



Universiteit
Leiden
The Netherlands

Requirements towards optimizing analytics in industrial processes

Zeiser, A.Z.; Stein, B, van; Bäck, T.H.W.

Citation

Zeiser, A. Z., Stein, B. , van, & Bäck, T. H. W. (2021). Requirements towards optimizing analytics in industrial processes. *Procedia Computer Science*, 184, 597-605.
doi:10.1016/j.procs.2021.03.074

Version: Publisher's Version

License: [Creative Commons CC BY-NC-ND 4.0 license](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3277238>

Note: To cite this publication please use the final published version (if applicable).



The 4th International Conference on Emerging Data and Industry 4.0 (EDI40),
March 23 - 26, 2021, Warsaw, Poland

Requirements towards optimizing analytics in industrial processes

Alexander Zeiser^{a,b,*}, Bas van Stein^b, Thomas Bäck^b

^aBMW Group, Ohmstraße 2, 84030 Landshut, Germany

^bLeiden Institute of Advanced Computer Science, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands

Abstract

Modern production systems are composed of complex manufacturing processes with highly technology specific cause-effect relationships. Developments in sensor technology and computational science allow for data-driven decision making that facilitate efficient and objective production management. However, process data may only be beneficial if it is enriched with meta information and process expertise, reduced to relevant information and modelling results interpreted correctly. The importance of data integration in the heterogeneous industrial environment rises at the same momentum as new metrology techniques are deployed. In this paper, we focus on optimizing analytics, containing data-driven decision making for predictive quality and maintenance. We summarize key requirements for data analytics and machine learning application in industrial processes. With a use case from automotive component manufacturing we characterize industrial production, categorize process data and put requirements in context to a real-world example.

© 2021 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Industry 4.0; Predictive Modelling; Optimizing analytics; Information fusion; Machine Learning; Industrial processes

1. Introduction

Increase of computational power together with technologies like cloud or edge computing and decrease of respective hardware cost within recent years are substantial factors for enabling big data analytics in the industrial environment [24, 2]. Advancements of algorithms and methods to handle dimensionality allow for data-driven decision making and smarter manufacturing. One of the pillars of Industry 4.0 is data generated at high scale for control, monitoring and knowledge discovery related to manufacturing quality [22, 26]. In essence, data itself is only a means to an end and rather a prerequisite. The actual knowledge that is extracted after processing data for any purpose is what creates value for the production system. With goals of zero defect manufacturing (ZDM), maximization of overall equipment effectiveness (OEE) and cost minimization data is what fuels optimization capabilities.

* Corresponding author. Tel.: +49-151-6040-2080.

E-mail address: Alexander.AZ.Zeiser@bmw.de

Machine Learning (ML) has proven its effectiveness in data-driven applications for classification and regression problems [2, 15, 32]. Yet, surveys and literature addressing industrial applications reveal open issues for comprehensive integration of data analytics, ML, data-driven decision making and algorithm based process optimisation into real-world production. Examples are non-stationarity, highly imbalanced data sets, data heterogeneity, as well as quickly changing production conditions, expected decision transparency and operators qualification [19, 11, 12, 33]. In the present paper we take up previous works on challenges of industrial application of ML and provide a framework for categorisation of derived requirements. The core of this paper lies at the characterization of an industrial production environment with respect to predictive modelling. We present an industrial perspective onto developments in data analytics to bridge research and industrial application. Requirements that arise for intelligent manufacturing in terms of data generation, processing and analytics are summarized and put in reference to a real production process. In part 2, we provide background on predictive modelling and introduce the term optimizing analytics. Part 3 outlines key requirements for ML application in industry and refers to current research in literature. Part 4 concentrates on real-world, industrial processes. We introduce a casting process as use case from automotive component manufacturing to contextualize characteristics of series production and refer to requirements towards an optimizing analytics framework.

2. Evolution of data analytics towards optimizing analytics

The evolution of data analytics can be divided into four levels: *descriptive*, *diagnostic*, *predictive* and *prescriptive*. In terms of industrial maturation prescriptive analytics is still less prevalent. However, in literature the differentiation of predictive and prescriptive is indistinct as methods and aspects overlap. Extensions of predictive methods may contain prescriptive objectives, without being explicitly stated as such [25, 30]. A summary of predictive and prescriptive methods is provided in [25]. In our view, the advancement of predictive analytics includes objectives of guidance for the user. We characterize it rather as algorithm-supported, single or multi-objective optimization. From the industrial point of view, this includes the component of automatic optimization of manufacturing parameters or maintenance operations. The term we use in this context is *optimizing analytics* instead of *prescriptive*. The following list summarizes the four levels.

