DOI: 10.21062/mft.2023.036 http://www.journalmt.com

Overall Equipment Effectiveness-Related Assembly Pattern Catalogue based on Machine Learning

Péter Dobra (0000-0002-1805-5347)[1], János Jósvai (0000-0001-9417-5191)[2]
[1]Doctoral School of Multidisciplinary Engineering Science, Széchenyi István University, 9026 Győr, Egyetem tér 1, Hungary. E-mail: dobra.peter@sze.hu
[2]Department of Vehicle Manufacturing, Széchenyi István University, 9026 Győr, Egyetem tér 1, Hungary. E-mail: josvai@ga.sze.hu

Nowadays, a lot of data is generated in production and also in the domain of assembly, from which different patterns can be extracted using machine learning methods with the support of data mining. With the help of the revealed patterns, the assembly operations and processes can be further optimized, thus the profit achieved can be increased. This article attempts to explore the patterns related to the most used Key Performance Indicator (KPI) in manufacturing, the Overall Equipment Effectiveness (OEE) metric. The patterns and relationships discovered will be sorted into Assembly Pattern Catalogue (APC). Firstly, a literature review demonstrates scientific relevance. Secondly, it examines the circumstances and methods of samples in the Manufacturing Execution System (MES) data source and Enterprise Resource Planning (ERP) systems. In the third section, the detailed pattern catalogue is defined in the area of assembly. The novelty of the article is that beyond the generalization of patterns, it characterizes the pattern catalogue with mentioning practical industrial examples.

Keywords: Machine learning, Pattern catalogue, Assembly line, OEE

1 Introduction

Recently, enormous data amount is generated in production and also in the domain of assembly. Applying Industry 4.0 (I4.0) and Internet of Things (IoT) technology, more and more real-time on-site data is collected from assembly lines [1]. The current IT systems connect to machines, workers and products [2]. In industrial practice, Manufacturing Execution System (MES) is one of the most common digital tools to collect data about the entire production system, including efficiency figures. This data collection is supported by various smart sensors, barcodes, vision systems and wireless technologies [3]. MES provides support amongst other data collection, performance analysis, product tracking, process management, machine control, material and production logistics [4]. This execution system is an industrial software and a bridge between Enterprise Resource Planning (ERP) and controlling systems (e. g., PLC) [5]. In the smart factories, MES serves ERP with operational information such as assembled units, performance data, downtimes and scrap rates [6].

One of the effective ways to process and evaluate the large amount of manufacturing and assembly data is the use of data mining and machine learning methods. Data mining is a step in a Knowledge Discovery in Database (KDD) process when patterns are extracted from historical data [7]. This process known as knowledge extraction, information discovery, data archeology and data pattern processing [8]. Machine learning algorithms as a part of data mining can be supervised, unsupervised and reinforcement learning. Several machine learning algorithms are used for analysis and prediction among others decision trees, clustering, Bayesian, regression, regularization, instance-based tools, ensemble methods, neural networks, and deep learning [9-13]. During machine learning one of the aims is to reveal different normal and abnormal patterns with computer power which a human brain not necessarily would have found [14, 15]. Although there are plenty of tools and software available, industrial companies are not using enough data mining and machine learning methods to identify hidden patterns in manufacturing data. In daily practice, this is especially true for the Key Performance Indicator (KPI) used in the field of assembly lines as well as for Overall Equipment Effectiveness (OEE) [16].

Assembly lines such as dedicated assembly lines, flexible assembly lines, reconfigurable assembly lines are become more and more complex in manufacturing industry in the domain of automotive, electronics and complex equipment manufacturing. In a turbulent market environment semi-automatic or hybrid assembly lines in mass production with flexibility and changeability is essential [17, 18]. Smart planning and manufacturing process monitoring are the part of data-driven smart manufacturing. Data can be analysed by machine learning algorithms that identify the patterns of normal behaviour and identify unusual or risky events. [19]. During the entire optimization process several patterns can be found such as, material consumption patterns, energy consumption patterns and product relevant parameters (e. g., geometric, tolerance, machining parameters) [20, 21]. However, patterns have not been yet revealed in the case of OEE. This metric is an efficiency indicator that shows the difference between the ideal state and what has been achieved in reality. According to Nakajima the original formula for calculation Overall Equipment Effectiveness is written as:

$$OEE = a\ p\ q \tag{1}$$

Where:
a…Availability [%],
p…Performance [%],
q…Quality [%] [22].

A pattern is a local, usually complex structure in the dataset. At the domain of manufacturing, according to Ji et al. it is a big challenge to predict potential failures based on error patterns before the occurrence [23]. This can be a significant advantage for industrial companies. Based on another approach, manufacturing profiles as patterns are recording and applying in a proactive way for prediction, these profiles are saved as a probable environment behaviour. [24]. Bergman et al. for the tool failure pattern recognition used the following statistics during machine learning: count, mean, standard deviation, minimum, first quartile, median, third quartile, maximum, mode, maximum absolute step, maximum relative step, minimum absolute step, minimum relative step, value change, area under the curve and arc length [25]. Niedermann et al. optimized deep business automated processes and detected different patterns amongst other parallelization, decomposition and elimination [26]. The mentioned patterns with the connected processes were collected in a pattern catalogue. Pattern catalogue was defined as a consistent collection of patterns which mapped to a common formalism. Several business patterns were identified such as triage with clustering, automated decision with decision tree and resource selection with multiple regression [27].

