Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 107 (2022) 1204–1209

www.elsevier.com/locate/procedia

55th CIRP Conference on Manufacturing Systems

Tool failure recognition using inconsistent data

Júlia Bergmann[a,b], Klaudia Éva Zeleny[c,d], József Váncza[a,e], Andrea Kő[d]

[a] EPIC Center of Excellence in Production Informatics and Control, Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH), Budapest, Hungary
[b] Eötvös Loránd University, Doctoral School of Informatics, Budapest, Hungary
[c] EPIC InnoLabs Nonprofit Ltd., Budapest, Hungary
[d] Corvinus University of Budapest, Budapest, Hungary
[e] Budapest University of Technology and Economics, Dept. of Manufacturing Science and Technology, Budapest, Hungary

* Corresponding author. *E-mail address:* bergmann.julia@sztaki.hu

Abstract

Data is everything – at least this is one of the main messages of the ongoing industrial revolution. Manufacturing companies all over the world are expanding their digital infrastructure and knowledge on data analysis in the hope of increasing their KPIs with the help of artificial intelligence (AI). Although several well-designed data-driven solutions are available, the most crucial part, data preparation is still not fully supported. In this paper a framework is presented for processing sensor data of machining processes with variable cycle times in an unstable environment. Traditional and novel AI algorithms are tested on the data of a vulcanization process from the automotive industry, namely from tire manufacturing's curing phase. The process in question consists of several subprocesses, and the quality of curing is mostly dependent of the status of a specific type of machine tool. Conventional methods (e.g., examining the cured product manually) are currently used for failure recognition, however the examination is only feasible after a long delay due to the extreme level of heat, which leads to unnecessary and unwanted scrap production. Therefore, a more sophisticated and complex approach is required to increase quality score. A combination of mathematical methods is proposed combining t-SNE feature representation, convolutional neural network, and linear programming optimization. The model highly relies on the tool's continuous degradation characteristics. The threshold for the given binary classification is set by maximizing the accuracy of the detection model. The main contribution of the research is the method of inconsistent sensor data manipulation which supports a unique combination of AI models for early failure recognition.

Keywords: Type your keywords here, separated by semicolons ;
artificial intelligence; failure detection; data preparation; t-SNE; deep insight; linear programming;

1. Introduction

The evolution of manufacturing systems is driven by the *Adaptive Cognitive Manufacturing System* (ACMS) paradigm which emphasizes the role of cognitive faculties even in the highly digitized ecosystem of manufacturing [5]. No doubt, industrial data analytics, processing of big data by machine learning and visualization techniques have an immense potential in supporting or re-constructing human decision-making in a number of distinct fields of manufacturing, albeit the generic challenges are many [4, 19]. Along with the broadly and in-depth discussed issues of the 5V (i.e., volume, velocity, variety, veracity, and value) and security, which all are critical in manufacturing applications [6], here we would like to emphasize yet another aspect: knowledge distilled in any form from massive data sources should comply with the models and preliminary background knowledge of the technical (and business) system at hand. For the time being, just those algorithmic methods which are able to work over large datasets are hardly capable of considering – let alone exploiting – the available engineering background knowledge [8]. Furthermore, non-representable engineering knowledge, experience and intuition often provide very successful insights and heuristic solutions. Hence, capturing the human factor must be integral in any data-driven model development and adaptation.

The focus of this work is set to quality control in *tire manufacturing*, and in particular to the vulcanization process. The industrial motivation of our research is provided by the problem of *detecting leakage* in a tool – the so-called bladder – which is

2212-8271 © 2022 The Authors. Published by Elsevier B.V.

10.1016/j.procir.2022.05.132

used in one of the last stages of tire production. Wrong bladder may incorrigibly deteriorate the quality of the final product. The routine and most reliable solution of this problem is still based on manual inspection and human sensing (see section 2). Since explicit knowledge is hardly available in this problem domain, but data is generated continuously in real-time, and experience of previous quality checks are accumulating, it is expedient to apply data analytics in combination with the basic background knowledge of the production technology.