- *Descriptive*: visualization of mainly historic but also real-time (streaming) data; (“What has happened?”)
- *Diagnostic*: first root-cause analysis for events on historic data; (“Why did it happen?”)
- *Predictive*: forecast future events based on patterns in historic and real-time data; (“What will happen?”)
- *Optimizing analytics*: optimization based on predicted events; (“What should be done to optimize the system?”)

In the industrial context two directions emerge from predictive analytics: predictive quality and predictive maintenance [28]. Both focus on the utilization of manufacturing related data to foresee hostile situations, i.e., defects, faults or breakdowns, based on the actual state. Predictive maintenance includes both, diagnosis (current condition) and prognosis (projected future condition). In their review, Nguyen et al. [14] differentiate mainly between model-based, data-driven and hybrid approaches. The model-based approach relies on a deterministic, mathematical model that describes the degradation of a single tool or even the whole system. Data-driven approaches make use of statistical and machine learning techniques. They are prevalent in industry due to comparably easy development, adaptation to new machines and access to huge amount of data needed for application. In our work, we focus on data-driven approaches applied to industrial processes.

3. Machine Learning application in production industry: challenges and derived requirements

Research presents many applications of ML for quality enhancement and maintenance optimization based on predictions from learned data patterns [32]. Objectives are reduction of unplanned downtimes and system faults as well as productivity improvement by waste avoidance in terms of material, human and machinery resources [10]. However, transformation of advancements in computer science from academia to industry is still challenging. One of the main differences between academia and real-world production is the long-term scale and associated changing conditions. Applications under laboratory conditions demonstrate technical capability of methods. In turn, series production is

characterized by a highly multivariate and volatile system in terms of environmental influences and hidden cause-effect relationships. Industrial production needs models that are deployed on-line and capable to react to changing conditions, hence, regularly retrained to keep the needed performance and optimized further. In their industrial survey, Heimes et al. [19] name insufficient data basis as most significant hurdle for successful implementation of data analytics into series production. Most relevant requirements, hence, are stated as availability of process-descriptive data, reliable sensory equipment and sufficient data resolution. Another challenge, that is typically present in industrial production, are highly imbalanced data sets in terms of conformity of produced parts. With all effort and ambition towards a zero-defect manufacturing scrap rates are typically low, e.g. in casting ≤ 1 to 5 % depending on the technology. To realize machine learning models and exploit potentials of optimizing analytics a number of requirements must be met that are constraint by the production process, the organizational framework or by the analytical model. Table 1 presents different requirement categories that are described in detail in the following.

Table 1. Requirements for industrial ML applications.

| Requirement category | Elements | Conditioned by |
|----------------------|---|----------------|
| Data availability | Data is accessible, Data is timely updated [6] | Process |
| Data quality | Usability, Reliability, Readability, Relevance [6] | Process |
| Interdisciplinarity | Integration of domain experts, requirements definition in planning phase [19] | Organization |
| Data integration | Data management, information fusion [18, 23] | Model |
| Interpretability | Model-intrinsic, Post-hoc decision explainability [27, 13] | Model |
| Adaptability | Regular re-training, data drift detection, handling imbalance [16, 21] | Model |