This article focuses on the question: how OEE indicator patterns of the assembly lines can be described. The goal is to arrange the recognized patterns in Assembly Pattern Catalogue (APC) and to define the concept of the catalogue.

2 Material and methods

Previously, during manufacturing process optimization, the Manufacturing Pattern Catalogue was defined as typical optimization option (e. g., best practices) [28]. However, the authors of this article understand that a pattern catalogue is not the same as good manufacturing practice. During the creation of the Assembly Pattern Catalogue, human (e. g., selection of machine error codes), assembly lines (e. g., process parameters) and products (e. g., number of assembled units based on barcode) provide data to the various systems. This is a huge amount of data that the human brain cannot process efficiently, therefore machine learning is necessary. Figure 1 shows the process of generating an Assembly Pattern Catalogue.

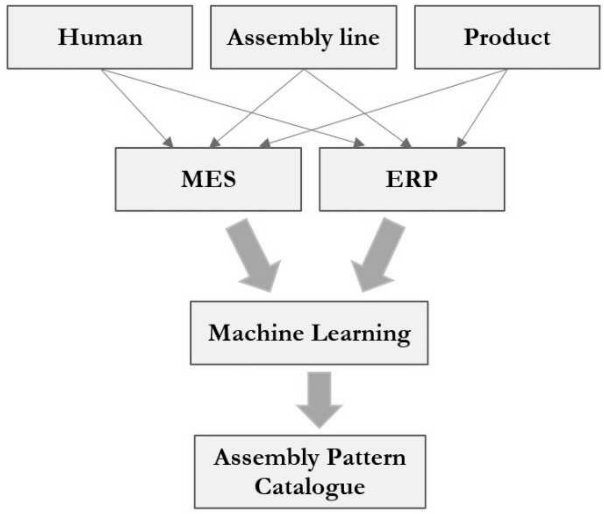

Fig. 1 *The process of making Assembly Pattern Catalogue*

The following primarily data sources were used during the recognition of assembly patterns for Overall Equipment Effectiveness:

- Manufacturing Execution System (MES), contains relevant historical data of all production stages, OEE values and its contributor data (quantity, planned assembly time, downtime, scrap, etc.);
- Enterprise Resource Planning (ERP), contains assembly batch data, planning data (e. g., assembly sequence);

Secondary data sources:

- SQL database, contains each product and process data, such as cycle time, product type, bottleneck workstation, etc.;
- Log files, which are usually typical values of a workstation (e. g., failure code).

Patterns can be explored from two perspectives, detected patterns and predicted patterns. The detected patterns refer to the present, showing a status on a historical basis, while the predicted patterns refer to the future, which contains uncertainty. The detected and predicted patterns show great similarity. In addition, it is advisable to focus on the bottleneck workstation first when revealing the patterns.

In order to interpret the Assembly Pattern Catalogue methodologically, it must be placed in a prediction model. This OEE prediction model is shown in Figure 2.

