Introduction to the Special Section on AI in Manufacturing: Current Trends and Challenges

Jefrey Lijffijt
Ghent University
Belgium
jefrey.lijffijt@ugent.be

Dimitra Gkorou,
Pieter Van Hertum,
Alexander Ypma
ASML
The Netherlands
{dimitra.gkorou, pieter.vanhertum, alexander.ypma}@asml.com

Mykola Pechenizkiy,
Joaquin Vanschoren
TU Eindhoven
The Netherlands
{m.pechenizkiy, j.vanschoren}@tue.nl

ABSTRACT

On 19 September 2022, the first workshop on AI for Manufacturing (AI4M Workshop) took place at ECML-PKDD, the European Conference on Machine Learning and Principles and Practice for Knowledge Discovery in Databases. The workshop brought together researchers and practitioners, from academia and industry, contributing their perspectives. This special section includes five articles in which Artificial Intelligence methods are used to address real problems in the manufacturing industry, ranging from the supply chain, to production, to quality insurance, and predictive maintenance. In this introduction, we present a high-level overview of the current state of the area: observed trends and the main open challenges. This overview is based on these papers, the keynote presentation, the panel discussion, and the discussion that emerged during the workshop.

1. INTRODUCTION

Process optimization, diagnostics, and maintenance of systems, just to name a few, are core to any manufacturing process. Currently these tasks depend largely on the knowledge and expertise of domain engineers. Artificial intelligence (AI), machine learning (ML), and data science (DS) techniques hold great premise to make these tasks (e.g., monitoring, diagnostics, control, planning, optimization, and design) more efficient, more accurate, and more reliable. Beyond increased automation and precision, AI may even bring entirely new possibilities. However, several key challenges are faced when applying AI/ML/DS techniques in this field:

- Often only small amounts of labelled data are available while typically the collected data is very high-dimensional, reflecting a complex problem space. Besides, this data may be biased or imbalanced to specific scenarios or contexts. To this end, techniques that maximize the efficiency of human experts involved in the process, as well as the integration of domain knowledge (in the form of knowledge bases, constraints, or surrogate models) in the learning or reasoning is vital.
- Also, most manufacturing use cases require interpretability and uncertainty quantifications because AI predictions are expected to contribute to decisions on fab processes with high financial impact. Thus, advanced visualisation methods and techniques to enhance interpretability and explainability are necessary.
- Finally, due to the physical complexity of manufacturing systems, problems like concept drift quickly arise, warping the problem space along with time.

These open opportunities and challenges were the motivation to organize a workshop around this topic at the ECML-PKDD conference this year.

This special section covers an introduction and five contributed papers. The introduction is organised as follows. We first summarize the contributions at the workshop: the eight papers accepted for presentation, with brief intro's to the top five papers that we included in this special section. Secondly, we present a summary of the keynote and panel discussion, highlighting current trends and open challenges. We conclude the introduction with a perspective of what the near future in AI for Manufacturing could entail, including aspects of AI relevant to manufacturing that were not touched upon much during the workshop.

2. CONTRIBUTED ARTICLES

We received eleven submissions to the workshop, of which eight were selected to the workshop program after peer-review. The program[1] also featured a invited talk and a panel discussion. Revised contributions of five article were invited for this special section. These five contributed articles span a variety of topics: predictive maintenance [1], fault detection and prediction [2], (acoustic) structural integrity assessment [3], industrial time-series classification [4], and supply chain link prediction [5].

Additionally, a poster was presented on few-shot learning to identify faults and concept drift, and there were two demo's. One demo introduced a system called 'Probe' to find and explore similar parts, based on 3D models, which is useful during design, and the other introduced the 'Data Curation Canvas', an approach to data curation that supports data-centric AI, leveraging modern AI techniques (self-supervision, explainability, uncertainty estimation).