3.1. Production process and organizational requirements

Data availability is the first prerequisite for implementation as the data to monitor and describe a manufacturing process must be principally accessible. Mutually, **Data quality** is important since wrongly placed sensors, inadequate sampling frequencies and faulty data transfer hamper prediction accuracy or even ruin the whole model capability. Low data quality has its causes at sampling level, storage level or processing level and can only be improved here. At model and interpretation level even most advanced techniques will not reach satisfying results if data quality is low. Data quality has several dimensions going beyond pure completeness, such as usability, reliability, readability and relevance [6]. Usability refers to how credible data is in terms of its units and values taken. Reliability summarizes data accuracy, consistency and integrity, that are measures of unambiguity. It refers also to the level of certainty that raw data cannot be manipulated and falsified at a later time. This is especially important to learning from historic data so that extracted patterns and rules are not distorted. Readability refers to the format in which data is represented and stored. Hence, how direct it can be integrated into a data analysis pipeline, that is defined by the structure (e.g. arrays, images, graphs or semi-structured data in XML or JSON format). Relevance refers to how well information content matches information needs. In industrial manufacturing this is related to how well sampled data can describe and therefore model the process. It is closely related to expert knowledge and technical hypotheses for root-cause analysis. Machine learning offers techniques to find hidden and implicit patterns in a vast amount of data. However, the more relevant the data is to the task the more precisely and efficiently a ML model can work. Therewith corresponding is the organizationally conditioned **interdisciplinarity**. The early integration of domain experts allows for better data and problem understanding. Additionally, the early definition of technical requirements for ML application and integration already into process planning and purchasing enables necessary data availability and data quality. This builds a framework resulting in accurate, problem-specific models and better performance during later series production [19].

3.2. Data modelling based requirements

Another requirement for successful implementation in industrial processes is **interpretability** (also referred to as explainability) of ML model results. In the manufacturing context of quality and process improvement the sole classification task of non-/conformity (binary) or defect type (multi-class) is not yet sufficient. Objectives are to derive respective measures, like machining parameter and maintenance operation optimization to avoid bad quality and

machine failures. To do so, results and decisions made by ML models must be comprehensible. There exist different approaches to allow more interpretation capabilities, like model-intrinsic and post-hoc interpretation [13, 27]. Model-intrinsic approaches incorporate interpretability directly into the design by (a) imposing constraints to make a model simpler, e.g. sparsity terms or decision tree pruning [13], by (b) applying mimic learning, i.e., approximating statistical properties of a complex model by a simpler, interpretable model, e.g. decision tree trained by a deep neural network to mimic input-output relations [5], or (c) by attention mechanisms that visualize where a deep learning model focuses on the input, e.g. images or time series, to discriminate and to make predictions [34, 15]. Post-hoc approaches are based on algorithms that are applied subsequently, after learning and prediction. These include feature importance ranking by measures like accuracy gain for traditional, feature-engineering-based ML algorithms, e.g. decision trees or logistic regression [13], and attribution methods like perturbation or back-propagation to calculate a gradient function [1].

Nonetheless, differentiation must be made between ML researchers and industrial end users. The presented approaches refer rather to decision logic in mathematical and statistical terms. For the industrial end user interpretability is directly connected to the expert domain. Model results are expected to serve as recommendations for improvement. Referring to automation of tasks the output of a ML model is input for execution of quality control and maintenance scheduling. Interpretability is indispensable for decision verification and trust into self-optimizing frameworks.

With the objective of data-driven decision making **data integration** and reduction to relevant attributes are fundamental. The industrial system presents a huge variety of sensors from multiple sources, process input variables and human interaction along the value-added chain. In this context information fusion is an important task to perform. The intention is to combine different sources and types of data to enhance the information content and reduce uncertainty and redundancy with respect to their individual processing [18]. Information fusion, also referred to as data fusion, is a multi-disciplinary field that implies wide-ranging research. Our objective is not a comprehensive overview of techniques but to introduce concepts in the context of our research.

Basically, three approaches of information fusion are discussed in literature: at (raw) *data level*, at *feature level* and at *model level* [18, 23]. Fusion at (raw) data level refers to, typically, timely or spatial synchronization, e.g. time series data sampled at different frequencies, and rather uniform data in terms of descriptive nature. An industrial example is the fusion of pressure signals in injection molding at different mold locations. In contrast, at the higher feature and model levels information from also mixed type data of very different structure is fused. Feature level fusion follows the objective to extract features from raw data separately and combine it into one feature vector. Common techniques are such as Canonical Correlation Analysis [18], deep learning based on convolutional neural networks (CNN) or multi kernel learning [20, 4]. On model level outputs of several models that process respectively one data type are combined. This approach is also called decision level fusion since, e.g. in classification tasks the final decisions of class affiliation per sample are compared. Techniques used are majority voting, known from ensemble learning, or inference based Bayes networks [18]. A comprehensive overview of fusion techniques can be found in [8].

Within machine learning research ensemble learning and hybrid methods are developments to improve robustness and performance over singular models. Dealing with heterogeneous data, especially hybrid methods can be advantageous due to integration of dedicated models. However, the interpretation of 'hybrid' is left to the developer of such method as there does not exist a clearly defined research area under this term [3]. We, therefore, propose a categorization for the hybrid machine learning framework: (a) Sequential, (b) Parallel and (c) Expert augmented.