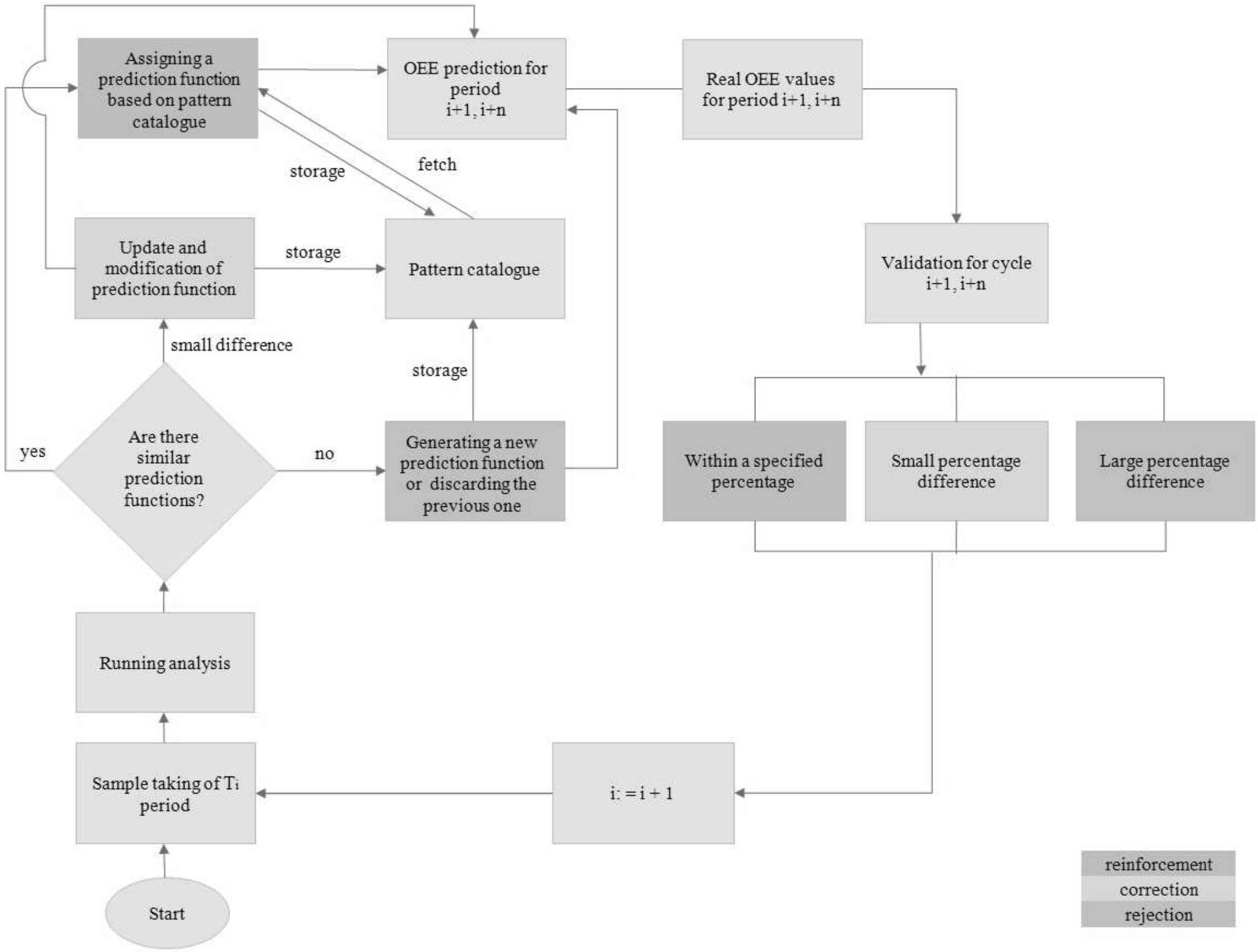

Fig. 2 *OEE prediction model with Pattern Catalogue*

After defining the sample period, the model performs an analysis (descriptive statistics, significance, adjusted R2, etc.) using machine learning tools and then examines the generated prediction function to see if there is the same, similar or completely different. Accordingly, a new prediction function is generated, or the previous one is modified or discarded. The prediction function is recorded in the pattern catalogue together with the samples and revealed patterns. The predicted OEE values is compared with the real OEE values, then validation takes place, where the applied prediction function is reused (reinforcement), updated (correction) or discard (rejection) in the next cycle. The purpose of this article is to present the Assembly Pattern Catalogue rather than explain the prediction model in more detail.

3 Results and discussion

In the field of assembly processes, patterns have not been revealed so far, much less compiled in any catalogue. Samples (from MES, ERP, SQL databases and log files), patterns and prediction functions (using machine learning tools) are available from the assembly lines, so these can be organized and stored in one structure, in Assembly Pattern Catalogue (APC). The schematic diagram of the Assembly Pattern Catalogue is shown in Figure 3.

The three main elements of Assembly Pattern Catalogue are:

- Samples (range of raw data, records that contain time series data);
- Patterns (revealed local useful structure by machine learning);
- Prediction functions (one or more mathematical function for prediction).

There is a one-way or two-way connection between the individual elements, which is represented by arrows in Figure 3. During the exploration of OEE-relevant assembly patterns, the patterns related to its main components, such as availability patterns, performance and quality patterns were prioritized. In addition to general descriptive statistics, these patterns were also partially explored by examining logical relationships. In the following, some discovered assembly patterns are detected and described that are affecting OEE values.