The application of machine learning (ML), and more recently, big data analytics have been suggested in quality control for quite a time (see [15], and [21]). Broad-sweeping reviews as well as generic frameworks designed for industrial data analytics point to fault diagnostics and predictive analytics as some of the most promising application fields of the novel techniques [4, 13, 14, 12]. Here, dealing directly with physical processes, one should expect that data comes from a highly nonstationary environment, from various and typically asynchronous data sources, burdened by noise and inconsistencies. Hence, it is a challenge how to adapt advanced ML techniques performing extremely well in other domains such as visual recognition or text processing [4, 12] To the best of our knowledge, the above targeted application of ML is still without preliminaries. E.g., using the terms ”machine learning” and ”tire production” in a Scopus database search resulted only in six articles, all related to different areas of production.

Recently, machine learning methods have been developed for *indirect tire pressure monitoring* [20]. Here, the loss of pressure could reliably be detected by a combination of decision tree and support vector machine techniques, using speed and vibration characteristics in the time and frequency domains. The predictor was suggested as a low-cost, redundant backup to expensive safety-sensitive sensor-based solutions. Tire performance prediction by means of various regression techniques was investigated in [7]. Note that in both studies the best method emerged from a numerical comparison of alternative ML techniques.

In what follows, after presenting the industrial background of tire manufacturing (section 2), we expose the specific research problem (section 3) and present in detail our research methodology (section 4). Experimental results are summarized in section 5, and conclusions including future work are discussed in section 6.

2. Industrial background

Numerous factories produce pneumatic tires around the world according to relatively standardized methods and machinery. Tires are a complex combination of various elements which require a wide range of ingredients. Fig. 1 summarizes the main steps of tire production, in which curing (vulcanization) is a key step followed by final quality check.

Tire manufacturing initializes with *processing the raw materials* namely natural or synthetic rubber, chemicals, steel and textiles. The composition of these raw materials varies depending on the size, flexibility, grip and resistance required. Next, various tire *components* such as fabric and steel cords, bead

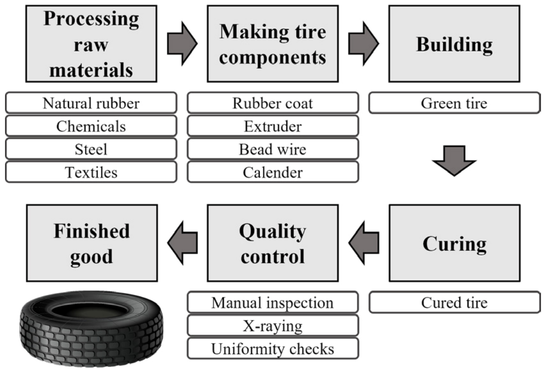

Fig. 1. Main phases of tire production.

wires, calender and extruder are manufactured. The next stage is the *assembly* when the tire is built from these components in a *tire building machine*. The result of this stage is the so-called *green tire*, which is still a flexible tire without any groove on the outer surface. In the following, *curing or vulcanization* stage, heat and pressure is applied to the green tire in order in form it into its final shape. During this stage, the green tire is placed by a robot into a mold inside the curing press machine and after a short idle time the curing process starts. A rubber bladder is inserted into the green tire and as it is filled with steam or gas it starts to inflate. The inflated bladder presses the green tire to the mold, taking on the tread pattern and sidewall lettering. Heating and annealing schedule, as well as processing time depend on the required features of the tire. The curing time of a standard green tire is between 10-15 minutes. The final stage is *quality control* which is in the focus of our current study.

After curing is complete, the tire is removed from the mold for cooling and then testing. Each tire is inspected for flaws such as bubbles in the rubber of the tread, sidewall, and interior. The most typical cause of flaws is a wrong, leaking bladder. This tool has a planned lifetime after which it is replaced, but sometimes it starts to leak before unexpectedly. When leakage happens, so-called pinholes or micro leaks appear on the inner side of the tire. Because of the homogeneity of its surface and the dark color of the tire the testing cannot be made by visual data processing and well-proven AI algorithms. Testing has to be done by human workforce, manually, by relying on *tactile sensing*. However, during the vulcanization process tires are heated up to 200 °C, hence manual inspection can be performed only after an hour. During this time, 4-6 more tires are being cured with the leaked bladder therefore scrap is produced. The goal of scrap production reduction motivates the current work to detect and tell by the end of a curing cycle whether the bladder is leaked or not.