- (a) The sequential framework refers to an architecture where different tasks, i.e. data fusion, feature extraction, classification or regression, are performed in a series of steps. Separate models are applied per dedicated task along the data processing pipeline. This type can be found in [31, 7, 23], where feature extraction and subsequent classification are performed by different models. Information fusion takes place at a feature level.
- (b) The parallel framework describes separate data processing pipelines, each handling a certain data type. Hence, dedicated capabilities of ML models are exploited. The information fusion takes place at model level by combining outputs into one single decision [3].
- (c) The expert augmented framework introduces another form of hybridization. Here, an interaction unit is complementing the ML pipeline to enhance prediction performance. Therein, human expert knowledge is extracted to supervise model training and validation in cases where target variables are missing and only unsupervised ML is applicable [9]. Another purpose is an improved generalization of the ML model when applied to various similar problems but with sparse data availability [17].

Another field of study that aims at integrating mixed type data is *multimodal ML*. This term is less common in industrial research but rather used in other disciplines like medicine, bio-informatics and robotics where perception tasks or sentiment analyses are performed. Modalities refer to the nature of how something is experienced. In terms of human sensation this includes vision, acoustics and touch [4]. On a technological level sensory capabilities can be abstracted to a modern production system with sensor networks, optical measurements and communication protocols. Hence, also the industrial environment can be defined as multimodal problem, where information fusion must be performed to advance data-driven decision making.

4. An industrial perspective from automotive component production

Typically, within production research a production system is characterized by its nature of process (discrete or continuous) and its way of volume creation (single, batch, mass). Further differentiation is then related to variety and organizational characteristics (job shop, manufacturing cell and flow line). From the perspective of predictive analytics another distinction may be introduced that refers to the level of consideration within the production system and scope of process optimization: (a) *machine level*, (b) *direct value chain level* and (c) *cross-process level*. Based on the level motivation and objectives of predictive analytics may change. However, levels partly build on one another.

- (a) On *machine level* one direct manufacturing procedure is considered. An immediate quality control allows for direct feedback and labelling. Data as generated by or directly related to the machine is used as input for predictive analytics. Optimization of parts quality and maintenance operations are likewise related to that machine.
- (b) On a *direct value chain level* more than one manufacturing procedure is considered but a quality check is only implemented at the end of a process section. In this case, early anomaly detection and quality prediction allow for prompt discharging of non-conform parts to save further value adding resources.
- (c) The *cross-process level* represents the combination of manufacturing and assembly. Different parts, either manufactured internally or externally, flow into the process to form an assembly. The generated data is highly heterogeneous, in terms of source, type and physical nature. A key aspect is the connection and relation between stations since the outcome of one step may influence the settings of a following. For information fusion and predictive analytics this level entails high complexity. However also optimization potential is high since real-time data processing and prediction allows for anticipatory adjustments.

4.1. Use case - Casting of engine components

Fig. 1 presents value and data stream of a typical casting process of engine components that can be summarized into casting, mechanical processing and quality control. Since the manufacturing is in focus here we do not go into detail with subsequent processes within engine and vehicle plant. The process refers to chill casting of an electric engine housing. Several sand cores are placed into the die to enable complex internal geometries, e.g. a water cooling jacket for the final engine. The melt is poured in with an injector. By a complex system of heating and cooling channels the solidification is regulated. Since risers and sand are still present a set of mechanical processing steps follows after casting, i.e. sawing, milling, further cooling, de-coring and washing. The processed casting then enters the quality control where through computer tomography abnormalities are detected. If non-conformities are repairable castings are reworked off-line and brought back into the value stream. If the casting was evaluated as quality-conform it leaves the foundry and is sent to the next step within the value stream. After heat treatment precision machining is performed within the engine plant before the engine is assembled. As final step the engine is assembled to powertrain and vehicle.