[1]The full programme and preprints of the papers not included in this special section can be found at `https://sites.google.com/view/ai4manufacturing/home/`

Investigating thresholding techniques in a real predictive maintenance scenario Giannoulidis et al. [1] present a study about unsupervised anomaly detection for predictive maintenance, on real streaming data coming from a cold-forming press. Anomaly detection methods frequently output a score, e.g., a probability, and then employ a threshold to decide whether the current data window warrants throwing an alarm message. In this case, the streaming data is not stable and changes occur over time. The study encompasses the evaluation of several thresholding techniques, including dynamic thresholding and a new method called *self-tuning*. The conclusion is that advanced thresholding methods indeed lead to higher F1 scores and the study informs on which method may be most appropriate given the characteristics of the data and the problem.

Feature Selection for Fault Detection and Prediction based on Log Analysis. Li and van Leeuwen [2] also consider fault/anomaly detection and prediction, but from event logs. Event logs can contain a wide variety of information and the study is about selecting a small set of log events that are most informative about the prediction task. Their approach consists of first *vectorizing* the log events, then feature selection using rank correlation and causality testing, and finally removal of redundant features. They evaluate this approach on 25 datasets and their results indicate this approach can lead to more accurate results, using a simpler model. Although not studied in the paper, the resulting models are probably substantially more interpretable.

Acoustic Structural Integrity Assessment of Ceramics using Supervised Machine Learning and Uncertainty-Based Rejection. Nunes et al. [3] show how supervised methods can be used in combination with uncertainty modeling to partially automate quality monitoring. Their application is integrity assessment of ceramic plates (cracked vs. uncracked), using an actuator and microphones, i.e., an audio signal in a controlled but not noise-free setting. The application context of a model rarely perfectly matches the context where the training data was gathered, and to this end uncertainty modelling may inform about the confidence of a classifier whether it can make a robust decision. To optimize the performance in this specific setting, several ML models and calibration are tested. The conclusions is that accurate performance can be obtained if the classifier is allowed to reject samples using uncertainty modeling.

Experiences with Contrastive Predictive Coding in Industrial Time-Series Classification. Gamage et al. [4] introduce a method for time-series classification using self-supervised learning, to tackle the problem only few samples may be labeled, while rich sensor measurement data is available. Specifically, they consider the use of contrastive predictive coding (CPC) to learn a suitable representation from unlabeled time-series, and then a multi-layer perceptron in supervised mode on that representation. They find that this can strongly boost performance if only few samples are labeled. On top of this, they consider to employ the learned representation for active learning to increase the label efficiency further, but they find only limited gain.

Supply Chain Link Prediction on Uncertain Knowledge Graph. Brockmann et al. [5] consider a very different type of problem that is nonetheless highly relevant to manufacturing: understanding the supply chain. Companies often do not have a full view of their supply chain or the relations between actors in their supply chain. They propose to construct an uncertain knowledge graph from web data using NLP, and then study how to do link prediction on this uncertain knowledge graph, using a graph neural network. They find strongly improved performance, and besides the scores from the uncertain link prediction may inform about the certainty of the predictions and these may enhance decision making based on the predicted links.

3. KEYNOTE TALK

The keynote talk dealt with the AI opportunities and challenges in semiconductor manufacturing. It was presented by Dr Alexander Ypma, Data and Analytics Manager at ASML, the leading supplier of lithography machines in the semiconductor industry.

Integrated Circuit (IC) manufacturing is a complex process where various machines and processing tools, such as deposition, coating, exposure or etching tools, are used. ICs are being fabricated on a thin silicon plate, called a wafer, which is processed in several layers. Latest generations of ICs can consist of up to 100s of layers. For instance, an advanced logic process could have from 600 to 1,000 steps and could take up to 3 months to be produced. Sensors monitor each step of this process, generating terabytes of data every hour. Due to the data-heavy nature of this process, it is a good candidate to be supported by ML and AI techniques. AI may be used to enhance the automated monitoring, diagnostics, and control, as well as to provide domain engineers with the necessary tools for more effective process optimization. The most promising use cases in IC manufacturing were grouped in three levels: fab level, system level and hardware component level.

