

PREDICTION OF OVERALL EQUIPMENT EFFECTIVENESS IN ASSEMBLY PROCESSES USING MACHINE LEARNING

DOBRA Péter¹, JÓSVAI János^{2*}

¹ Doctoral School of Multidisciplinary Engineering Sciences, Széchenyi István University, 9026 Győr, Hungary;
dobra.peter@sze.hu

² Department of Vehicle Manufacturing and Technology, Széchenyi István University, 9026 Győr, Hungary;
josvai@ga.sze.hu

Abstract: 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 support of various modern technical and Information Technology (IT) tools, the recording, storage and processing of large amounts of data is now a routine activity. Based on machine learning, efficiency metrics including Overall Equipment Effectiveness (OEE), can be partially predicted, but industrial companies need more accurate and reliable methods. The analyzed algorithms can be used in general for all production units or machines where production data is recorded by Manufacturing Execution System (MES) or other Enterprise Resource Planning (ERP) systems are available. This paper presents and determinates which most used machine learning methods should be combined with each other in order to achieve a better prediction result.

KEYWORDS: machine learning; OEE; assembly; prediction

1 Introduction

Nowadays, one of the most complex tasks in the field of planning and management of production systems is predicting the efficiency of assembly operations, despite the fact that real-time measurement, tracking and monitoring are usually automated activities using MES (Manufacturing Execution System - MES) [1]. With the help of the Overall Equipment Effectiveness (OEE) indicator, which is the most used in industrial practice, the production planning and management areas of companies are able to measure performance and make forecasts [2, 3]. Accurate prediction has an impact on stock levels, delivery times, resource planning, maintenance activities, required investments, and thus financial results [4-6]. Based on these, it is extremely important to be able to predict the expected efficiency of semi-automatic assembly lines as accurately, quickly and reliably as possible during production planning.

Within the framework of Industry 4.0, a lot of data is constantly being generated in the field of assembly, the processing of which can no longer be efficient with the help of human resources alone, however, the application of machine learning (ML) methods has opened up new opportunities to predict OEE values [7-10]. Many algorithms are capable of prediction, however, the prediction of the OEE indicator consisting of the product of availability, performance and quality is a special case. Due to limitations (0% - 100%), stochastic nature, and continuous development activities, the OEE base will vary, as a result of which it will be difficult to compare data from two different times.

The aim of this paper is to combine individually used machine learning methods with each other and determine which combinations achieve the best prediction results.

2 Materials and Methods

A production unit and within it the efficiency of assembly operations can be measured in several ways [11]. Overall Equipment Effectiveness (OEE) as a commonly used standard and best practice metric was presented within the Total Productive Maintenance concept by Nakajima [12]. The equation for calculation OEE is written as:+

$$OEE = a \cdot p \cdot q \quad (1)$$

where,

a—availability [%]

p—performance [%]

q—quality [%].

Based on the machine learning scientific literature, the prediction of OEE has been approached among others:

- using decision tree, neural networks, support vector machine at intraday forecasting of OEE [13]
- simple moving average with Holt's double exponential smoothing method using Python [14]
- recurrent neural networks to reveal hidden correlation that affecting to OEE [15]
- Person correlation and one-sample T-test with Design Expert software [16]
- using grid search algorithm by decision tree, K-nearest neighbor, neural networks and support vector machine [17]
- support vector machine, random forest, gradient boost, deep learning [18, 19]
- OEE prediction by Bayes-based machine learning, decision tree, support vector machine and logistic regression [20, 21].

2.1 Applied machine learning methods

The primary consideration for the selection of machine learning methods was to choose supervised learning methods that are widely known and accepted in the scientific literature for the analysis of time series. In addition to these, industrial applicability was also important, so procedures requiring less computing capacity and requiring predictable time were preferred. Another aspect was that all the selected methods could be used within a given program package, which was made possible by R and RStudio. The chosen procedures are shown in Table 1.

Table 1 Applied machine learning methods.

	Machine learning methods	Abbreviation
1	Multiple linear regression	MLR
2	Polynomial regression (simple)	Pol1
3	Polynomial regression (complex, degree>2)	Pol2
4	Lasso regression	Lasso
5	Ridge regression	Ridge
6	Random forest regression	RF
7	Gradient boost regression	GB
8	Mixed GAM computation vehicle with automatic smoothness estimation regression	MGCV
9	K-nearest neighbors	KNN

2.2 Evaluation and comparability

During the estimation, evaluation and comparing of the predicted results were carried out using root mean squared error (RMSE) based on the next formula:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\bar{y}_i - y_i)^2}{n}} \quad (2)$$

where,

n – number of fitted points

y_i – actual value

$\bar{y}_i$ – predicted value.