Each manufacturing step generates a set of data, either referable directly to the casting produced, i.e. *process monitoring data* and *process output data*, or referable only indirectly, i.e. *inter-cycle data*. Table 2 gives an overview of the process data. The process monitoring data consists of about 150 time series from sensors placed at various locations in and around the mold as well as casting unit and devices. During mechanical processing data from sawing machines, vibrating plates and hammers are sampled. Static variables refer to various measurements at defined control instants in the process. The process output data consists of the components' quality appraisal data, i.e. CT images and derived defect type, location and size. Outcome of assessment is the respective label of OK (conform) or NOK (non-conform) per casting produced. Inter-cycle data does not describe the manufacturing process in-situ but defines

fundamental production conditions in terms of recipe and machine settings, tool repairs and cleaning. It contains valuable information about how the component is manufactured, how process data is sampled and how to interpret it. We also refer expert knowledge to this category that is further distinguished into implicit and explicit knowledge. Especially in its implicit form knowledge is difficult to include in data analytics and ML models since it is not tangible and dependent on single operators.

Table 2. Manufacturing process data.

| Data category | Representation | Process relation |
|---------------------------|--|------------------|
| Time series | Mold temperatures, filling velocity, cooling flow rate, axes feed rates, current | Monitoring |
| Static variables | Control measurements at process instants (e.g. step durations, min/max, position casting unit) | Monitoring |
| Image data | Grayscale 3D scan | Output |
| Quality data | Conformity label, defect type, location, size | Output |
| Meta data | Tool IDs, recipe, machine settings, sampling frequencies | Inter-cycle |
| Maintenance documentation | Tool life, exchanges, repair, cleaning | Inter-cycle |
| Expert knowledge | Implicit (experience), explicit (instructions, process documentation) | Inter-cycle |

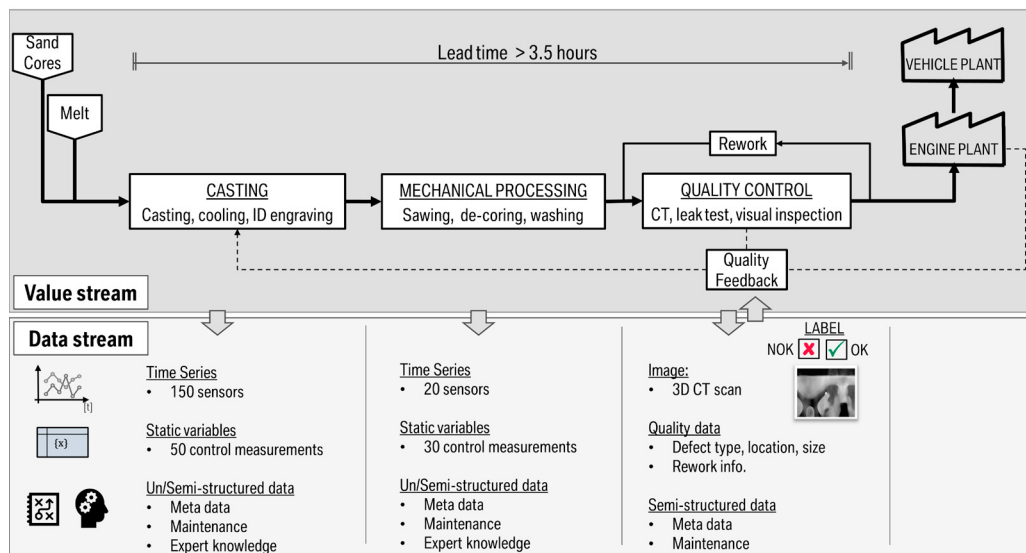


Fig. 1. Value and data stream of an engine component casting process.

The difficulty for quality appraisal in casting methods is that human operators can only assess external and surface defects. Due to the components' complex internal geometry and nature of the process, i.e. change of phase, a full inspection is only feasible with imaging test methods. Computer tomography or X-ray allow insights into the internal metallic structure to find porosities or shrink holes. The process from start of casting to the final appraisal has a lead time of about 3.5 hours. More precisely, a casting produced receives all the value adding steps before its quality is evaluated. Additionally, quality feedback and derived reactions to defects are delayed by the lead time to the next quality control step. The dashed line in Fig. 1 represents this feedback, both from internal quality control and subsequent engine plant. The engine plant as internal customer may detect further defects during fine machining. If defects are systematic or recurring, many more castings may already be affected since the cycle time here is about 170 seconds. This exposes the potential and need for early, real-time quality prediction. In casting of aluminum, typically, material cost are much lower than value-adding process cost since materials can be re-melted and sand reused for new cycles. Data-driven quality prediction moves the appraisal upstream, enables direct discharge and avoidance of further value-adding. Including process parameter optimization the optimizing analytics framework facilitates end-to-end

