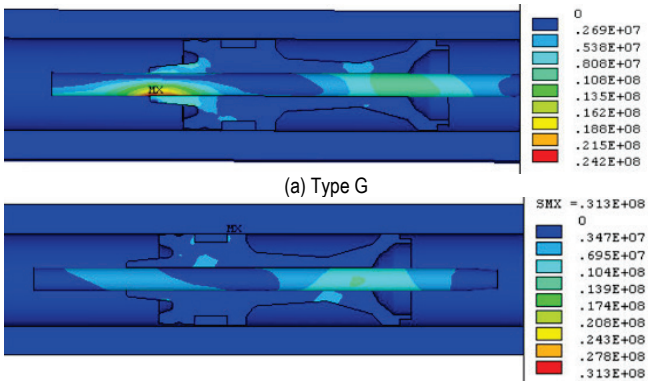


Figure 7 Deformation of penetrator centerline (t = 3 ms)

3.2 Stress of Penetrator

The stress at the point where the moving time of the projectile is about 3 ms for types G and G+O is presented in Fig. 8. In the case of type G, the maximum von Mises stress (σ_v) occurs in the front area between the sabot and projectile. The stress at this time is about 200 MPa. The reason for the difference in stress in the same moving time is as follows. The curvature of the barrel is constant in type G. However, in type G+O, it is judged that the barrel centerline is inflected in 3 ms.

The von Mises stress state in the warhead and tail during the entire process of the penetrator in the overlapped barrel shows in Fig. 9. The solid line indicates the stress state of the warhead according to the penetration time, while the symbol · indicates the stress state of the fin tail. The stress of the warhead gradually increases with time. The stress of the warhead increases greatly at the time of 3.6 ms, just before departure from the barrel. Comparing the results of the previous deflection with the results of Fig. 7, the deflection of the warhead is greater than that of the other parts.



(a) Type G

(b) Type G+O

Figure 8 von Mises stress distribution

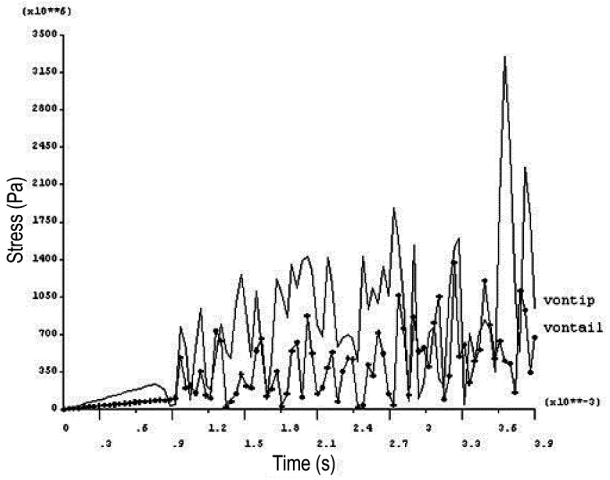


Figure 9 von Mises stress at full path (mixed centerline)

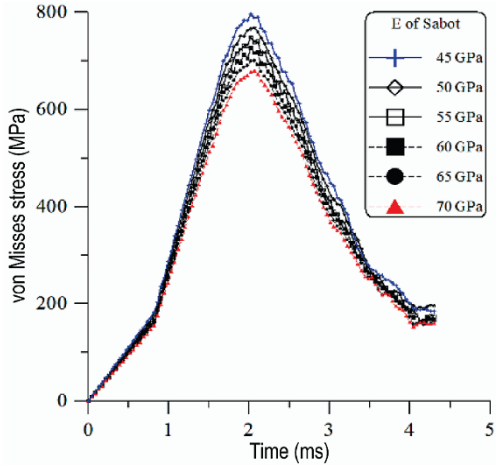


Figure 10 von Mises stress with the variation of elastic modulus of sabot

Fig. 10 shows the change in the maximum σ_v of the projectile according to the Young's modulus (E) of the sabot. The von Mises stress of the penetrator varies according to E of the sabot. The maximum stress occurs around 2 ms, and

when the elastic modulus of the sabot is 70 GPa, the stress of the penetrator is 630 MPa. When the modulus of elasticity of the sabot is 45 GPa, the stress of the penetrator becomes 800 MPa; and as the modulus of elasticity of the sabot increases, the von Mises stress of the penetrator decreases.

4 CONCLUSION

Numerical analysis was performed on the deformation and stress of the penetrator in the barrel that is deflected only by gravity, and in the barrel where the deformations of gravity and machining error are superposed. The penetrator is surrounded by a sabot, and the sabot proceeds while in contact with the barrel. The deflection and stress in the penetrator were analyzed according to the change in the characteristics of the sabot. The obtained results are as follows.

In the case of bending due to only its weight, the degree of deflection in the entire length of the penetrator at each position within the barrel was generally minimal. The deflection degree of the small elastic modulus of the sabot was smaller than that of the large elastic modulus. If the elastic modulus of the sabot is large, the deflection of the warhead is greater than the deflection of the fin in most cases proceeding within the barrel.