At the fab level: a wafer passes multiple process steps that could cause errors and variations. Multiple tools need to be monitored and controlled at the same time. Variations and instabilities may lead to wafer deformations. The physical nature of most deformations can be described by a distinct pattern, the so-called fingerprint. Lithography has the most advanced control knobs and so, it can control variation patterns from other tools such as etcher or deposition tools. The following two use cases may benefit from AI:

- **Adaptive Wafer Level Control**: With the help of ML, lithography machine may monitor deformation patterns on a wafer, and link them to fab tools causing them. Deformation patterns may propagate when stacking layers. With the help of context data, we may identify the layers that are responsible for the deformation. Fab tools have different fingerprints, after being able to link fingerprints to tools, wafer level corrections can be applied accordingly. Wafer level control results in tighter process optimization. Pattern recognition, clustering, and control methods are important for this use case.

- **Virtual Metrology**: The quality of an exposure is measured via dedicated metrology tools. Metrology requires extra time and thus, is a bottleneck in production. Predicting the quality of an exposure, so-called virtual metrology, may increase throughput and reduce the production cost. However due to the same restrictions on time and the metrology cost, only small high-dimensional datasets are available for training. Feature engineering and active learning are crucial to boost the performance of ML models. By building advanced visualisation, domain experts can interact and engineer features that have a physical meaning.

At the system level: a lithography machine consists of hundreds of hardware components. The high physical complexity and nanometer resolution lead to many interactions between modules, making system monitoring and diagnostics a big challenge. Especially, when we are diagnosing performance drops instead of hard failures, it becomes increasingly difficult to identify the root-cause of a problem. In this context, ML may enhance diagnostics via the following use cases:

- **Diagnostics with causal reasoning**: To find the root cause of a failure or performance problem of a lithography machine, it is crucial to combine domain expertise with sensor data to allow for causal reasoning in this complex domain. To this end, diagnostics tooling use interfaces that can capture the relevant domain knowledge of domain experts and frameworks that use the knowledge to automatically create custom end-to-end diagnostics pipelines that can interpret data, identify health metrics and produce a root-cause diagnosis for a specific set of problems. The main components of this tool are designing the causal graphs representing the system components and their interactions in the right abstraction level, doing probabilistic inference with appropriate graphical models, and presenting the results to domain experts.

- **Monitoring Equipment Availability**: Proper monitoring of machines trough their event logs is necessary to ensure functionality and availability. Process mining is used to recognize sequences and service actions. The pipeline processes the data, finds relevant information in the machine logs, and transforms it into actionable insights for the domain experts. These event-logs often consist of terabytes of data, making scalability an important factor. MLOPs technologies are essential to enable operating at this scale.

At the hardware component level: it is important to monitor the functionality of each hardware component separately for better diagnostics and prognostics. Next to this, ML can also support tuning and calibration of the individual modules. On component level, there are two use cases:

- **Predictive Maintenance:** The performance of certain hardware components in a machine can degrade over time due to ageing. If untreated, this can lead to performance problems, or machine failure, resulting in yield loss. Sensor measurements monitor the health state of hardware components. Classification models use the high dimensional sensor data to estimate (future) health states of a component and time series forecasting and regression help estimating the Remaining Useful Lifetime. The physical complexity leads to concept drift and noisy labels, introducing a high need for smart interactive labeling and advanced MLOps [6] to enable strong monitoring, quick retraining, and the orchestration of multiple interacting models.

- **Surrogate modeling:** To calibrate a lithography machine, physical simulations are often needed. These simulations need to be fast, in order to calibrate models in their relevant operating conditions (which often requires hyperparameter optimization that involves many expensive function evaluations). With surrogate modeling, a ML model is trained to simulate the physical model at a much higher computational efficiency. By offloading the expensive simulations offline, more computational resources can be exploited to support the expensive simulation and training process.

Through these use case, the following ML challenges are identified:

- **Data quality and availability**: Typically, labeled data is rare and often only partially labeled due to a high label acquisition cost. Data is often very imbalanced because out of spec behaviour is very rare, leading to an unequal risk of misclassification of the fault class, together with an imbalanced cost of misclassification. These challenges offer opportunities for semi-supervised active learning with a human in the loop. The challenges on imbalanced data invite techniques like one-class classification, autoencoders, and manifold learning.