3. Problem statement

Instead of putting the product to delayed, expensive and tedious manual testing, our goal is to anticipate the result of the quality test from production related process and sensory data. Of course, there is a dilemma: indirect testing of the product involves risks, but, on the other hand, it can be done without delay, hence the use of a wrong tool (i.e., bladder) can be averted. If false negative decisions can reliably be avoided, then false positive decisions can still be followed by manual testing. Hence, resulting in less scrap, increased yield, shorter cycle times and less demanding working conditions, the overall impact of the automatic quality check can be positive.

All in all, our problem statement is as follows:

- *Given* basic parameters of tyres, time-stamped information on the status of tyre building machines, sensory data of the cyclic production processes, as well as records of the quality checks of the tyres made during the respective cycles,
- *Find* a predictor that indicates by using solely the product and real-time process data if the risk of using a wrong bladder exceeds a threshold,
- *Such that* prediction accuracy of human inspectors is approached while false negative decisions are minimized.

This is a high-dimensional, non-linear problem containing mixed-type (i.e., numerical and categorical) features and multiple datasets. The bulk of the data is so-called data-in-motion [4] generated in the course of a cyclic manufacturing process. In this problem domain, subsequent data records attached to the same builder machine are strongly related in time. Hence, when looking for useful patterns in historical data records, this neighborhood or locality is to be exploited. This assumption gives a hint as for which ML methods to investigate as candidates. Secondly, it is a technologically feasible hypothesis that a bladder, once leaking, cannot undergo a self-healing process. Hence, the probability of leakage in a bladder can but increase in time, let it be generated by any method.

4. Methodology

This research is based on data collected by sensors located in several curing press machines in a tire manufacturing plant. To systematically carry out the research steps, we followed the CRISP-DM methodology [2], adapted to data-rich applications in manufacturing [10], as shown also in Fig. 2 below:

1. *Business/domain understanding* focuses on understanding the research objectives and requirements from a business/domain perspective, which we detailed in section 1.
2. *Data understanding* starts with an initial data collection and proceeds with steps to get familiar with the data, to identify data quality problems, to discover first insights into the data, as discussed in section 4.1.
3. *Data preparation/manipulation* covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Section 4.2 is summarizing these activities.
4. *Modeling* selects and applies the appropriate modelling techniques, determines the key parameters and calibrates them to optimal values. In our case, models are presented in section 4.3.
5. *Evaluation* assesses the validity of results and reviews the steps executed to construct the model, to be certain it properly achieves the business objectives. ML model evaluation is discussed in section 5.
6. *Deployment* embeds the model into its application environment. The current work does not include deployment challenges.

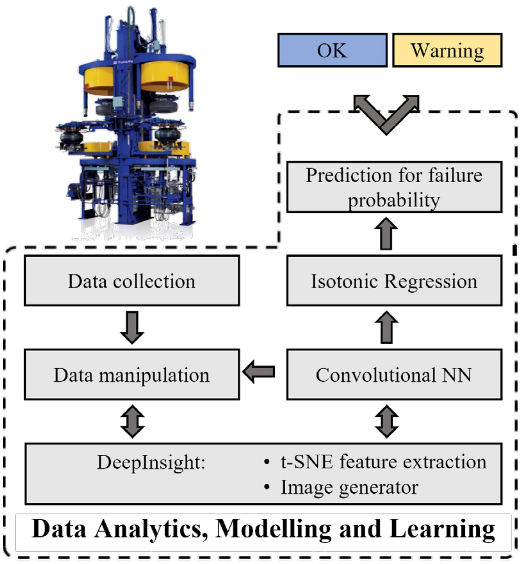

Fig. 2. Workflow of the analytical and learning process.

4.1. Data understanding

Data is generated continuously in real time from three kinds of sources: (1) the controllers of the curing presses (registered in the Manufacturing Execution System), (2) the status register of the machines, as well as (3) sensors attached to the machines. Each data record is timestamped, but the sources are not synchronized. In a harsh physical production environment like this, data transmission delays, missing data, outliers are all common.

First, there are data available from production. This data source provides information on cycle level, where each cycle is characterized by the following:

- The cycle start timestamp.
- The exact press machine ID on which the cycle was running.
- The bladder ID the cycle was made with.
- The type of tire cured in the cycle.

Secondly, there is data on the *status* of each curing press machine. This is a simple timestamped binary (on/off) data indicating the start of the curing process.

Third, data is collected also by physical *sensors*. Altogether there are five sensors in each machine which collect data continuously, measuring the *temperature* and *pressure* in different locations inside the press. Each sensory data record contains 3 attributes: a timestamp, the sensor name and the measured value.