The deflection of the penetrator is similar at the beginning, regardless of the bending shape of the barrel. As the penetrator pushes forward, the difference in deflection of the penetrator in the barrel where the deformation is superimposed becomes large. The difference in deflection between the fin and the warhead is greater than when the elastic modulus of the sealer is small in the stage just before the penetrator leaves the barrel. Considering the deflection of the projectile, the larger the modulus of the sabot the more stable for the penetrator behavior.

As E of the sabot increases, the stress generated in the penetrator decreases. When the results of stress and deflection are considered, it can be seen that the larger the E of the sabot, the better the structural stability of the penetrator.

Acknowledgements

The author would like to thank Hanseo University for its substantial support (Project code: 2022065).

5 REFERENCES

[1] Sana, Z., Emad, U., & Aamir, M. (2020). Experimental investigation of impact of long rod eroding projectiles against rolled homogeneous armor. *Proceedings of the Institution of Mechanical Engineers Part L Journal of Materials Design and Applications, 234*(10), 1335-1342. https://doi.org/10.1177/14644207 20940766

[2] Sreelal, M., & Rajesh, G. (2012). Trajectory predictions of new lift separation sabots. *Defence Technology, 17*(4), 1361-1373. https://doi.org/10.3969/j.issn.2214-9147.2021.04.023

[3] Tolga, D. (2020). Effects of Projectile and Gun Parameters on the Dispersion. *Defence Science Journal*, 70(2), 166-174. https://doi.org/10.14429/dsj.70.14 922

[4] Cason, N. B., Mark, F. C., & Benjamin, T. D. (2022). Aerodynamic Stabilization of a Finless Projectile with Active Nose Deflection. *J Spacecraft and Rockets, 59*(5), 1592-160. https://doi.org/10.2514/1.A35352.

[5] Alexey, M. L., Stanislav, A. K., & Ivan, G. R. (2017). Optimization of aerodynamic form of projectile for solving the problem of shooting range increasing. *AIP Conference Proceedings*, 1893, 030085, 1-9. https://doi.org/10.1063/1.5007543

[6] Thomas, F. E., Brendan, J. P., & Alan, F. H. (2001). Dispersion analysis of the XM881 APFSDS projectile. *Shock Vib*, 8, 183-191. https://doi.org/10.1155/2001/585930

[7] Bruce, P. B., Drysdale, W. H., Hoppel, C. P., & Bogetti, T. (2001). The development of composite sabots for kinetic energy projectiles. *Proc. 19th International Symposium on Ballistics*.

[8] Zhang, W., Huang, X., & Qi, Y. (2017). Experimental investigation on ballistic stability of high-speed projectile in sand. *AIP Conference Proceedings*, 1793. https://doi.org/10.1063/1.4971587

[9] Alexander, E. Z., Paul, W., & John, B. (2000). Electromagnetic and aeromechanical analysis of sabot discard for rail gun projectiles. *J Spacecraft and Rocket, 37*(2), 257. https://doi.org/10.2514/2.3554

[10] Wang, Y., Cheng, J., Yu, J. Y., & Wang, X. M. (2016). Influence of yawing force frequency on angular motion and ballistic characteristics of a dual-spin projectile. *Defence Technology, 12*(2), 124-128. https://doi.org/10.1016/j.dt.2015.12.007

[11] Huang, Z., Xia, C., Cao, Y., Chen, Z., & Zhang, H. (2018). Numerical investigations on the sabots discard process of an APFSDS at different angles of attack. *19th International Conference of Fluid Power and Mechatronic Control Engineering*, 373-378. https://doi.org/10.1049/joe.2018.9006

[12] Lisy, P. & Bridik, L. (2019). Modal Analysis of Medium Calibre Barrels. Problems of Mechatronics. Armament, Aviation. *Safety Engineering, 2*(36), 9-22. https://doi.org/10.5604/01.3001.0013.2113

[13] David, L. H. (2018). The Effect of Projectile Weight on the Optimum Launch Angle and Range. *The Physics Teacher*, 56, 584. https://doi.org/10.1119/1.5080567

[14] Mehmet, A. K., İsmail, E., & Yusuf, Ç. (2016). Tip Deflection Determination of a Barrel for the Effect of an Accelerating Projectile Before Firing Using Finite Element and Artificial Neural Network Combined Algorithm. *Latin American Journal of Solids and Structures, 13*(10), 1968-1995. https://doi.org/10.1590/1679-78252718

[15] Newill, J., Burns, B., & Wilkerson, S. (1998). Over View of Gun Dynamics Numerical Simulations. *US Army Research Lab. Report*, TR-1760.

[16] Lyon, D., & Soencksen, K. (1994). Radial Stiffness an In Bore Balloting Analysis for the M900 Projectiles. *US Army Research Lab*. Rept., TR-593.

[17] Newill, J. & Bundy, M. (2003). Gun Tube Selection Methodologies for Accuracy Testing. *US Army Research Lab*. Rept., TR-3080.