- **System complexity**: System of systems level interactions, together with the nanometer resolution lead to high complexity. Together with the heterogeneous irregularly sampled and partially observed data in this high-dimensional space, this is a small sample size problem. Data distributions and operating conditions of these machines drift and so, adaptive methods, monitoring, data-centric learning, and MLOps practices are essential.

- **Modeling challenges**: Solutions need to be plausible and explainable for use in the larger semiconductor ecosystem. The need to interpret models and to solve inverse problems drives AI to utilise physics domain knowledge. To achieve this, a combination of physics-driven and data-driven ML is needed. Areas of interest include knowledge representation and elicitation methods, physics-informed ML, surrogate modelling, causal inference, and hyper-parameter optimization.

The invited talk concluded that in semiconductor manufacturing, great opportunities for applying ML and AI exist, needing state-of-the-art techniques, solutions and tooling.

4. PANEL DISCUSSION

The panel included participants from manufacturing industry and academia (Micron, Intel, Imec, KU Leuven). Its goal was to get insights from industrial practitioners and academic partners on the adoption, common challenges, and existing limitations of AI in manufacturing.

The panelists recognized the potentiality of AI in manufacturing but also admitted that AI currently supports only a narrow spectrum of applications in prognostics and diagnostics. Currently, most manufactured products would require distinctly trained ML models. However, they expect greater adoption as increasing volumes of data become available, data pipelines improve, and the ability to transform manufacturing problems into ML problems matures. The challenges of AI adoption include the need for interpretable model predictions, small data sets available for training ML algorithms, and the ability to scale fast ML models. In the first two challenges, collaboration between data scientists and domain experts is a key. For the latter challenge especially proper ML engineering is crucial.

The first highlighted challenge was interpretability. Important in manufacturing is to minimise defects, downtime or anything that adversely affects production goals. In such cases it is not enough to predict but also explain predictions so appropriate action can be taken on it. As a result, next to predictive analyticts, the field shows an increasing interest in prescriptive analytics, which will prescribe the right action to take. To face this challenge data science teams collaborate closely with the domain experts. The domain experts interact with the models via advanced visualisations, embeddings, dimensionality reduction techniques and explanations of model decisions.

Second, to work with the small datasets often encountered, the role of domain experts is crucial. Invariably trying to predict adverse outcomes results in small datasets and class imbalance, since in high volume manufacturing, there are only a few of them. Domain experts help clean noisy labels, dive into uncertain cases, support feature engineering, and define the constraints of a problem. In many use cases, the labels are not clear and the data scientists start with unsupervised approaches. With the help of domain experts and extra analysis, labels can be obtained or generated. By closely interacting with domain experts, data scientists are able to better define their use case, construct their data sets and properly constrain their problem. On the other hand, through the interaction with data scientists, domain experts may get new insights in their problems and learn more on the statistical properties of their domain.

The last highlighted challenges was about scalability. ML solutions need to support process optimization during manufacturing in order for AI to be widely adopted. In high volume production, terabytes of data are produced every hour. Scaling ML solutions with proper deployment including MLOPs, solid data pipelines, high performance data / analytics infrastructure, and proper orchestration is crucial for actually using it in a factory. Towards this end, a collaboration between data scientists, data engineers, IT experts, domain experts, and ML engineers is important.

Concluding, the panelists mentioned that learning how to properly apply AI is a journey for each company. They highlighted that the collaboration between domain experts and data scientists is the key factor for the success of an AI project. It is important to start with small projects with tangible goals because AI is new technology for manufacturing and thus, risky. After some small successes, bigger projects with increasing complexity become possible. This approach enables the teams to learn incrementally and to improve before tackling big projects.