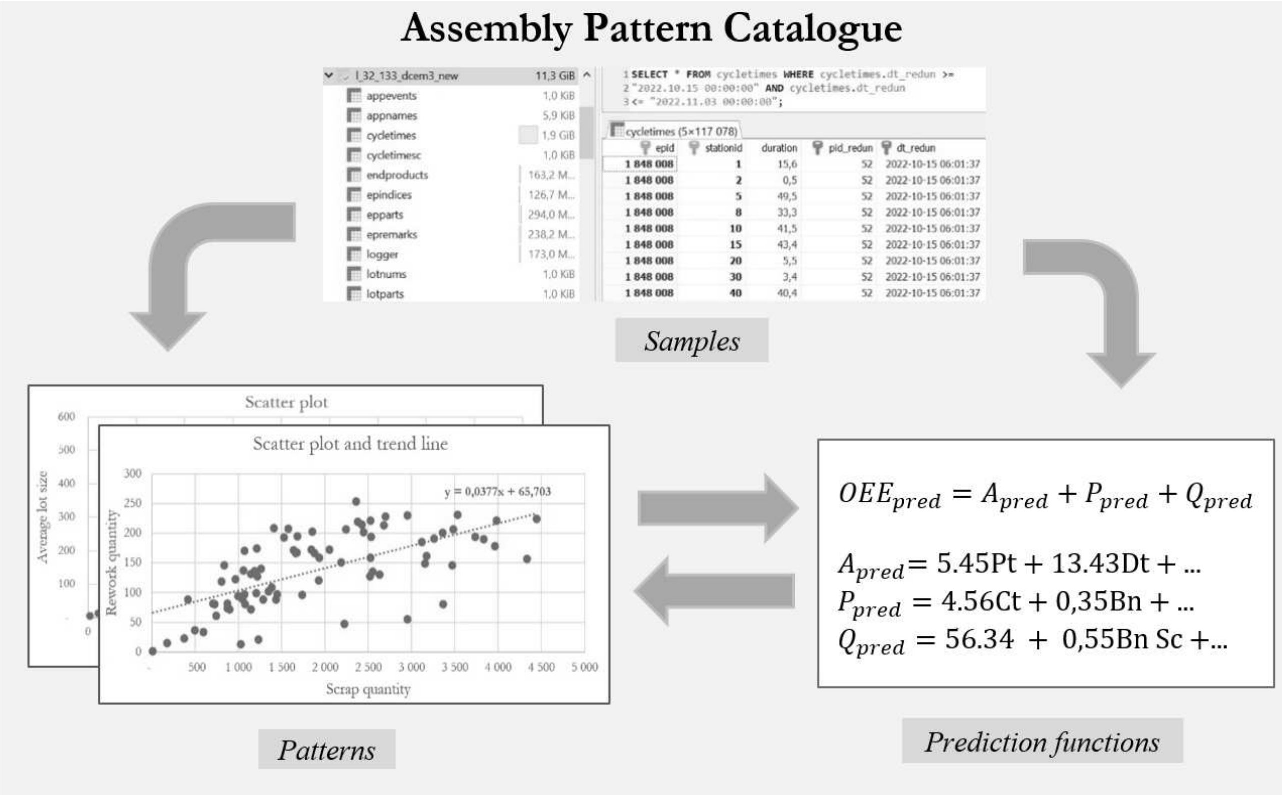

Fig. 3 *Elements of the Assembly Pattern Catalogue*

Availability patterns revealed by machine learning:

- Typical periods of machine failures (e. g., start of assembly process, first shift);
- Planned and unplanned downtimes distribution;
- Effect of maintenance strategy (the time and duration of the planned activities).

Performance patterns revealed by machine learning:

- Human and machines performance ripple curve and cyclicity;
- Recognition of normal pace and pulsation of the assembly operations;
- Variable cycle times on critical stations and takt times on the assembly line;
- Best of best values and circumstances;
- Deviation from assembly technology (e. g., swapped assembly operations).

Quality patterns revealed by machine learning:

- Number of scrap units and scrap rate after type changes and/or machine failures;
- Rejected pieces one after the other (workstation, failure code, series, etc.);
- Appearance of visual failures on the products (e. g., place, form, etc.);
- Revealed connection between SPC values and machines setup, adjustment.

3.1 Practical implementation of Assembly Pattern Catalogue

One of the main goals of creating the Assembly Pattern Catalogue is to predict the OEE value as accurately as possible. This is difficult mainly because some of the assembly operations are stochastic and the prediction of technical errors is often uncertain, but machine learning can overcome.

In this chapter, the previously presented Assembly Pattern Catalogue is applied in practice through a real industrial example. Data from the seat structure semi-automatic assembly line of a European automotive company from the years 2021 and 2022 were used. The authors choose RStudio program as a machine learning environment to recognize the assembly patterns. The regression model used was MGCV (Mixed GAM Computation Vehicle) regression, where the data of 769 records were analysed. Figure 4 shows the histogram of OEE values.