[18] David, L., Mark, G., & Miljenko, L. (2021). Analysis of the effect of bore centerline on projectile exit conditions in small arms. *Defence Technology*, 1-10. https://doi. org/10.1016/j.dt.2021.09.008

[19] Zhengui, H., Chenchao, X., Yuan, C., Zhihua, C., & Huanhao, Z. (2019). Numerical investigations on the sabots discard process of an APFSDS at different angles of attack. *J Eng*., 13, 373-378. https://doi.org/10.1049/joe.2018.9006

[20] Bundy, M., Newill, J. F., Marcopoli, V., Ng, M., & Wells, C. (2000). A Methodology for Characterizing Barrel Flexure Due to Tank Motion. *US Army Research Laboratory*, ARL-MR479.
[21] Tomasz, B. & Mariusz, M. (2021). Numerical Analysis of Kinetic Energy Projectile Sabot Structure Influencing the Armour Penetration Depth. Part I - Projectile Basic Option. *Problemy Techniki Uzbrojenia, 155*(4), 23-48. https://doi.org/10.5604/01.3001.0014.8998

Authors' contacts:

Jin Bong Kim, PhD, Professor
University Hanseo,
Taean Campus 236-49, Gomseom-ro, Nam-Myeon, Taean-Gun,
Chungcheongnam-do, Republic of Korea 32158
jbkim@hanseo.ac.kr

ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20230429094940

Subject review
Received: 2023-04-29, Accepted: 2023-05-02

Lean Product Development Tools for Promotion of Sustainability Integration in Product Development

Ivana Cukor*, Miro Hegedic

Abstract: This article aims to enhance existing understanding of incorporating sustainability aspects during the product development and seeks to fill the gaps regarding the relationship between LPD and sustainability aspects. An up-to-date literature review was performed. More specifically, the expansion of current knowledge covers finding instances in earlier studies that explain the meaning of a sustainability aspect, such as environmental, social and economic aspect. Also, this article focuses on exploring various sustainability aspects using lean product development (LPD) tools and practices. LPD tools and practices that would enable the achievement of sustainability objectives are presented. The findings suggest that the chance for the integration of sustainability in product development comes when integrating sustainability aspects in LPD methods and tools that are used in companies daily. An analysis of the impact of every single LPD tools on individual aspects of sustainability is lacking. The paper concludes with recommendations for future research.

Keywords: lean product development; lean tools; sustainability; sustainability aspect

1 INTRODUCTION

Organizations that are responsible for product development processes and manufacturing have the potential to make a significant contribution to the achievement of sustainability goals within society and can obtain business economic benefits [1]. The new product development (NPD) has to build in the sustainability elements: environment, economic, and social—a sustainable product that is necessary for the industry [2, 3].

Since elementary product properties and characteristics are established during the initial stages, it is crucial to integrate sustainability into the product development process [4, 5]. The sustainability of the future product is determined in this phase [6]. Considering the environmental, social, and economic aspects from the start and include these aspects in the product design is, therefore, the most beneficial method for developing sustainable products. In any case, it is crucial to offer designers and engineers adequate assistance throughout this process [7]. The objectives of developing a sustainable product are to satisfy customers, achieve sustainability in business and meet stakeholders' demands on the industry. Nevertheless, adopting the criteria of sustainable products in the NPD industry is rarely present, especially in the automotive industry [2]. Therefore, the incorporation of sustainability into NPD is still in its infancy [8] and continues to pose challenges for organisations [5].

Lean Product Development (LPD) involves application of lean methodologies to product development, with the goal of developing new or improved products that successfully meet the needs of the market [9]. The increasing adopting of LPD by organizations opens up space to involve sustainable elements in its processes, methods, and tools. This will enable new products to become economically successful, environmentally correct, operationally secure, socially fair, and culturally accepted [10]. Lean philosophy could catalyse to promote better sustainability performance [11-13]. Inclusion of LPD into organizations lean management presents a promising opportunity to also take into account sustainability aspects in product development [14].

Although Lean and Sustainability have been used in many organisations, their integration to the development and design of new products still presents a challenge [5]. Lean and sustainability are increasingly used concepts in companies at all levels. They are used together, unitary and complementary, more and more as a single tool for reducing waste, being also an actual interdisciplinary theme of research [15].

The existing literature and application have not adequately investigated the possible conflicts, interactions, or overlaps between LPD and sustainability. Therefore, it is necessary to adopt a holistic systems approach that promotes integration and synergy in the usage of sustainable product development methods and tools [10]. Examining the relations between lean and sustainability seems like a practical approach to address the inquiry of how to combine and consequently implement these two initiatives [16].

Notwithstanding that, the investigation of lean and sustainability interrelationships has been perhaps one of the most researched subjects in recent times, most of the research efforts do not adequately answer the question: which aspects of sustainability can LPD tools affect? The limited viewpoint of previous investigations results in a fragmented understanding of the synergies between LPD tools and different types of sustainability aspects. Also, this article focuses on exploring various sustainability aspects using LPD tools and practices. It aims to find out what aspects of sustainability engineers in the industry consider in product development.

There is a gap in the literature relating to LPD tools and affected sustainability aspects. Hence, an attempt has been made in this paper to fill this gap.

The paper is structured as follows. After an introduction, the next section provides the research methodology. The following section presents research results followed by a descriptive analysis of the papers gathered and the most important findings. Finally, the conclusions, implications, and future research are outlined.

2 RESEARCH OBJECTIVES AND RESEARCH METHODOLOGY

The objective of this research is to bridge the gaps in the existing literature regarding the relationship between LPD and various sustainability aspects by addressing the following research questions:

- RQ1: Which aspects of sustainability can be influenced in the product development phase?
- RQ2: Which lean product development tools can be used to influence aspects of sustainability?