After the analysis, the individual RMSE values were compared with each other, so that the lower value was considered better.

2.3 Real assembly environment

The data used in this research work comes from the MES system of an automotive company. The assembly data used are real, accurate, continuous and scaled with no factors. There are 15 workplaces on the hybrid or semi-automatic assembly line, the planned takt time is 60 seconds, seat structure assembly operating in two shifts.

2.4 Prediction approach and main settings

During the prediction, a rolling horizon approach was used for the selected machine learning procedures. In the rolling horizon approach, the selected training data use a different time window for each estimation, which moves forward continuously with one week data.

One week data contains 10 shift (2 shift, 5 days) OEE values. The detailed prediction settings are shown in Table 2.

Table 2 Applied main prediction settings

parameter	value
rolling window	1 week: 10 shifts OEE data
training set	8 weeks: 80 shifts OEE data
test set	6 weeks: 60 shifts OEE data
independent variables	12 (downtime, cycle time, etc.)

The amount of available data (767 shifts) supported the above settings, so the rolling events could be performed in 67 cases. The prediction for 6 weeks was determined based on industrial practical needs.

3 OEE prediction with alternating algorithms

Obviously, the combination of two procedures shows a better result than the independent application. The main question is which two machine learning method should be combined alternately in order to achieve an even more accurate prediction. The selected methods were alternately combined based on the algorithm presented in Figure 1.

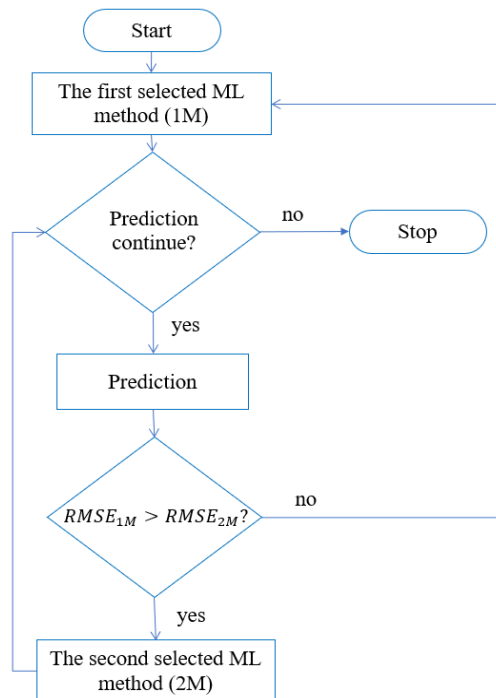


Fig. 1 Prediction algorithm of alternating approach.

The estimation is based on the following steps:

- in the case of the first estimation, the RMSE value corresponding to the estimated OEE values of the first selected algorithm is accepted
- in the next step, the sequence of prediction steps can be continued or stopped after any number of cycles

- after performing the OEE estimation with the algorithm for each step, the RMSE values can be calculated and compared with each other
- in the given cycle, the RMSE values ($RMSE_{1M}$ and $RMSE_{2M}$) of the two selected algorithms are compared
- after the first estimation, which algorithm has a lower RMSE value, the prediction continues until the RMSE value of the other procedure is lower
- the number of switching to another algorithm with a lower RMSE value is not limited.

The performed calculations and comparisons confirmed that the average RMSE value decreases mainly due to the alternating. In addition, it can be established that there are typically smaller blocks (predictions number 6–10) before the changes, so that constant switches do not dominate. The results obtained by combining machine learning are shown in the Table 3.

Table 3 RMSE values based on combination of machine learning methods.

	MLR	Pol 1	Pol 2	Lasso	Ridge	RF	GB	MGCV	KNN
MLR	-	0.0488	0.0492	0.0471	0.0497	0.0465	0.0469	0.0489	0.0478
Pol 1	0.0488	-	0.0543	0.0490	0.0507	0.0503	0.0510	0.0488	0.0512
Pol 2	0.0492	0.0543	-	0.0496	0.0509	0.0509	0.0517	0.0493	0.0518
Lasso	0.0471	0.0490	0.0496	-	0.0489	0.0488	0.0486	0.0471	0.0491
Ridge	0.0497	0.0507	0.0509	0.0489	-	0.0490	0.0484	0.0497	0.0491
RF	0.0465	0.0503	0.0509	0.0488	0.0490	-	0.0529	0.0464	0.0538
GB	0.0469	0.0510	0.0517	0.0486	0.0484	0.0529	-	0.0470	0.0551
MGCV	0.0489	0.0488	0.0493	0.0471	0.0497	0.0464	0.0470	-	0.0478
KNN	0.0478	0.0512	0.0518	0.0491	0.0491	0.0538	0.0551	0.0478	-

Based on the given estimation environment, the best results (0.0464 and 0.0465) can be achieved by using Random Forest (RF) and Multiple linear regression (MLR) or Random Forest and MGCV machine learning method alternately. It is necessary to note that the operation of MLR and MGCV is very similar.