data-driven quality control. The monitoring of processes by in-situ sampled variables is the evident, straightforward way for quality control. However, the origin of occurring defects may not be that easily identifiable. In a complex system of controllable and uncontrollable factors it is a composition of unfavorable instances that lead to abnormal machine state and finally to a defect or machine breakdown. Hence, a complex multivariate situation must be analyzed that incorporates also inter-cycle data as valuable state description. Process-oriented developments, so far, focus mostly on monitoring data but ignore the information content of inter-cycle data [29]. Future research for industrial applications, therefore, must address approaches of integrating information from maintenance, experts and other unstructured or implicit sources.

4.2. Optimizing analytics framework

As shown in Fig. 2, the reference analytics architecture on machine level consists of three, partly overlapping modules. The data management comprises all aspects of data acquisition and storage. For the long run, edge computing is utilized for pre-processing, feature extraction and lean organization of raw data streams. The actual analytic pipeline is within the integration and modelling part. Here, the data is further processed and integrated and the predictive model is built. While the model is deployed for the running system it must be constantly retrained to adapt to changing conditions. A data drift detection algorithm, e.g. on the basis of Principal Component Analysis (PCA) or adaptive windowing can be used to determine moment and timeframe [16]. Lastly, the information and control part closes the feedback loop. A visualisation component allows workers to access process information for supervision and decision making. Based on results' interpretation optimization actions and adjustments on machining parameters are initiated. Depending on certainty and production policy these adjustments can be either implemented directly or by human operators. The closed loop optimization is both adaptable for predictive quality as well as maintenance on a machine focused level.

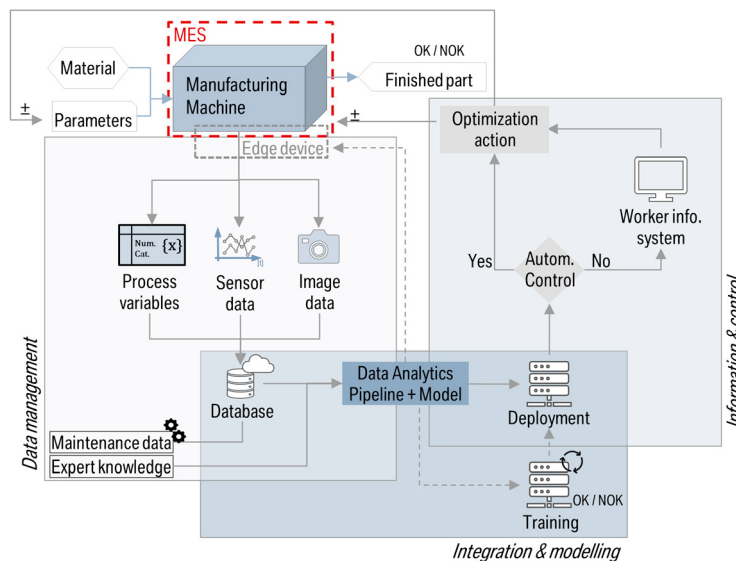


Fig. 2. Optimizing analytics on machine level.

5. Conclusion

With developments towards industry 4.0 the industrial sector has evolved from a purely mechanical to a highly computerized and automated discipline. Following data-driven decision making we summarized key requirements for data analytics and machine learning application in series production. We categorized the industrial production system and demonstrated requirements' relevance based on a real use case from automotive component manufacturing. Along

the value stream, we characterized the production process also by its data stream and pointed out key potentials for ML in such processes. Due to decreasing cost of sensors and devices with simultaneously growing capabilities it is evident that the variety of data sources and hence modalities in industrial processes will even substantially grow. Applications presented in literature are based mostly on process data. In the industrial environment, inter-cycle data contains highly valuable information but is not yet integrated systematically into ML pipelines. Information fusion techniques and multimodal machine learning have been proven effectively in different domains. However, in an industrial context of optimization comprehensive studies for applicability under series production conditions are yet to conduct.