The first research question will be answered through finding the instances in earlier studies that explain the meaning of a sustainability aspect, such as environmental, social and economic aspect. The second research question focuses on the LPD tools and practices that would enable the achievement of sustainability objectives. In particular, we want to contribute to this field of knowledge by exploring the different types of tools and LPD practices and recognising they can be used to improve sustainability aspects based on the literature. It should be noted that sustainability is considered the unity of all three dimensions: environmental, social and economic.

A comprehensive analysis of the most current literature available was carried out. The purpose of the literature review was to recognise the LPD tools and sustainability aspects used in the product development process. Papers were gathered from international peer-reviewed journal articles and conferences and were extracted from the following online databases: Web of Science and Science Direct. This aligns with the recommendations to use a minimum of two databases for research purposes [12].

Table 1 Sustainability aspect grouped according to the sustainability dimensions and authors

Sustainability aspects	Reference	Sustainability dimension
emissions	[20] , [21], [18], [16], [4], [22]	Environmental
pollution	[20], [21], [18], [16], [23], [22]	
natural habitat conservation	[20], [21], [4]	
consumption of water	[24] , [25], [21], [18], [26], [16], [27], [22]	
consumption of material/ using replenishable resources / reducing the weight / using recyclable materials	[24], [25], [21], [26], [18], [16], [4], [23], [27], [22]	
consumption of energy	[28],[25], [21], [26], [18], [16], [4], [23], [27]	
land use	[24]	
adherence to regulations and certifications regarding material usage /environmental regulations and standards	[29], [25], [21], [4], [22]	
End-Of-Life strategy	[29], [25], [21], [18]	
energy efficiency while using the product /design for transport	[30], [25], [21], [26], [18], [16], [4], [23], [24]	
choosing materials that are not harmful or toxic / phasing out hazardous substances/ avoidance of conflict minerals for product parts and/or its manufacturing	[30], [21], [26], [18], [16], [23], [28], [24], [31], [4], [22]	
promote repair and upgrading/serviceability	[26], [18], [23]	
welfare of employees (including health and safety, career growth opportunities and contentment with the organization)	[20], [21], [26], [16], [23], [31], [24], [4], [25], [22]	Social
welfare of customers (including their health and safety during the production and usage of the product, ensuring customer contentment and rights are met)	[20], [21], [26], [16], [24], [22]	
welfare of community (participation in community development programs, human rights, equity, and anti-corruption measures)	[20], [21], [26], [27], [22]	
avoidance of forced labour, child labour, and corrupt practices	[32], [19]	
availability of quality drinking water, the right to form and join trade unions, promoting gender equality, and access to fundamental knowledge and education	[19]	
quality and durability of the product	[29], [23],	
functional performance	[29]	
product safety and health impact	[29], [21], [18], [16], [31], [24], [4], [25]	
product meets End-Of-Life standards and certifications	[29], [25]	
organizational learning	[21], [16]	
growing divided into: financial gain, expenses, and investments that an organization makes, return on investment	[20], [21], [18], [18], [16], [23], [24], [29], [24]	Economic
innovation potential	[4], [21], [16]	
increased competitiveness, competitive advantage, create economic opportunities	[4], [21], [16], [28]	
energizing employees	[4], [21], [16]	
Research and Development expenses	[29], [18]	
direct and indirect expenses, such as expense of labour and material	[29], [18], [18], [23]	
market profit and quality of products	[29], [21], [27]	
productivity, improvement in employee performance, Design for Manufacturing	[21], [18]	

To ensure that the literature review included current information, the search was restricted to articles and conference papers published in English from 2010 to present. An initial query string was built using keywords (lean AND product development AND sustain*), and a second one using synonyms (lean AND product design AND sustain*). Only the engineering domain was set relevant (At this stage, papers from fields such as arts and humanities, astronomy, and medicine were excluded). The final selection of documents was based on the following criteria: title of the publication or

the text outlined in the paper's abstract. If the initial criteria were not sufficient for exclusion, the next step involved reading the introduction and conclusion of the paper, and if necessary, the entire document was read to make a final decision. The resulting sample was enriched through snowballing.

3 RESULTS AND DISCUSSION

3.1 Sustainability Aspects

It is essential for both professionals and academics to recognize that all three dimensions of sustainability must be addressed at the same time in order to achieve substantial outcomes [17]. The answer to the first research question is given in Tab. 1, which represents an overview of the different sustainability aspects presented in the literature review in a structured way. Aspects of sustainability are listed in table 1 so that they are grouped according to pillars: environmental, social or economic, which is indicated in the third column. The second column shows authors who deal with a certain aspect in their works.

Typically, engineers and designers are not trained to identify the environmental and social consequences of the products they design. To address this issue, straightforward and effective tools are necessary to enable designers and engineers to consider the environmental and social implications of their work while still accounting for the complex nature of sustainability concerns [18].

J. P. Schöggl et al. [18] introduced a qualitative decision support tool for evaluating environmental, economic, and social aspects during the initial stages of product development.

Incorporating sustainability criteria into decision support for the product innovation process is a crucial factor in efficiently integrating a sustainability point of view in the initial stages of product development [4].