5. CONCLUSIONS AND OUTLOOK

There exist a plethora of potential uses of AI in the manufacturing industry, yet many challenges studied actively in the AI, machine learning, and data science communities are highly relevant to make use cases implementing AI a success. Methods for control, optimization, and planning have been used in the sector for a long time already [7]. However, there is a need to use state of the art AI methods, as resources and time are more scarce than ever. Fahle et al. [8] provide a review of machine learning techniques that have been applied in a range of use cases, while Usuga Cadavid et al. [9] review machine learning methods specifically for production planning and control. As an outlook towards the future topics that could be covered in the domain, we provide a high-level grouping of the relevant challenges:

Complex data and context: although there is rich data collection in manufacturing processes, the problem space is often very high-dimensional and the context often changes over time (changing environment, concept drift [10], dynamic production processes, per-batch customization), so the data, of which *time* is almost always a crucial aspect, remains sparse. AI methods for time-series [11; 12], event sequences [13], and data streams [14] are particularly relevant in this domain. Data sparsity may be alleviated partially through hybrid (e.g., physics-informed) models [15], the use of surrogate models, simulation data, or digital twins. The changing environment may be addressed using anomaly detection, change detection, transfer and online learning. Robustness and generalization capability of models are crucial for operation in dynamic contexts.

Scarcity of domain experts and feedback: obtaining labels for data is often time consuming and costly. Hence, only a small portion of the available data is labeled, and besides when constructing AI/ML pipelines, there is a need to integrate expert knowledge. There is a strong need to make maximal use of domain experts' time. Firstly, this requires the use of explainable models (simplicity and causal learning are important for this) or explainability methods [16], including the visualisation of data and predictions, to give experts insight into the working of the model, such that they can provide useful feedback. Secondly, systems may guide the experts to provide maximally useful feedback, in the form of active learning and more generally human-in-the-loop modes of operation [17], and helping experts to understand problems or anomalies (e.g., through root-cause analysis). Worth to mention here is also the trend on data-centric AI [18; 19], which emphasises the role of data quality and curation in building AI systems. Thirdly, small amounts of labeled data may be combined with large amounts of unlabeled data through few-shot and self-supervised learning [20]. Finally, related to domain experts but slightly different and not discussed during the workshop are new possibilities for AI systems and robots to work together with humans in a production environment (so-called cobots; these come with their own challenges, such as safety, human-machine interaction, and dialogue systems).

Autonomy and collaboration: in future editions of the workshop, it would be interesting to also touch upon the topic of autonomy of production systems that integrate or use AI. AI methods that operate in the edge, i.e., within a machine or even a component, need to either be very robust, or for example integrate uncertainty modeling (e.g., as considered for a specific application by Nunes et al. [3]), such that the system itself knows when it needs attention from an expert. A set of machines working together can be modeled as a multi-agent system, to facilitate their cooperation [21]. Finally, beyond the company level, federated learning may facilitate the exchange of information or models, improving the operations of all parties without having to share confidential information [22].

New applications: AI has already brought forward some

use cases in the manufacturing industry that would not be possible without. Notably, predictive maintenance and real-time automated quality monitoring. Brockmann et al. [5] also provide pointers toward extended supply chain analytics. AI may also lead to the next levels of accuracy/precision in manufacturing, either at smaller scales, or with fewer variation or defects (e.g., zero-defect manufacturing). As the final application, AI can facilitate higher efficiency or effectiveness in design [23], as a form of augmented intelligence, also highlighted in one of the presented demo's.
Security: As the final point, the trend in AI is to use data from outside sources or pre-trained models for various tasks. Although these are most prominent in AI on text and images, this may soon become a trend for time-series data. High-stakes applications like in manufacturing seem to warrant extra attention to issues of security and the study of adversarial robustness. As models are complex, these are prone to manipulation for a multitude of reasons [24].

We hope you will enjoy reading the papers on AI in Manufacturing in this special section and find them an inspiration for further research in this area.

Acknowledgements

The workshop took place with the help from ASML, Intel, Micron, Imec, KU Leuven, TU Eindhoven, and Ghent University. We would like to thank the panelists (William Susanto, Micron; Hai Liang 'Duncan' Lee, Intel; Bappaditya Dey, Imec; and Mathias Verbeke, KU Leuven), all authors who submitted and presented their work, the PC members, and all the participants for their contributions.

6. REFERENCES

[1] A. Giannoulidis, N. Nikolaidis, A. Naskos, A. Gounaris, and D. Caljouw. Investigating thresholding techniques in a real predictive maintenance scenario. *SIGKDD Expl.*, 24(2), 2022.