In manufacturing, still, major decisions are confined to human operators, although advancements in computer and data science, process monitoring and automation are increasingly supporting decision making. Use cases from industry and literature reveal not only technological challenges of bringing developments into production. Another big step towards the self-optimizing machine or system is to build trust into artificial intelligence. Research on transparency and explainability of complex machine learning models is expected to facilitate acceptance and confidence. Additionally, research on expert augmented ML will enhance applicability and interpretability of advanced ML models in industry.

References

- [1] Ancona, M., Ceolini, E., Öztireli, C., 2018. Towards better understanding of gradient-based attribution methods for Deep Neural Networks, in: *Int Conf on Learning Representations (ICLR)*. URL: <https://openreview.net/forum?id=Sy21R9JAW>.
- [2] Angelopoulos, A., Michailidis, E.T., Nomikos, N., 2020. Tackling faults in the industry 4.0 era—a survey of machine-learning solutions and key aspects. *Sensors (Switzerland)* 20, 1–34. doi:10.3390/s20010109.
- [3] Ardabili, S., Mosavi, A., 2020. Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods - Engineering for Sustainable Future 101, 384. URL: <http://link.springer.com/10.1007/978-3-030-36841-8>, doi:10.1007/978-3-030-36841-8.
- [4] Baltrusaitis, T., 2019. Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41. URL: <https://ieeexplore.ieee.org/document/8269806/>, doi:10.1109/TPAMI.2018.2798607.
- [5] Bastani, O., 2018. Interpretability via Model Extraction, in: *Workshop on Fairness, Accountability, and Transparency in Mach. Learn.*
- [6] Cai, L., Zhu, Y., 2015. The challenges of data quality and data quality assessment in the big data era. *Data Science Journal* 14, 1–10. doi:10.5334/dsj-2015-002.
- [7] Carballal, A., 2020. Transfer learning features for predicting aesthetics through a novel hybrid machine learning method. *Neural Computing and Applications* 32, 5889–5900. URL: <https://doi.org/10.1007/s00521-019-04065-4>, doi:10.1007/s00521-019-04065-4.
- [8] Castanedo, F., 2013. A Review of Data Fusion Techniques. *The Scientific World Journal* URL: <http://dx.doi.org/10.1155/2013/704504>, doi:10.1155/2013/704504.
- [9] Cho, S., May, G., Tourkogiorgis, I., 2018. A Hybrid Machine Learning Approach for Predictive Maintenance in Smart Factories of the Future, in: *IFIP Int Conf APMS*, pp. 311–317. doi:10.1007/978-3-319-99707-0_{ }39.
- [10] Dalzochio, J., 2020. Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry* 123. URL: <https://doi.org/10.1016/j.compind.2020.103298>, doi:10.1016/j.compind.2020.103298.
- [11] Diez-Olivan, A., Ser, J.D., Galar, D., 2019. Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Information Fusion* 50, 92–111. doi:10.1016/j.inffus.2018.10.005.
- [12] Dogan, A., Birant, D., 2021. Machine learning and data mining in manufacturing. *Expert Systems With Applications* 166, 114060. URL: <https://doi.org/10.1016/j.eswa.2020.114060>, doi:10.1016/j.eswa.2020.114060.
- [13] Du, M., 2019. Techniques for Interpretable Machine Learning. *Communications of the ACM* doi:<https://doi.org/10.1145/3359786>.
- [14] Duc Nguyen, V., Kefalas, M., Yang, K., Al., E., 2019. A Review: Prognostics and Health Management in Automotive and Aerospace. *Int J of Prognostics and Health Management* 10. URL: <https://www.klm.com/corporate/en/publications/2015>.
- [15] Fawaz, H.I., Forestier, G., Weber, J., 2019. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery* 33. doi:10.1007/s10618-019-00619-1.
- [16] Gemaque, R.N., Costa, A.F.J., Giusti, R., 2020. An overview of unsupervised drift detection methods. *WIREs Data Mining and Knowledge Discovery* 10, 1–18. URL: <https://onlinelibrary.wiley.com/doi/10.1002/widm.1381>, doi:10.1002/widm.1381.
- [17] Gennatas, E.D., Friedman, J.H., Ungar, L.H., 2020. Expert-augmented machine learning. *PNAS* 117. URL: www.pnas.org/cgi/doi/10.1073/pnas.1906831117, doi:10.1073/pnas.1906831117/-/DCSupplemental.
- [18] Haghighat, M., 2016. Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition. *IEEE Transactions on Information Forensics and Security* 11. doi:10.1109/TIFS.2016.2569061.
- [19] Heimes, H., Kampker, A., Bühner, U., Krottil, S., 2019. Potenziale und Hürden von Data Analytics in der Serienfertigung. *Industrie 4.0 Management* 2019, 57–60. URL: <http://library.gito.de/oa-im1-19-heimes.html>, doi:10.30844/I40M_{ }19-1_{ }S57-60.
- [20] Janani, T., Ramanan, A., 2017. Feature fusion for efficient object classification using deep and shallow learning. *International Journal of Machine Learning and Computing* 7, 123–127. doi:10.18178/ijmlc.2017.7.5.633.
- [21] Kong, J., Kowalczyk, W., Menzel, S., Bäck, T., 2020. Improving Imbalanced Classification by Anomaly Detection, in: *Parallel Problem Solving from Nature – PPSN XVI, 16th International Conference*. volume 1, pp. 512–523. doi:10.1007/978-3-030-58112-1.
- [22] Kusiak, A., 2018. Smart manufacturing. *International Journal of Production Research* 56, 508–517. URL: <https://www.tandfonline.com/doi/full/10.1080/00207543.2017.1351644>, doi:10.1080/00207543.2017.1351644.