R. Gould et al. [19] designed support for using social principles to analyse product concepts.

The conclusion is that there is no lack of works that state aspects of sustainability, but these aspects are mostly included in various models for sustainability assessment. The mentioned articles and models lack a component that would evaluate the impact of the individual tools that designers already use in their work on the mentioned aspects of sustainability. This would make engineers more aware of their own impact on product sustainability through the tools they use and the decisions they make while using these tools daily.

3.2 Lean Product Development tools

Tab. 2 presents the compilation of LPD tools and practices retrieved from the literature. The answer to the second research question is given in Table 2, where are the first column listed LPD tools and practices which can be used to improve sustainability aspects. The second column listed all authors who cited certain LPD tool as influencing sustainability.

The similarities between lean principles and sustainability outweigh their differences. It seems that sustainability, driven by lean principles, still has a significant amount of unexplored potential [17].

The implementation level of methods and tools for supporting sustainability considerations in product development is generally low, possibly due to inadequate practical applicability and incomplete sustainability coverage. One alternative solution is to incorporate sustainability aspects into existing methods and tools that are commonly used in organizations [14].

To effectively tackle sustainability challenges and design products for a more environmentally friendly future, the initial requirement is to train employees with a new perspective that incorporates not only economic considerations but also environmental and social aspects into their daily works [5].

Using tools related to lean thinking can assist in recognizing economic, environmental, and social waste. The search for entirely new methodologies may not be the best solution to enhance the sustainability of products. The primary issue is not the scarcity of methods and tools, but rather their application, and further reflection is needed on how they can aid each aspect of sustainability [10].

The study conducted by K. F. Barcia et al. [21] identifies the frequently used lean six sigma approaches implemented to enhance sustainability. Apart from [21], no other paper states which aspects of sustainability can be influenced by certain lean tools.

Table 2 Authors grouped according to the tools and models analyzed

LPD Tool	Reference
A3	[33], [34], [5]
AHP	[35], [36]
Cradle to cradle (C2C) design	[37]
Design for sustainability (DfS)	[5], [38], [39], [40]
Design of experiments (DOE)	[33], [36]
Functional modelling	[33]
Employee involvement	[41], [21], [42]
Pull & Just-in-time (JIT)	[43], [17], [41], [21], [42], [40]
Kaizen (Continuous Improvement)	[33], [17], [43], [41], [21], [42], [40]
Plan, Do, Check, Act (PDCA)	[43], [17]
Poka Yoke	[43], [17], [40]
Preference set-based design (PSD) / Set based concurrent engineering	[14], [44], [38]
Product Development VSM	[43], [17], [41], [33], [38], [36]
Six sigma, QFD	[41], [38], [42], [40], [36], [33]
Standardization, Standardized work	[43], [17], [34], [40]
Trade-off curves	[33], [34]
Visual Management	[43], [17], [40]
5S	[41], [17], [43], [42], [40], [36]
FMEA	[36], [33]

By combining LPD with their lean management approach, organizations have a better chance of considering sustainability in their product development processes. Opportunities for integrating sustainable views into product development arise when sustainability aspects are incorporated into LPD methods and tools that are already utilized by companies on a regular basis. An analysis of the impact of every single LPD tools on individual aspects of

sustainability is lacking. It would also be necessary to examine whether certain aspects can be incorporated into known popular LPD tools to avoid the creation of new methods and tools.

4 CONCLUSION AND FUTURE RESEARCH

Theoretically, our study results contribute to advancing the literature in the field of LPD and sustainability. The study provided important information on relationship between LPD tools and sustainability aspects.

This manuscript helps foster the evolving debate of integrating the principle of lean and sustainable in product development and contributes to a deeper understanding of their relationship from the literature's examination. In conclusion, the research results provide valuable information for companies to make informed decisions. The continuous improvement teams of companies can utilize this research to identify the most commonly used lean methods and tools that have a positive impact on sustainability.

This study establishes a theoretical basis for further investigations on these topics, highlighting opportunities for future research. It also gives researchers and managers interested in enhancing sustainability the chance to obtain valuable insights on the types of LPD methods and tools that could be utilized. It points out the most utilised LPD tools and sustainability aspects across the economic, social, and environmental dimensions.

To provide directions for future research and taking into account the gaps found in their research, authors have identified the directions for future research:

1) Measurement of the impact of every single LPD tool on sustainability aspects
2) Investigate how can a sustainability aspect be integrated into LPD tools which are already used in companies daily

This would have practical implications for LPD as it would support the recognition of the tools and prioritise their influence so that improvement plans for sustainability could be made.

Like any research, this research has its own limitations, and it's crucial to recognise them. Firstly, this study's search process only covered two major database, which means that some relevant information might have been overlooked. Conducting a more thorough investigation and gathering data from additional databases could have resulted in a more comprehensive analysis of the connections between lean and sustainability.

5 REFERENCES

[1] Mesquita, P. L. & Missimer, M. (2021). Social sustainability work in product development organizations: An empirical study of three Sweden-based companies. *Sustain., 13*(4), 1-21. https://doi.org/10.3390/su13041986

[2] Ahmad, H. M. A. H. (2018). The criteria's of sustainable product development and organizational performance. *Proc. Int. Conf. Ind. Eng. Oper. Manag., 2018*(JUL), 1391-1391.