Comparing the RMSE values without alternation with the results achieved with alternation, it can be concluded that in all cases, a better estimation result can be reached with the alternating algorithm. Based on Table 4, the rate of improvement is between 3.8% and 19.2%. To predict OEE, the proposed combination is to use Random Forest and MGCV machine learning alternating procedures. Without alternating data comes from previously published research results [22].

Table 4 Results of without alternating and with alternating approach.

	MLR	Pol 1	Pol 2	Lasso	Ridge	RF	GB	MGCV	KNN
without alternating	0.0489	0.0546	0.0569	0.0489	0.0513	0.0540	0.0559	0.0489	0.0570
alternating with	RF	MLR or MGCV	MLR	MLR or MGCV	GB	MGCV	MLR	RF	MLR or MGCV
	0.0465	0.0488	0.0492	0.0471	0.0484	0.0464	0.0469	0.0464	0.0478
$\Delta\%$	-5.2%	-11.9%	-15.7%	-3.8%	-6.0%	-16.4%	-19.2%	-5.4%	-19.2%

The combined procedure should also be used in the case of lower-performing methods (e.g. KNN, Pol 2), because the accuracy of the estimation can be further improved. The limitation of the study is that other prediction settings (e.g. test set, training set) should be tried and validation is required for other similar assembly lines.

CONCLUSION

This paper predicted the value of Overall Equipment Effectiveness using nine different selected supervised machine learning tools by combining two methods. Applying the rolling approach, the accuracy of the estimates was compared based on the RMSE indicator. It can be confirmed by using alternating approach prediction algorithm that the combination of two procedures shows a better result than the independent application.

Based on the defined estimation environment, the best result can be achieved by using Random Forest (RF) and Multiple linear regression (MLR) or Random Forest and MGCV machine learning method alternately. Using the alternating way the rate of prediction improvement is between 3.8% and 19.2%.

The authors plan to examine further machine learning settings as training sets, test sets and rolling window size. It may be necessary to confirm and validate the results on other similar types of semi-automatic assembly lines.

REFERENCES

- [1] Beregi, R., Pedone, G., Háý, B., Váncza, J. “Manufacturing Execution System integration through the standardization of a common service model for Cyber-Physical Production System”, *Applied Sciences* 11, 7581, **2021**. DOI: 10.3390/app11167581
- [2] Cadavid, J. P. U., Lamouri, S., Grabot, B., Pellerin, R., Fortin, R. “Machine learning applied in production planning and control: A state-of-the-art in the era of Industry 4.0”, *Journal of Intelligent Manufacturing* 31, pp. 1531 – 1558, **2020**. DOI: 10.1007/s10845-019-01531-7
- [3] Zhou, Z.H. “Machine Learning”, Springer Nature: Singapore, **2021**. DOI: 10.1007/978-981-15-1967-3
- [4] Prasetyawan, Y., Giffary, F., Putera, B. S. A. “The proposed OEE - Sigma prediction for increased profits”, *Materials Science and Engineering* 847, pp. 1 – 9, **2020**. DOI: 10.1088/1757-899X/847/1/012034
- [5] Arinez, J. F., Chang, Q., Gao, R. X., Xu, C., Zhang, J. “Artificial intelligence in advanced manufacturing: Current status and future outlook”, *Journal of Manufacturing Science and Engineering* 142, pp. 1 – 16, **2020**. DOI: 10.1115/1.4047855
- [6] Omar, F., Alexandre, K., Alekcie S. “Predicting the Burnishing Force for Cylindrical Workpieces with Amodified Surface Layer”, *Journal of Mechanical Engineering* 72, pp. 35 – 48, **2022**. DOI: 10.2478/scjme-2022-0004
- [7] Plathottam, S. J., Rzonca, A., Lakhnori, R., Iloeje, C. O. “A review of artificial intelligence applications in manufacturing operations”, *Journal of Advanced Manufacturing and Processing* 5 (3), pp. 1 – 19, **2023**. DOI: 10.1002/amp2.10159
- [8] Song, Z., Luo, S. “Application of machine learning and data mining in manufacturing Industry”, *International Journal of Computer Science and Information Technology* 2 (1), pp. 425 – 436, **2024**. DOI: 10.62051/ijcsit.v2n1.45
- [9] Akshansh, M. “Machine Learning Algorithm for Surface Quality Analysis of Friction Stir Welded Joint”, *Strojnícky časopis – Journal of Mechanical Engineering* 70 (2), pp. 11 – 20, **2020**. DOI: 10.2478/scjme-2020-0016
- [10] Chen, T., Sampath, V., May, M. C., Shan, S., Jorg, O. J., Aguilar Martín, J. J., Stamer, F., Fantoni, G., Tosello, G., Calaon, M. “Machine learning in manufacturing towards,