It is important to highlight that the beginning of a cycle and beginning of a curing process are not exactly same. The cycle includes other auxiliary activities besides the curing process, such as application and removal of the tire or bladder change.

4.2. Data manipulation

Merging the production and sensor data together and sorting by timestamp allows to split the sensor data by cycle, i.e. to collect all the corresponding observations during the curing process. Since also the length of each intermediate step is also known, the sensor data between each cycle can be broken down into further steps. After this transformation a list of the measured sensor values for each sensor type by press, cycle and step is created. Figure 3 presents the structure of the raw data and the data after this transformation.

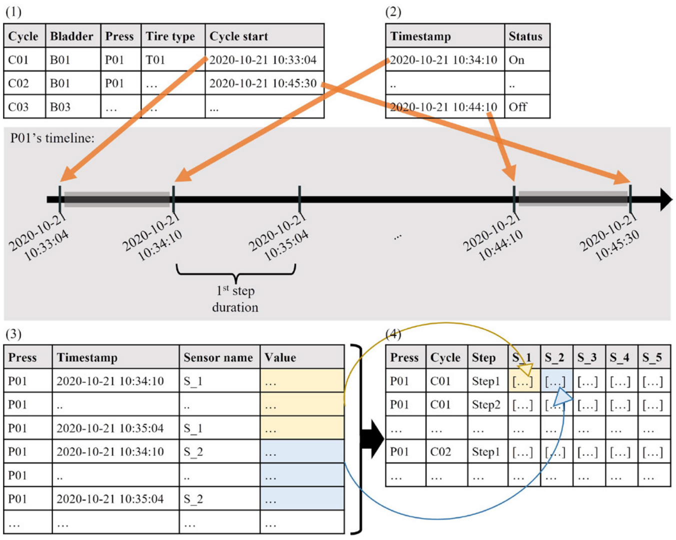

Fig. 3. Structure and relations of data available. The given tables are: (1) production data, (2) process status data and (3) sensor value data. Table (4) contains lists of sensor values of a specific cycle separated by intermediate process steps.

The lists defined by this way do not have the same length (because of varying cycle times and missing sensor logs), therefore several statistics are calculated to characterize the measured temperatures and pressures in each step. The chosen statistics are: *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*. The goal is to detect the failure (leakage) at the end of each cycle therefore a last transformation is required to turn the data into a certain structure:

- **features**: A cycle ID, and the corresponding information following from production data, such as press machine, type of tire etc. and the statistics for each sensor and step, such as the average internal temperature in step 1, the minimum internal temperature in step 2 and so on.
- **label**: A binary variable that indicates if there was a leakage in that cycle.

All in all for each cycle there are around 650 features, since there were 5 sensor types, 16 statistics and 7 or 8 steps for each cycle.

4.3. Modelling

DeepInsight

Due to the large number of features, it is reasonable to use some dimensionality reduction technique. On the other hand neural networks are widely used and have promising results on image-related problems. Combining so-called t-SNE feature reduction and convolutional neural networks (CNNs) is a powerful tool called *DeepInsight*, published in details in [17]. The idea is to generate point clouds from standard tabular data based on their similarities and then feed a CNN with these generated images. The architecture of the trained CNN is presented in Fig. 4.

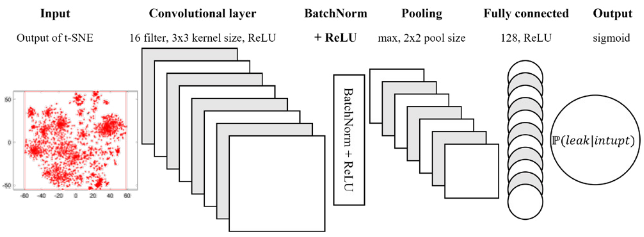

Fig. 4. A proposed architecture of the CNN model inspired by [17].

Isotonic regression

Isotonic regression or monotone regression is a method of fitting the closest monotone increasing function over a set of observations [1]. Practically speaking, isotonic regression can be formalized as the solution of a small quadratic programming model:

$$\text{minimize} \sum_{i=0}^{N} (x_i - \hat{x}_i)^2 \tag{1}$$

$$\text{subj. to: } \hat{x}_i \geq \hat{x}_{i-1} \qquad \forall i = 1..N \tag{2}$$

The objective function (1) is the sum of element-wise differences between the given observations and the fitted values, therefore minimizing this sum returns the closest non-decreasing function. The monotony is insured by constraint (2).