[3] Abdul-Rashid, S. H., Sakundarini, N., Raja Ghazilla, R. A., & Thurasamy, R. (2017). The impact of sustainable manufacturing practices on sustainability performance: Empirical evidence from Malaysia. *Int. J. Oper. Prod. Manag., 37*(2). https://doi.org/10.1108/IJOPM-04-2015-0223

[4] Hallstedt, S. I. (2017). Sustainability criteria and sustainability compliance index for decision support in product development. *J. Clean. Prod., 140*, 251-266. https://doi.org/10.1016/j.jclepro.2015.06.068

[5] Flores, M., Maklin, D., Ingram, B., Golob, M., Tucci, C., & Hoffmeier, A. (2018). Towards a sustainable innovation process: Integrating lean and sustainability principles. *IFIP Advances in Information and Communication Technology, 535*. https://doi.org/10.1007/978-3-319-99704-9_5

[6] Pigosso, D. C. A. & McAloone, T. C. (2015). Supporting the development of environmentally sustainable PSS by means of the ecodesign maturity model. *Procedia CIRP 2015, 30*. https://doi.org/10.1016/j.procir.2015.02.091

[7] Bocken, N. M. P., Farracho, M., Bosworth, R., & Kemp, R. (2014). The front-end of eco-innovation for eco-innovative small and medium sized companies. *J. Eng. Technol. Manag. - JET-M, 31*(1). https://doi.org/10.1016/j.jengtecman.2013.10.004

[8] Gmelin, H. & Seuring, S. (2014). Achieving sustainable new product development by integrating product life-cycle management capabilities. *Int. J. Prod. Econ., 154*. https://doi.org/10.1016/j.ijpe.2014.04.023

[9] Welo, T. & Ringen, G. (2016). Beyond Waste Elimination: Assessing Lean Practices in Product Development. *Procedia CIRP 2016, 50*, 179-185. https://doi.org/10.1016/j.procir.2016.05.093

[10] de Souza, J. P. E. & Dekkers, R. (2019). Adding sustainability to lean product development. *Procedia Manuf., 39*(2019), 1327-1336. https://doi.org/10.1016/j.promfg.2020.01.325

[11] Johansson, G. & Sundin, E. (2014) Lean and green product development: Two sides of the same coin? *J. Clean. Prod., 85*, 104-121. https://doi.org/10.1016/j.jclepro.2014.04.005

[12] Mittal, V. K., Sindhwani, R., & Kapur, P. K. (2016). Two-way assessment of barriers to Lean–Green Manufacturing System: insights from India. *Int. J. Syst. Assur. Eng. Manag., 7*(4). https://doi.org/10.1007/s13198-016-0461-z

[13] Dieste, M., Panizzolo, R., Garza-Reyes, J. A., & Anosike, A. (2019). The relationship between lean and environmental performance: Practices and measures. *J. Clean. Prod., 224*, 120-131. https://doi.org/10.1016/j.jclepro.2019.03.243

[14] Zetterlund, H., Hallstedt, S., & Broman, G. (2016). Implementation Potential of Sustainability-oriented Decision Support in Product Development. *Procedia CIRP 2016, 50*, 287-292. https://doi.org/10.1016/j.procir.2016.05.011

[15] Tăucean, I. M., Ivașcu, L., Șerban, M., & Negruț, M. (2019). Synergies between lean and sustainability: A literature review of concepts and tools. *Qual. - Access to Success, 20*(S1), 559-564.

[16] Martínez León, H. C. & Calvo-Amodio, J. (2017). Towards lean for sustainability: Understanding the interrelationships between lean and sustainability from a systems thinking perspective. *J. Clean. Prod., 142*, 4384-4402. https://doi.org/10.1016/j.jclepro.2016.11.132

[17] Tasdemir, C. & Gazo, R. (2018). A systematic literature review for better understanding of lean driven sustainability. *Sustain., 10*(7). https://doi.org/10.3390/su10072544

[18] Schöggl, J. P., Baumgartner, R. J., & Hofer, D. (2017). Improving sustainability performance in early phases of product design: A checklist for sustainable product development tested in the automotive industry. *J. Clean. Prod., 140*, 1602-1617. https://doi.org/10.1016/j.jclepro.2016.09.195

[19] Gould, R., Missimer, M., & Mesquita, P. L. (2017). Using social sustainability principles to analyse activities of the extraction lifecycle phase: Learnings from designing support for concept selection. *J. Clean. Prod., 140*, 267-276. https://doi.org/10.1016/j.jclepro.2016.08.004

[20] Joung, C. B., Carrell, J., Sarkar, P., & Feng, S. C. (2013). Categorization of indicators for sustainable manufacturing. *Ecol. Indic., 24*. https://doi.org/10.1016/j.ecolind.2012.05.030

[21] Barcia, K. F., Garcia-Castro, L., & Abad-Moran, J. (2022). Lean Six Sigma Impact Analysis on Sustainability Using Partial Least Squares Structural Equation Modeling (PLS-SEM): A Literature Review. *Sustain., 14*(5). https://doi.org/10.3390/su14053051

[22] Bertoni, M. (2019). Multi-criteria decision making for sustainability and value assessment in early PSS design. *Sustain., 11*(7). https://doi.org/10.3390/su11071952

[23] Deutz, P., McGuire, M., & Neighbour, G. (2013). Eco-design practice in the context of a structured design process: An interdisciplinary empirical study of UK manufacturers. *J. Clean. Prod., 39*, 117-128. https://doi.org/10.1016/j.jclepro.2012.08.035

[24] Paulson, F. (2018). Inclusion of Sustainability Aspects in Product Development at Manufacturing Companies. *Licentiate Dissertation*, Linköping University.