- Industry 4.0: From 'For Now' to 'Four-Know'", *Applied Sciences* 13, pp. 1 – 32, **2023**. DOI: 10.3390/app13031903
- [11] Jordan, M. I., Mitchell, T. M. "Machine learning: Trends, perspectives, and prospects", *Science* 349, pp. 255 – 260, **2015**. DOI: 10.1126/science.aaa8415
- [12] Nakajima, S. "Introduction to TPM: Total Productive Maintenance", Productivity Press: Cambridge, UK, **1988**.
- [13] Mittal, S., Khan, M. A., Romero, D., Wuest, T. "Smart manufacturing: Characteristics, technologies and enabling factors", *Proceedings of the Institution of Mechanical Engineers, Journal of Engineering Manufacture* 233, pp. 1342 – 1361, **2019**. DOI: 10.1177/0954405417736547
- [14] Anusha, C. H., Umasankar, V. "Performance prediction through OEE-model", *International Journal of Industrial Engineering and Management* 11, pp. 93 – 103, **2020**. DOI: 10.24867/IJIEM-2020-2-256
- [15] Lepenioti, K., Pertselakis, M., Bousdekis, A., Louca, A., Lampathaki, F., Apostolou, D., Mentzas, G., Anastasiou, S. "Machine learning for predictive and prescriptive analytics of operational data in smart manufacturing", *Advanced Information Systems Engineering Workshop*. Springer International Publishing, Cham, pp. 5 – 16, **2020**. DOI: 10.1007/978-3-030-49165-9_1
- [16] Okpala, C. C., Anozie, S. C., Mgbemena, C. E. "The optimization of overall equipment effectiveness factors in a pharmaceutical company", *Heliyon* 6 (4), pp. 1 – 9, **2020**. DOI: 10.1016/j.heliyon.2020.e03796
- [17] Souza, B. V., Santos, S. R. B., Oliveira, A. M., Givigi, S. N. "Analyzing and predicting Overall Equipment Effectiveness in manufacturing industries using machine learning", In *Proceedings of the 2022 IEEE International Systems Conference (SysCon)*, Montreal, QC, Canada, 25–28 April, pp. 1 – 8, **2022**. DOI: 10.1109/SysCon53536.2022.9773846
- [18] El Mazgualdi, C., Masrour, T., El Hassani, I., Khoudi, A. "Machine learning for KPIs prediction: A case study of the overall equipment effectiveness within automotive industry", *Soft Computing*. 25, pp. 2891 – 2909, **2021**. DOI: 10.1007/s00500-020-05348-y
- [19] El Mazgualdi, C., Masrour, T., ElHassani, I., Khoudi, A. "Using machine learning for predicting efficiency in manufacturing industry", In *Advanced Intelligent Systems for Sustainable Development*, Springer International Publishing, Cham, pp. 750 – 762, **2020**. DOI: 10.1007/978-3-030-36671-1_68
- [20] Imane, M., Aoula, E. S., Achouyab, E. H. "Using Bayesian Ridge Regression to predict the Overall Equipment Effectiveness performance", In *Proceedings of the 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, Morocco, 3–4 March 2022; pp. 1 – 4, **2022**. DOI: 10.1109/IRASET52964.2022.9738316
- [21] Imane, M., Aoula, E. S., Achouyab, E. H. "Support Vector Regression to predict the Overall Equipment Effectiveness indicator", In *Proceedings of the 2022 International Conference on Intelligent Systems and Computer Vision (ISCV)*, Fez, Morocco, 18–22 May 2022; pp. 1 – 5, **2022**. DOI: 10.1109/ISCV54655.2022.9806111

- [22] Dobra, P., Jósvali, J. “Cumulative and rolling horizon prediction of Overall Equipment Effectiveness (OEE) with machine learning”, *Big Data and Cognitive Computing* 7, pp. 1 – 16, **2023**. DOI: 10.3390/bdcc7030138