In the current work isotonic regression is used as a refinement of the predicted values. As the last layer of the CNN structure (Fig. 4) is a single perceptron with a sigmoid activation function, the output is a probability estimation of the failure event. Assuming that the probability of the failure of the same tool cannot drop overtime, it is possible to perform isotonic regression on the series of these predicted probabilities in order to eliminate error terms of the prediction.

5. Evaluation and discussion

In this section six classical binary classification models are compared through different approaches to the proposed mixture of the DeepInsight technique and isotonic regression. The classical models are logistic regression, k-neighbors, decision tree, random forest, XGBoost and LigtGBM [3, 11, 16]. The data manipulation process is the same for all cases therefore all models have the same normalized input. A test set is separated before the learning process, all the ML models are trained on the same dataset and tested on a different but common dataset.

The comparison of several models is not a strictly defined process, however it is quite common to start by checking basic metrics such as the precision, recall, balanced accuracy and F1-score [9]. Due to confidentiality reasons, the concrete values cannot be shown, however Fig. 5 shows the relationship of those metrics of the compared models. All the groups are scaled by the maximum value in the metric group. This way the models can be compared as the model with the best performance will have the highest relative value (which is 1), and all the others will show lower relative metrics proportional to the original difference between the metric scores.

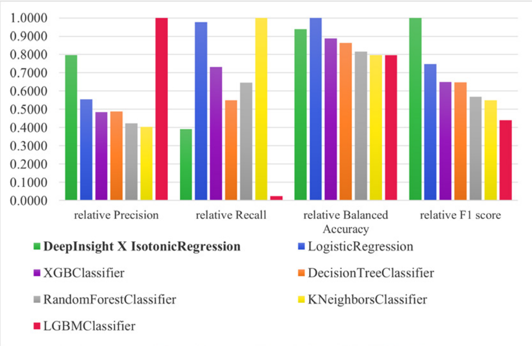

Fig. 5. Relative precision, recall, balanced accuracy and F1 score of the compared models.

In real life use cases the absolute winner rarely exists and the current situation shows the same. The proposed method (*DeepInsight X IsotonicRegression*) performs among the firsts except for in case of recall. While the LightGBM model has the best precision value, it has quite poor performance concerning the other metrics. In situations like this it is one option to bring in more evaluation metrics into the analysis, then note the ranks of the models for the chosen metrics [18]. The rank is simply the index of the model in the decreasingly sorted vector of the metric values (Fig. 6). The row means of the rank matrix returns the average rank of each model which is shown in Fig. 7. This ordering of the models includes the information of several evaluation metrics all of which have different meanings, therefore combining these allows us to retrieve a better picture of the overall performance of each model.

One of the accuracy metrics, namely the F1 score, is already a combination of two other metrics: it is the harmonic mean of recall and precision. Its more generic version is the F_β score which values its two components differently. β is an arbitrary positive real number meaning that recall is β times as important as precision. Fig. 8 shows the relative F_β scores of the compared models for different β values. Again the real values are scaled by the maximum of all F_β scores.

Since the comparison of multiple ML models is not carved in stone, an other nontraditional method is proposed now. Take the false negative (FN) and the false positive (FP) counts from the confusion matrix of the models. Scale down these counts by the minimum of FNs and FPs respectively, and let us call the new values relative false negatives (RFNs) and relative false positives (RFPs). By doing so the scaling difference between FNs and FPs disappears. RFN and RFP technically denotes the multiplier between the given model's FN and FP and the best possible FN and FP counts. Fig. 9 plots the RFN - RFP pairs of the compared models.

	ACC	PPV	TPR	TNR	NPV	BAL ACC	F1
DeepInsight X IsotonicRegression	2	2	6	2	4	2	1
Logistic Regression	6	3	2	6	1	1	2
XGB Classifier	4	5	3	4	2	3	3
Decision Tree Classifier	3	4	5	3	5	4	4
LGBM Classifier	1	1	7	1	7	6	7
Random Forest Classifier	5	6	4	5	6	5	5
Kneighbors Classifier	7	7	1	7	3	7	6

Fig. 6. Rank matrix
Ranks of the compared models concerning different metric scores: accuracy, positive predictive value (precision), true positive rate (recall), false positive rate, negative predictive value, balanced accuracy and F1-score. Lower rank (greener shade) signifies higher value therefore better performance.