[25] Letens, G. (2015). Lean Product Development—Faster, Better … Cleaner? *Front. Eng. Manag., 2*(1), 52-59. https://doi.org/10.15302/J-FEM-2015007

[26] Luttropp, C. & Lagerstedt, J. (2006). EcoDesign and The Ten Golden Rules: generic advice for merging environmental aspects into product development. *J. Clean. Prod., 14*(15), 1396-1408. https://doi.org/10.1016/j.jclepro.2005.11.022

[27] Souza, J. P. E. & Alves, J. M. (2018). Lean-integrated management system: A model for sustainability improvement. *J. Clean. Prod., 172*, 2667-2682. https://doi.org/10.1016/j.jclepro.2017.11.144

[28] Poulikidou, S., Björklund, A., & Tyskeng, S. (2014). Empirical study on integration of environmental aspects into product development: Processes, requirements and the use of tools in vehicle manufacturing companies in Sweden. *J. Clean. Prod., 81*. https://doi.org/10.1016/j.jclepro.2014.06.001

[29] Shuaib, M., Seevers, D., Zhang, X., Badurdeen, F., Rouch, K. E., & Jawahir, I. S. (2014). Product Sustainability Index (ProdSI). *J. Ind. Ecol., 18*(4). https://doi.org/10.1111/jiec.12179

[30] Issa, I. I., Pigosso, D. C. A., McAloone, T. C., & Rozenfeld, H. (2015). Leading product-related environmental performance indicators: a selection guide and database. *J. Clean. Prod., 108*(PartA). https://doi.org/10.1016/j.jclepro.2015.06.088

[31] Tingström, J., Swanström, L., & Karlsson, R. (2006). Sustainability management in product development projects - the ABB experience. *J. Clean. Prod., 14*(15-16). https://doi.org/10.1016/j.jclepro.2005.11.027

[32] Ekener-Petersen, E. & Finnveden, G. (2013). Potential hotspots identified by social LCA - Part 1: A case study of a laptop computer. *Int. J. Life Cycle Assess., 18*(1). https://doi.org/10.1007/s11367-012-0442-7

[33] Lermen, F. H., Echeveste, M. E., Peralta, C. B., Sonego, M., & Marcon, A. (2018). A framework for selecting lean practices in sustainable product development: The case study of a Brazilian agroindustry. *J. Clean. Prod., 191*, 261-272. https://doi.org/10.1016/j.jclepro.2018.04.185

[34] Oliveira, G. A., Tan, K. H., & Guedes, B. T. (2018). Lean and green approach: An evaluation tool for new product development focused on small and medium enterprises. *Int. J. Prod. Econ., 205*, 62-73. https://doi.org/10.1016/j.ijpe.2018.08.026

[35] Arroyo, P., Tommelein, I. D., & Ballard, G. (2012). Comparing multi-criteria decision-making methods to select sustainable alternatives in the AEC industry. *International Conference on Sustainable Design, Engineering, and Construction, ISCDEC 2012*. https://doi.org/10.1061/9780784412688.104

[36] Kaswan, M. S. & Rathi, R. (2020). Green Lean Six Sigma for sustainable development: Integration and framework. *Environ. Impact Assess. Rev., 83*, 106396. https://doi.org/10.1016/j.eiar.2020.106396

[37] Zhang, W. (2009). Sustainability and manufacturing philosophy - From mass production to intelligent energy field manufacturing. *Proceedings of the ASME International Manufacturing Science and Engineering Conference, MSEC2009, 2*. https://doi.org/10.1115/MSEC2009-84304

[38] Sorli, M., Sopelana, A., Salgado, M., Pelaez, G., & Ares, E. (2012). Balance between lean and sustainability in product development. *Key Eng. Mater., 502*, 37-42. https://doi.org/10.4028/www.scientific.net/KEM.502.37

[39] Nahkala, S. (2013). Aligning product design methods and tools for sustainability. https://doi.org/10.1007/978-981-4451-48-2_9

[40] Cherrafi, A., Elfezazi, S., Chiarini, A., Mokhlis, A., & Benhida, K. (2016). The integration of lean manufacturing, Six Sigma and sustainability: A literature review and future research directions for developing a specific model. *J. Clean. Prod., 139*. https://doi.org/10.1016/j.jclepro.2016.08.101

[41] Ciannella, S., Santos, L. C., & Morioka, S. N. (2019). Does lean mean sustainable? Exploring linkages through a systematic literature review. https://doi.org/10.14488/enegep2019_ti_st_290_1634_37597

[42] Caldera, H. T. S., Desha, C., & Dawes, L. (2017). Exploring the role of lean thinking in sustainable business practice: A systematic literature review. *J. Clean. Prod., 167*. https://doi.org/10.1016/j.jclepro.2017.05.126