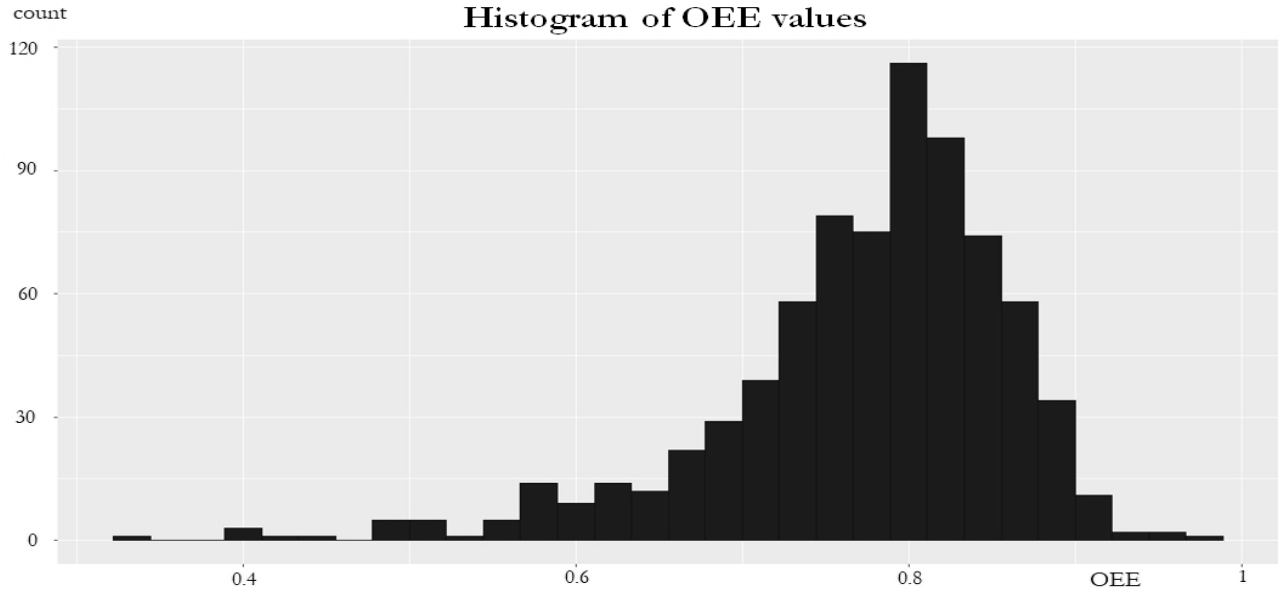

Fig. 4 *Histogram of OEE values*

Each record covers one shift (eight hours) of data for a semi-automatic assembly line. The following independent variables were considered for regression: process failure downtime, break downtime, technical downtime, changeover downtime, quality reason downtime, logistics reason downtime, not planned downtime, other downtime reason, number of changeover, average cycle time, number of assembled units and number of scrap units. OEE, availability, performance and quality values are the dependent variables. Figure 5 shows the correlation matrix for the entire dataset.

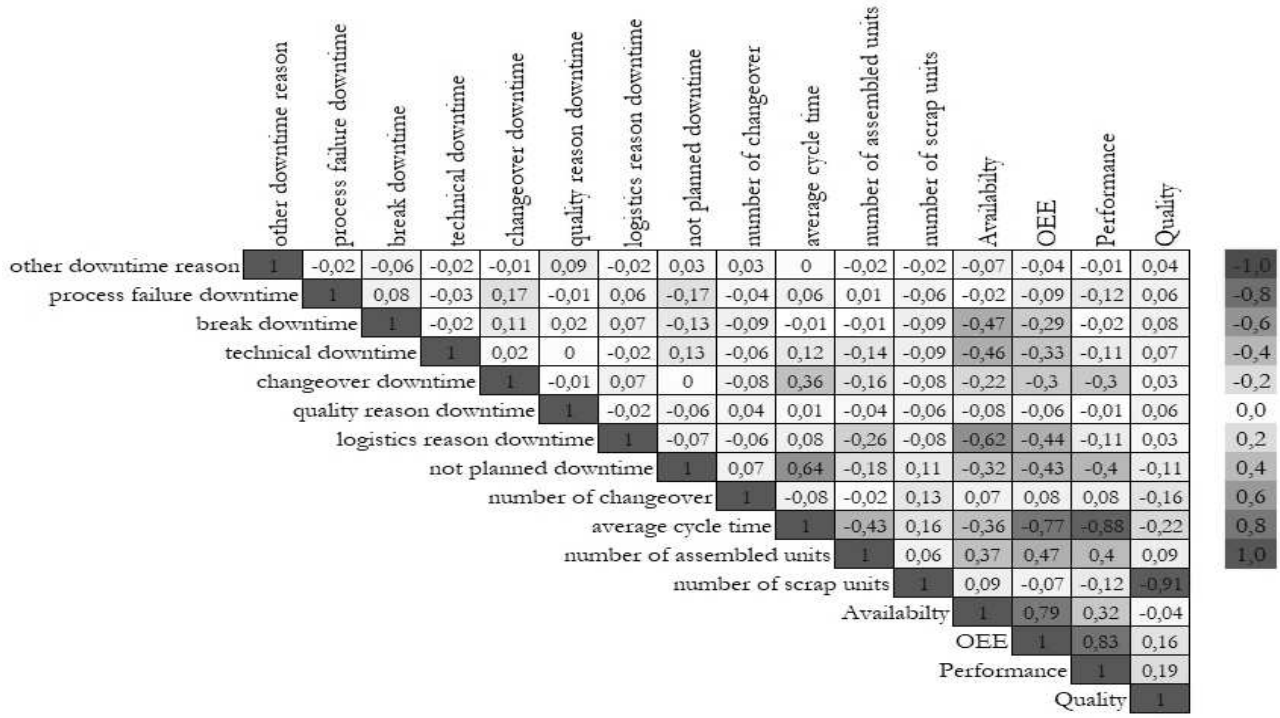

Fig. 5 *Correlation matrix*

In the following, one prediction case is presented as an example together with their characteristics.