[2] Z. Li and M. van Leeuwen. Feature selection for fault detection and prediction based on log analysis. *SIGKDD Expl.*, 24(2), 2022.

[3] M.L. Nunes, M. Barandas, H. Gamboa, and F. Soares. Acoustic structural integrity assessment of ceramics using supervised machine learning and uncertainty-based rejection. *SIGKDD Expl.*, 24(2), 2022.

[4] S. Gamage, B. Klöpper, and J. Samarabandu. Experiences with contrastive predictive coding in industrial time-series classification. *SIGKDD Expl.*, 24(2), 2022.

[5] N. Brockmann, E. Elson Kosasih, S. Baker, I. Blair, and A. Brintrup. Supply chain link prediction on an uncertain knowledge graph. *SIGKDD Expl.*, 24(2), 2022.

[6] D. Sculley et al. Hidden technical debt in machine learning systems. In *NIPS*, pages 2503–2511, 2015.

[7] W. Zijm. Towards intelligent manufacturing planning and control systems. *OR Spektrum*, 22:313–345, 2000.

[8] S. Fahle, C. Prinza, and B. Kuhlenköttera. Systematic review on machine learning (ML) methods for manufacturing processes – identifying artificial intelligence (AI) methods for field application. *CIRP*, 93:413–418, 2020.

[9] J.P. Usuga Cadavid, S. Lamouri, B. Grabot, R. Pellerin, and A. Fortin. Machine learning applied in production planning and control: a state-of-the-art in the era of industry 4.0. *J Intell Manuf*, 31:1531–1558, 2020.

[10] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang. Learning under concept drift: A review. *IEEE TKDE*, 31(12):2346–2363, 2019.

[11] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller. Deep learning for time series classification: a review. *Data Min Knowl Disc*, 33:917–963, 2019.

[12] Q. Ma, J. Zheng, S. Li, and G.W. Cottrell. Learning representations for time series clustering. In *NeurIPS*, volume 32. Curran Associates, Inc., 2019.

[13] K. Diba, K. Batoulis, M. Weidlich, and M. Weske. Extraction, correlation, and abstraction of event data for process mining. *WIREs DMKD*, page 10:e1346, 2020.

[14] H. Murilo Gomes, J. Read, A. Bifet, J.P. Barddal, and J. Gama. Machine learning for streaming data: state of the art, challenges, and opportunities. *SIGKDD Expl.*, 21:6–22, 2019.

[15] L. von Rueden et al. Informed machine learning - a taxonomy and survey of integrating prior knowledge into learning systems. *IEEE TKDE*, 2021.

[16] R. Marcinkevics and J.E. Vogt. Interpretability and explainability: A machine learning zoo mini-tour. arXiv:2012.01805, 2020.

[17] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and A. Fernández-Leal. Human-in-the-loop machine learning: a state of the art. *Artif Intell Rev*, 2022.

[18] E. Strickland. Andrew Ng, AI minimalist: The machine-learning pioneer says small is the new big. *IEEE Spectrum*, 59(4):22–50, 2022.

[19] H. Patel, S. Guttula, R.S. Mittal, N. Manwani, L. Berti-Equille, and A. Manatkar. Advances in exploratory data analysis, visualisation and quality for data centric ai systems. In *KDD'22*, page 4814–4815, 2022.

[20] Y. Wang, Q. Yao, J.T. Kwok, and L.M. Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.*, 53(3), 2020.

[21] M.-A. Dittrich and S. Fohlmeister. Cooperative multi-agent system for production control using reinforcement learning. *CIRP Annals*, 69(1):389–392, 2020.

[22] T. Li, A.K. Sahu, A. Talwalkar, and V. Smith. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Mag.*, 37(3):50–60, 2020.

[23] W.J.C. Verhagen, P. Bermell-Garcia, R.E.C. Van Dijk, and R. Curran. A critical review of knowledge-based engineering: An identification of research challenges. *Advanced Engineering Informatics*, 26(1):5–15, 2012.

[24] R. Bommasani et al. On the opportunities and risks of foundation models. ArXiv:2108.07258, 2021.