- [23] Lee, G., Kang, B., Nho, K., Sohn, K.A., Kim, D., 2019. MildInt: Deep learning-based multimodal longitudinal data integration framework. *Frontiers in Genetics* 10, 1–7. doi:[10.3389/fgene.2019.00617](https://doi.org/10.3389/fgene.2019.00617).
- [24] Lee, J., 2020. *Industrial AI*. Springer Singapore, Singapore. doi:[10.1007/978-981-15-2144-7](https://doi.org/10.1007/978-981-15-2144-7).
- [25] Lepenioti, K., Bousdekis, A., Apostolou, D., Mentzas, G., 2020. Prescriptive analytics: Literature review and research challenges. URL: <https://doi.org/10.1016/j.ijinfomgt.2019.04.003>, doi:[10.1016/j.ijinfomgt.2019.04.003](https://doi.org/10.1016/j.ijinfomgt.2019.04.003).
- [26] Lichtblau, K., 2015. *Industrie 4.0 Readiness*. VDMA Impuls-Stiftung .
- [27] Murdoch, W.J., Singh, C., Kumbier, K., 2019. Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences of the United States of America* 116, 22071–22080. doi:[10.1073/pnas.1900654116](https://doi.org/10.1073/pnas.1900654116).
- [28] Psarommatis, F., May, G., Dreyfus, P.A., 2020. Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. *International Journal of Production Research* 58, 1–17. doi:[10.1080/00207543.2019.1605228](https://doi.org/10.1080/00207543.2019.1605228).
- [29] Reis, M.S., Gins, G., 2017. Industrial process monitoring in the big data/industry 4.0 era: From detection, to diagnosis, to prognosis. *Processes* 5, 35. URL: <http://www.mdpi.com/2227-9717/5/3/35>, doi:[10.3390/pr5030035](https://doi.org/10.3390/pr5030035).
- [30] Soltanpoor, R., 2016. Prescriptive Analytics for Big Data, in: *Lecture Notes in Computer Science*. doi:[10.1007/978-3-319-46922-5](https://doi.org/10.1007/978-3-319-46922-5).
- [31] Trinks, S., Felden, C., 2019. Smart Factory – Konzeption und Prototyp zum Image Mining. *HMD Praxis der Wirtschaftsinformatik* 56, 1017–1040. URL: <http://link.springer.com/10.1365/s40702-019-00529-2>, doi:[10.1365/s40702-019-00529-2](https://doi.org/10.1365/s40702-019-00529-2).
- [32] Weichert, D., 2019. A review of machine learning for the optimization of production processes. *International Journal of Advanced Manufacturing Technology* 104. doi:[10.1007/s00170-019-03988-5](https://doi.org/10.1007/s00170-019-03988-5).
- [33] Wuest, T., Weimer, D., Irgens, C., Thoben, K.D., 2016. Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research* 4, 23–45. doi:[10.1080/21693277.2016.1192517](https://doi.org/10.1080/21693277.2016.1192517).
- [34] Xu, K., Ba, J.L., Kiros, R., 2015. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention, in: *International Conference on Machine Learning*. URL: <https://arxiv.org/abs/1502.03044>.