The next settings were used during the case study:

- Training set: 100-150 records (50 records),
- Test set: 151-180 records (30 records).

A detail of the descriptive statistics of the training set is shown in Table 1.

The elements of the prediction function are the independent variables, whose parametric coefficients are shown in table 2.

ISSN 1213–2489
e-ISSN 2787–9402

***Tab. 1** Descriptive statistics of the training set (detail)*

	Break downtime [s]	Logistics reason downtime [s]	Not planned downtime [s]	Number of changeover	Average cycle time [s]	OEE
Minimum	0	0	289	0	62.42	0.5645
1st Quartile	1344	387	497.5	1	67.10	0.7008
Median	1770	557	651	1	69.90	0.7357
Mean	1562	683	754	1.549	70.77	0.7258
3rd Quartile	1862	903	884	2	73.67	0.7607
Maximum	2144	2316	2094	7	85.87	0.8082

***Tab. 2** Parametric coefficients*

	Estimate	Std. Error	t value	Pr(>|t|)	sig.
Intercept	7.366e-01	6.805e-02	10.825	2.59e-13	***
other downtime reason	1.090e-04	1.304e-04	0.836	0.40831	
process failure downtime	-1.598e-04	7.825e-05	-2.042	0.04800	*
break downtime	-1.804e-05	1.110e-05	-1.626	0.11208	
technical downtime	-2.378e-05	1.086e-05	-2.190	0.03455	*
changeover downtime	-4.586e-05	2.200e-05	-2.084	0.04374	*
quality reason downtime	0.000e+00	0.000e+00	-	-	
logistics reason downtime	-2.360e-05	1.529e-05	-1.543	0.13081	
not planned downtime	-6.710e-05	1.913e-05	-3.507	0.00116	**
number of changeover	-9.535e-05	4.232e-03	-0.023	0.98214	
average cycle time	2.629e-05	3.825e-05	0.687	0.49590	
number of assembled units	2.794e-04	1.304e-04	2.143	0.03841	*
number of scrap units	-8.276e-03	4.040e-03	-2.049	0.04728	*
Signif. codes: 0 '***' ; 0.001 '**' ; 0.01 '*' ; 0.05 '.', 0.1 ' '; 1					

Based on Table 2, the significant factors in the examined period are process failure downtime, technical downtime, changeover downtime, not planned downtime, number of assembled units and number of scrap units. Figure 6 shows MGCV regression for OEE values where training set was 100-150 and test set was 151-180. The blue line represents the predicted values and the red line represents the actual OEE values.

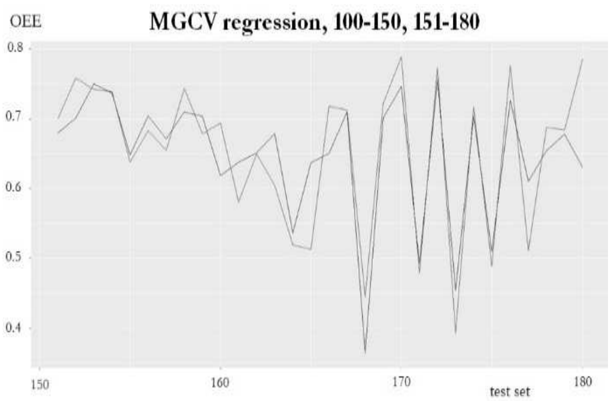

***Fig. 6** MGCV regression*

The accuracy of the prediction is 6.7398 according to MAPE (Mean Absolute Percentage Error), 0.0409 according to MAE (Mean Absolute Error), 0.0030 according to MSE (Mean Squared Error) and 0.0554 according to RMSE (Root Mean Squared Error).

Many patterns can be revealed during the analysis and prediction, which can be recorded in the pattern catalogue. Figure 7 and Figure 8 show examples of the revealed patterns. Figure 7 shows a scattered plot of the OEE values and assembled parts. Marked in blue, the assembly of two product types with different cycle times can be separated, and it can be seen that the OEE value is mostly around 0.8 (80%). The high OEE values in addition to the lower number of assembled units mean that the production did not take place in a full shift, but only in part. From the point of view of production planning, the assembly line should not be planned with 90% capacity and lower OEE values should also be considered.

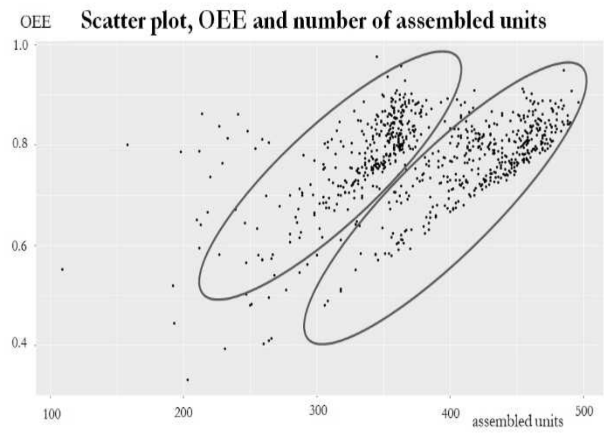

***Fig. 7** OEE and assembled units scatter plot*

Figure 8 shows a scattered plot of the OEE values and total downtimes. The green arrow indicates that a higher OEE value can be achieved with less downtime, but it is noticeable that in most cases there is about 2500 seconds (41.6 minutes) of downtime per shift, which is why a higher OEE value cannot be achieved. In addition to these, there are many cases where even higher downtimes occur.