[43] Vinodh, S., Arvind, K. R., & Somanaathan, M. (2011). Tools and techniques for enabling sustainability through lean initiatives. *Clean Technol. Environ. Policy, 13*(3). https://doi.org/10.1007/s10098-010-0329-x

[44] Inoue, M., Lindow, K., Stark, R., & Ishikawa, H. (2010). Preference Set-Based Design Method for Sustainable Product Creation. In: Pokojski, J., Fukuda, S., Salwiński, J. (eds) New World Situation: New Directions in Concurrent Engineering. *Advanced Concurrent Engineering*. Springer, London. https://doi.org/10.1007/978-0-85729-024-3_42

Authors' contacts:

Ivana Cukor, MSc Mech. Eng.
(Corresponding author)
Yazaki Europe Limited,
Slavonska avenija 26/6, 10000 Zagreb, Croatia
cukor.ivana@gmail.com

Miro Hegedić, dr. sc.
Department of Industrial Engineering,
Faculty of Mechanical Engineering and Naval Architecture,
University of Zagreb,
Ivana Lučića 5, 10000 Zagreb, Croatia
miro.hegedic@fsb.hr


COMPLAS 2023
XVII International Conference on Computational Plasticity
Fundamentals and Applications
5–7 September 2023, Barcelona, Spain
ECCOMAS Thematic Conference
iacm Special Interest Conference
A CONFERENCE CELEBRATING
THE 70th BIRTHDAY OF EUGENIO OÑATE
Co-Chairs
MICHELE CHIUMENTI, CIMNE / UNIVERSITAT POLITÈCNICA DE CATALUNYA, SPAIN
CARLOS AGELET DE SARACIBAR, CIMNE / UNIVERSITAT POLITÈCNICA DE CATALUNYA, SPAIN
DJORDJE PERIC, SWANSEA UNIVERSITY, UNITED KINGDOM
EDUARDO DE SOUZA-NETO, SWANSEA UNIVERSITY, UNITED KINGDOM
Conference Secretariat
CIMNE Congress Bureau
www.cimne.com
Phone +34 – 93 405 46 94
complas_sec@cimne.upc.edu
CIMNE

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

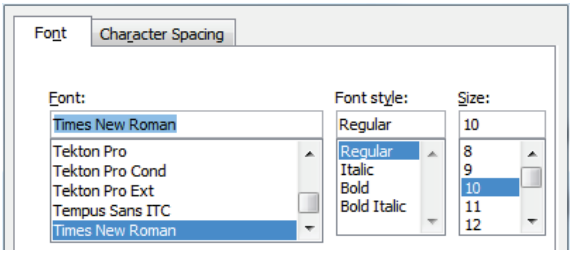


Figure 1 Text under the figure [1]
(Style: Arial Narraw, 8pt, Align Centre)

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

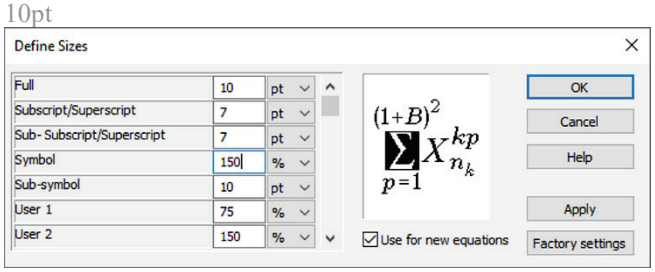


Figure 2 The texts under figures

(Style: Arial Narraw, 8pt, Align Centre)

10pt

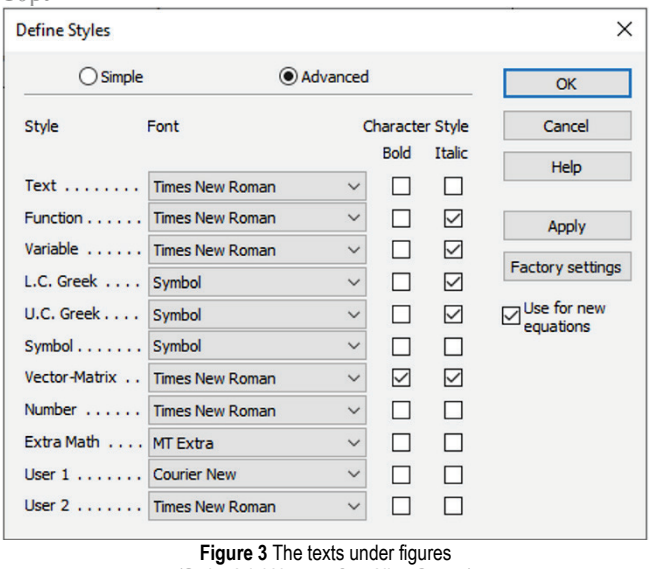


Figure 3 The texts under figures

(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.

TEHNIČKI GLASNIK / TECHNICAL JOURNAL – GODIŠTE / VOLUME 17 – BROJ / ISSUE 2

LIPANJ 2023 / JUNE 2023 – STRANICA / PAGES 153-303

Sveučilište Sjever

SVEUČILIŠTE SJEVER / UNIVERSITY NORTH – CROATIA – EUROPE

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

TEHNICKIGLASNIK@UNIN.HR – HTTP://TEHNICKIGLASNIK.UNIN.HR