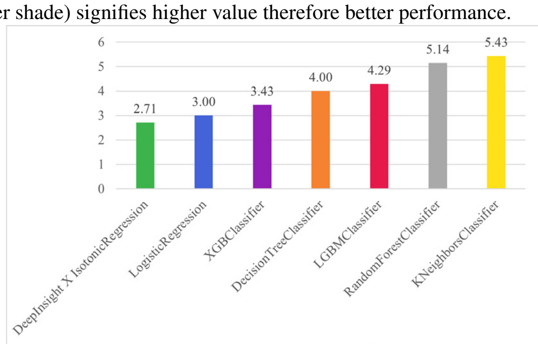

Fig. 7. Average ranks of the compared models based on the results of the rank matrix (Fig. 6).

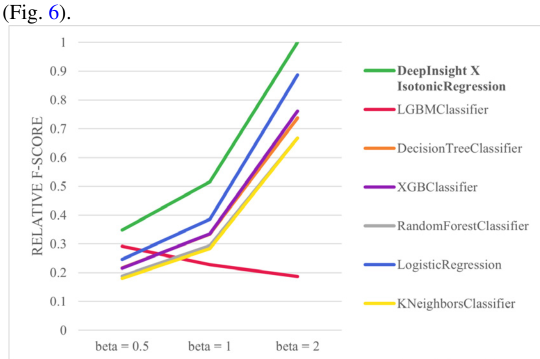

Fig. 8. Relative F-scores for all $\beta \in \{0.5, 1, 2\}$ of the compared models.

In an ideal world there would be one best model with RFN = RFP = 1 i.e. one model would minimize both the FNs and FPs. However real world scenarios are not so straightforward, so an intuitive idea is to sort the models based on their Euclidean distance from the ideal $(1, 1)$ point. This distance tells how far is the given model from both the best FN and FP counts.

Based on the discussion in this section and the results shown in the referred figures, it is safe to conclude that the proposed method in section 4 outperforms the other ML models in the comparison. However the absolute best model does not exists as

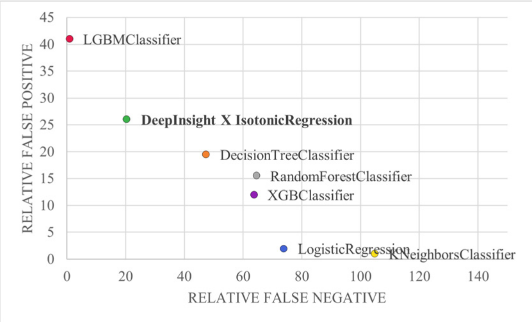

Fig. 9. Relative false negative (RFN) (and positive (RFP)) values
RFN and RFP values are the false negative (and positive) counts normalized by the minimal false negative (and positive) amount.

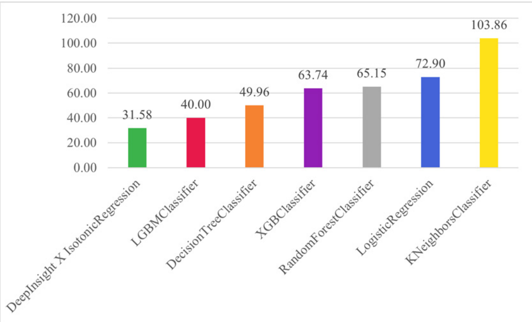

Fig. 10. Euclidean distance from the hypothetical (1, 1) point on the RFN-RFP map (Fig. 9).

(based on the selected metrics' outcomes) some models precede others from one point of view but drops behind from another.

6. Conclusion and future work

The paper suggested a method to transfer well-proven ML techniques to the in-line quality check of vulcanized tyres. The method is generic as far as the processing of data coming from cyclic manufacturing are concerned. Albeit the results for indirect tool monitoring are promising, their routine industrial application requires future work in two directions: (1) false negative predictions should further be decreased, and (2) the method has to be embedded into a *manufacturing big data ecosystem* [4], where the generation and secure storage of data, date processing like analytics and visualisation in particular, systematic update and management of the results, as well as their inclusion into the standard quality control workflow are seamlessly integrated. All this requires an intensive collaboration of the industrial and academic partners. We will also evaluate other, more complex CNN structures and even complement them with transfer learning, too. An alternative research path is to formalize a method for ML model comparison in ambiguous situations.