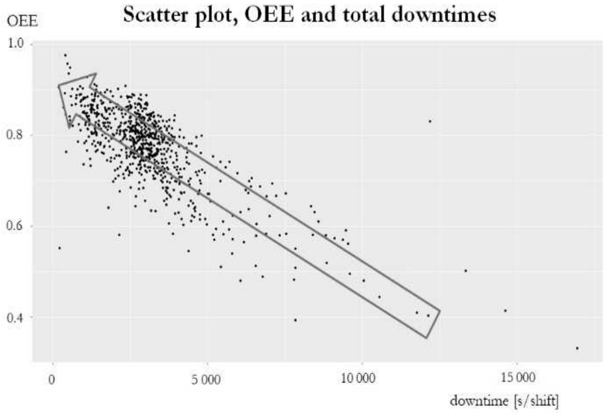

Fig. 8 *OEE and total downtimes scatter plot*

Beyond these patterns, further patterns among others can be logical patterns, seasonality patterns, trends and extreme values. It is important to examine the appearance conditions, frequency, length, priority, occasional characteristics and abnormality of the patterns.

4 Conclusions

In industry, it is crucial to monitor and improve assembly performance based on the patterns revealed by machine learning. This article focused on how OEE indicator patterns of the assembly lines can be described. These recognized patterns supported by data of Manufacturing Execution System (MES), Enterprise Resource Planning (ERP), different databases and log files. The samples, patterns and prediction functions were organized and stored in one structure named Assembly Pattern Catalogue (APC). During exploration of OEE-relevant assembly patterns, the main components of this metrics such as availability, performance and quality were prioritized.

This paper presented a model which includes APC to predict assembly efficiency by Mixed GAM Computation Vehicle) regression as supervised machine learning. Histogram of OEE values, correlation matrix, descriptive statistics, parametric coefficients, regression and scatter plots were presented to illustrate the practical usability at an automotive company. By creating and using the systematic Assembly Pattern Catalogue, more accurate and faster production and capacity planning can be achieved. With the continuous expansion of APC, more and more patterns are becoming known, by taking them into account, production and assembly can become more balanced. A further research direction could be the exploration of additional patterns in the case of other production or logistics metrics such as OTDP (On Time Delivery Performance).

Acknowledgement

The APC was funded by Széchenyi István University, Győr, Hungary, Publication Support Program.

References

[1] KANG, Z., CATAL, C., TEKINERDOGAN, B. (2020). Machine learning applications in production lines: A systematic literature review. *Computers and Industrial Engineering*, Vol. 149, 106773. ISSN 0360-8352.

[2] ENKE, J., GLASS, R., KREß, A., HAMBACH, J., TISCH, M., METTERNICH, J. (2018). Industrie 4.0 – Competencies for a modern production system. *Procedia Manufacturing*, Vol. 23, pp. 267-272. ISSN 2351-9789.

[3] KUSIAK, A. (2017). Smart manufacturing must embrace big data. *Nature*, Vol. 544, pp. 23-25. ISSN 1476-4687.

[4] BEREGI, R., PEDONE, G., HÁY, B., VÁNCZA, J. (2021). Manufacturing Execution System integration through the standardization of a Common Service Model for Cyber-Physical Production Systems. *Applied Sciences*, Vol. 11, 7581. ISSN 2076-3417.

[5] MANTRAVADI, S., MOLLER, C. (2019). An overview of next-generation Manufacturing Execution Systems: How important is MES for Industry 4.0? *Procedia Manufacturing*, Vol. 30, pp. 588-595. ISSN 2351-9789.

[6] OLÁH, J. (2019). Framework of Industry 4.0 Technologies. *International Journal of Engineering and Management Sciences*, Vol. 4. No. 4, pp. 213-223. ISSN 2498-700X.

[7] FAYYAD, U., PIATETSKY-SHAPIRO, G., SMYTH, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, Vol. 17, No. 3, pp. 37-54. ISSN 0738-4602.

[8] FAYYAD, U., PIATETSKY-SHAPIRO, G., SMYTH, P. (1996). The KDD process for extracting useful knowledges from volumes of data. *Communications of the ACM*, Vol. 39, No. 11, pp. 27-34. ISSN 0001-0782.

[9] WUEST, T., WEIMER, D., IRGENS, C., THOBEN, K.D. (2016) Machine learning in manufacturing: Advances, challenges, and applications. *Production and Manufacturing Research*, Vol. 4, pp. 23-45. ISSN 2169-3277.

[10] PEJIC, A., MOLCER, P.S. (2021). Predictive machine learning approach for complex problem-solving process data mining. *Acta Polytechnica Hungarica*, Vol. 18, pp. 45-63. ISSN 2064-2687.

[11] CADAVID, J.P.U., LAMOURI, S., GRABOT, B., PELLERIN, R., FORTIN, A. (2020). Machine learning applied in production planning and control: A state-of-the-art in the era of Industry 4.0. *Journal of Intelligent Manufacturing*, Vol. 31, pp. 1531-1558. ISSN 1572-8145.

[12] MIENYE, I.D., SUN, Y., WANG, Z. (2019). Prediction performance of improved decision tree-based algorithms: A review. *Procedia Manufacturing*, Vol. 35, pp. 698-703. ISSN 2351-9789.

[13] SHUAILIANG, G., HAN, Z., XIANGZENG, L., LIZHI, G. (2021). Comparison on milling force model prediction of new cold saw blade milling cutter based on deep neural network and regression analysis. *Manufacturing Technology*, Vol. 21, No. 4, pp. 456-463. ISSN 1213-2489.

[14] BUER, S.V., FRAGAPANE, G.I., STRANDHAGEN, J.O. (2018). The data-driven process improvement cycle: Using digitalization for continuous improvement. *IFAC PapersOnLine*, Vol 51, No. 11, pp. 1035-1040. ISSN 2405-8963.

[15] HARDING, J.A., SHAHBAZ, M., SRINIVAS, S., KUSIAK, A. (2006). Data mining in manufacturing: A review, *Journal of Manufacturing Science and Engineering*, Vol. 128, pp. 969- 976. ISSN 1528-8935.

[16] HADDAD, T., SHAHEEN, B. W., NÉMETH, I. (2021). Improving Overall Equipment Effectiveness (OEE) of extrusion machine using Lean manufacturing approach. *Manufacturing Technology*, Vol. 21, No. 1, pp. 56-64. ISSN 1213-2489.

[17] KERN, W., RUSITSCHKA, F., BAUERNHANSL, T. (2016). Planning of workstation in a modular automotive assembly system. *Procedia CIRP*, Vol. 57, pp. 327-332. ISSN 2212-8271.

[18] GYULAI, D., MONOSTORI, L. (2017). Capacity management of modular assembly systems. *Journal of Manufacturing Systems*, Vol. 43, No. 1, pp. 8-99. ISSN 0278-6125.

[19] BENCZÚR, A. (2018). Data-driven methodologies and Big Data. *In: Security challenges in the 21th century*, pp. 351-365. Dialóg Campus Kiadó, Budapest, Hungary. ISBN 978-615-5920-76-9.

[20] TAO, F., QI, Q., LIU, A., KUSIAK, A. (2018). Data-driven smart manufacturing. *Journal of Manufacturing Systems*, Vol. 48, pp. 157-169. ISSN 0278-6125.

[21] CÉZOVA, E. (2023). Economical and statistical optimization of the maintenance in the production process. *Manufacturing Technology*, Vol. 23, No. 1, pp. 32-39. ISSN 1213-2489.

[22] NAKAJIMA, S. (1988). *Introduction to TPM: Total Productive Maintenance*, Productivity Press, Cambridge, UK. ISBN 978-0915299232.

[23] JI, W., WANG, L. (2017). Big data analytics-based fault prediction for shop floor scheduling. *Journal of Manufacturing Systems*, Vol. 43, pp. 187-194. ISSN 0278-6125.

[24] KLÖS, V., GÖTHEL, T., GLESNER, S. (2018). Be prepared: Learning environment profiles for proactive rule-based production planning. *44th Euromicro Conference on Software Engineering nd Advanced Applications (SEAA),* Prague, Czech Republic, 2018, pp. 89-96. ISBN 978-1-5386-7384-3.

[25] BERGMANN, J., ZELENY, K.É., VÁNCZA, J., KŐ, A. (2022). Tool failure recognition using inconsistent data. *Procedia CIRP*, Vol. 107, pp. 1204-1209. ISSN 2212-8271.

[26] NIEDERMANN, F., RADESCHÜTZ, S., MITSCHANG, B. (2010). Deep business optimization: A platform for automated process optimization. *In: Abramowicz, W., Alt, R., Fähnrich, K.P., Franczyk, B., Maciaszek, L.A. Informatik 2010, Business Process and Service Science – Proceedings of ISSS and BPSC* pp. 168-180.

[27] NIEDERMANN, F., SCHWARZ, H. (2011). Deep business Optimization: Making business process optimization theory work in practice. *In: Enterprise, Business-Process and Information System Modelling*, Vol. 81, pp. 88-102. ISBN 978-3-642-21759-3.

[28] GRÖGER, C., NIEDERMANN, F., MITSCHANG, B. (2012). Data mining-driven manufacturing optimization. *Proceedings of the World Congress of Engineering 2012*, Vol. 3, July 4-6, 2012, London, UK. ISBN 978-988-19252-2-0.