7. Acknowledgement

This research was supported by the Doctoral Student Scholarship Program of the Co-operative Doctoral Program of the Ministry of Innovation and Technology, financed by NRDIO. The authors also acknowledge the support of the EC EPIC grant No. 739592, and the Hungarian the ED18-2-2018-0006 grant.

References

[1] Barlow, R.E., Brunk, H.D., 1972. The isotonic regression problem and its dual. Journal of the American Statistical Association 67, 140–147.

[2] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R., 1999. The CRISP-DM user guide, in: 4th CRISP-DM SIG Workshop in Brussels in March.

[3] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., et al., 2015. Xgboost: extreme gradient boosting. R package version 0.4-2 1, 1–4.

[4] Cui, Y., Kara, S., Chan, K.C., 2020. Manufacturing big data ecosystem: A systematic literature review. Robotics and Computer-Integrated Manufacturing 62, 101861.

[5] ElMaraghy, H., Monostori, L., Schuh, G., ElMaraghy, W., 2021. Evolution and future of manufacturing systems. CIRP Annals 70, 635–658.

[6] Gao, R.X., Wang, L., Helu, M., Teti, R., 2020. Big data analytics for smart factories of the future. CIRP Annals 69, 668–692.

[7] Gutiérrez-Gómez, L., Petry, F., Khadraoui, D., 2020. A comparison framework of machine learning algorithms for mixed-type variables datasets: A case study on tire-performances prediction. IEEE Access 8, 214902–214914.

[8] Gödri, I., Kardos, C., Pfeiffer, A., Váncza, J., 2019. Data analytics-based decision support workflow for high-mix low-volume production systems. CIRP Annals 68, 471–474.

[9] Hossin, M., Sulaiman, M.N., 2015. A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process 5(2), 1–11.

[10] Huber, S., Wiemer, H., Schneider, D., Ihlenfeldt, S., 2019. DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. Procedia CIRP 79, 403–408.

[11] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems 30, 3146–3154.

[12] Kozjek, D., Vrabič, R., Rihtaršič, B., Lavrač, N., Butala, P., 2020. Advancing manufacturing systems with big-data analytics: A conceptual framework. Int. Journal of Computer Integrated Manufacturing 33, 169–188.

[13] Krauß, J., Frye, M., Beck, G.T.D., Schmitt, R.H., 2019. Selection and application of machine learning- algorithms in production quality, in: Beyerer, J., Kühnert, C., Niggemann, O. (Eds.), Machine Learning for Cyber Physical Systems, Springer Berlin Heidelberg. pp. 46–57.

[14] Lee, K.T., Lee, Y.S., Yoon, H., 2019. Development of edge-based deep learning prediction model for defect prediction in manufacturing process, in: 2019 International Conference on Information and Communication Technology Convergence (ICTC), IEEE. pp. 248–250.

[15] Monostori, L., Markus, A., Van Brussel, H., Westkämpfer, E., 1996. Machine learning approaches to manufacturing. CIRP Annals 45, 675–712.

[16] Sharda, R., Delen, D., Turban, E., 2016. Business intelligence, analytics, and data science: a managerial perspective. Pearson.

[17] Sharma, A., Vans, E., Shigemizu, D., Boroevich, K.A., Tsunoda, T., 2019. Deepinsight: A methodology to transform a non-image data to an image for convolution neural network architecture. Scientific Reports 9, 1–7.

[18] Van Erp, M., Schomaker, L., 2000. Variants of the borda count method for combining ranked classifier hypotheses, in: 7th International Workshop on frontiers in handwriting recognition, International Unipen Foundation. pp. 443–452.

[19] Wang, J., Xu, C., Zhang, J., Zhong, R., 2021. Big data analytics for intelligent manufacturing systems: A review. Journal of Manufacturing Systems .

[20] Wei, L., Wang, X., Li, L., Yu, L., Liu, Z., 2021. A low-cost tire pressure loss detection framework using machine learning. IEEE Transactions on Industrial Electronics 68, 12730–12738.

[21] Zhang, D., Xu, B., Wood, J., 2016. Predict failures in production lines. 2016 IEEE International Conference on Big Data , 2070–2074.