International Journal of Production Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tprs20

Federated explainable artificial intelligence (fXAI): a digital manufacturing perspective

Andrew Kusiak

To cite this article: Andrew Kusiak (2023): Federated explainable artificial intelligence (fXAI): a digital manufacturing perspective, International Journal of Production Research, DOI: 10.1080/00207543.2023.2238083

To link to this article: https://doi.org/10.1080/00207543.2023.2238083

Published online: 23 Jul 2023.

INTERNATIONAL JOURNAL OF PRODUCTION RESEARCH
https://doi.org/10.1080/00207543.2023.2238083

Taylor & Francis
Taylor & Francis Group

Federated explainable artificial intelligence (fXAI): a digital manufacturing perspective

Andrew Kusiak

Department of Industrial and Systems Engineering, The University of Iowa, Iowa City, IA, USA

ABSTRACT

The industry has embraced digitalisation leading to a greater reliance on models derived from data. Understanding and getting insights into the models generated by machine learning algorithms is a challenge due to their non-explicit nature. Explainable artificial intelligence (XAI) is to enhance understanding of the digital models and confidence in the results they produce. The paper makes two contributions. First, the XRule algorithm proposed in the paper generates explicit rules meeting user's preferences. A user may control the nature of the rules generated by the XRule algorithm, e.g. degree of redundancy among the rules. Second, in analogy to federated learning, the concept of federated explainable artificial intelligence (fXAI) is proposed. Besides providing insights into the models built from data and explaining the predicted decisions, the fXAI provides additional value. The user-centric knowledge generated in support of fXAI may lead to discovery of previously unknown parameters and subsequently models that may benefit the non-explicit and explicit perspectives. The insights from fXAI could translate into new ways of modelling the phenomena of interest. A numerical example and three industrial applications illustrate the concepts presented in the paper.

ARTICLE HISTORY
Received 2 January 2023
Accepted 11 July 2023

KEYWORDS
Explainable artificial intelligence (XAI); Federated XAI; Digital manufacturing; Data science; Decision-making

1. Introduction

The interest in applications of artificial intelligence (AI) is growing across industrial and service domains. Machine learning, in particular neural networks, have taken the front stage of the emerging applications. For example, software tools and algorithms such as deep, adversarial, and broad neural networks have been applied to build complex models from large volumes of data of different origins. However, the progress made in the development of these software tools and algorithms for building models is not well balanced with the research on the tools enhancing the transparency of the decisions produced by these models. The research in explainable AI (XAI) offers a user-centered view of the data-derived models and the underlying processes. The research reported in this paper makes two contributions to the XAI research. First, the XRule algorithm proposed in the paper generates explicit rules to satisfy preferences set by the users interested in the insights of predictive models and providing clarity of the predicted outcomes. Second, a concept of federated explainable artificial intelligence (fXAI) is proposed. It parallels the idea of federated learning that applies to the model-building phase. While federated learning focuses on preserving data privacy, fXAI aims at providing insights into the models built by machine learning algorithms and explaining the predicted results. The deliberate requirement in federated learning to limit the data exchange between the model developers, which contrasts the knowledge sharing notion among the model users in fXAI. In addition, the model users in fXAI actively communicate with the experts in other domains, including the model developers.

The paper is structured in seven sections. Section 1 introduces the research topic. The literature on explainable artificial intelligence is surveyed in Section 2. The XAI progress, in general, is reviewed in Section 2.1, while Section 2.2 is focused on the manufacturing applications. The implementation of the XAI concept supported with the XRule algorithm developed in the research reported in this paper is included in Section 3. The steps of the XRule algorithm are presented in Section 3.1. An example illustrating the XRule algorithm is provided in Section 3.2. Section 4 introduces federated learning and reviews the most recent developments in this important area of research. The concept of federated explainable artificial intelligence (fXAI) is proposed in Section 5. Section 6

CONTACT Andrew Kusiak andrew-kusiak@uiowa.edu Department of Industrial and Systems Engineering, The University of Iowa, Iowa City, IA 52242–1527, USA

discusses the concept of federated XAI in digital manufacturing. Section 7 concludes the paper.

2. Explainable artificial intelligence (XAI) in the literature

The majority of machine learning algorithms produce models that are complex and non-explicit. While data analysts may find limited ways to interpret the data-derived models and their inputs and outputs, most users do not share the same experience. The users expect to be presented with insights into the models and be offered justification of the outcomes generated by these models.

Explainable AI (XAI) is to enhance model understanding and confidence in the decisions produced by the models extracted with machine learning algorithms.

Doran, Schulz, and Besold (2017) classified the data-derived models into three categories:

(a) Comprehensible models – Both, the insides into the model and the output are explained, e.g. decision tree models;
(b) Interpretable models – The relationship between the inputs and the output is expressed with a formal model, e.g. a linear regression model;
(c) Opaque models – A user cannot comprehend the model, e.g. a neural network model.

An exhaustive review of the explainable artificial intelligence (XAI) literature was published by Barredo Arrieta et al. (2020). The survey provided a taxonomy of XAI for deep learning algorithms. The following four groups of goals of XAI were identified:

(a) Trustworthiness and causality;
(b) Transferability and informativeness;
(c) Confidence, fairness, and accessibility;
(d) Interactivity and privacy awareness.

In addition, the following terms used to characterise machine learning models were defined:

(a) Understandability;
(b) Comprehensibility;
(c) Interpretability;
(d) Explainability; and
(e) Transparency.

The transparency of six different models, i.e. linear regression; decision tree; rule based; k-nearest neighbour; Bayesian, and generalised additive models, was illustrated with numerical examples and graphics. Also, fairness, ethics, privacy, accountability, transparency, and security and safety of responsible AI were discussed.

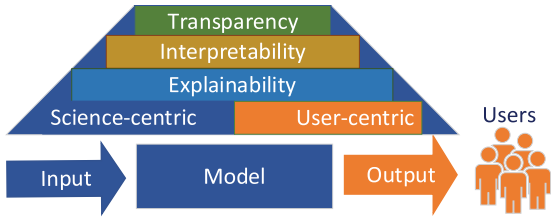

Figure 1. The scope of transparency, interpretability, and explainability.

The recent progress in explainable artificial intelligence is discussed in Section 2.1.

2.1. Recent developments in explainable AI

The paper by Li et al. (2022) reviewed and categorised the XAI methods as data-driven and knowledge-aware. The explanation characteristic in the former methods was expanded into instance based, local, and global categories using the task-related data. The knowledge-aware methods were categorised as knowledge-based and general knowledge methods. The XAI solutions deployed in industry were surveyed. Mohamed, Sirlantzis, and Howells (2022) surveyed visualisation techniques in explaining the architecture, logic, biases, and behaviour of intelligent systems. The XAI benefits such as increased transparency, safety, and confidence in the results were emphasised. In the survey paper, Roscher et al. (2020) focused on three characteristics of XAI, transparency, interpretability, and integration of domain knowledge. The latter involves the explainabilty of a model and its inputs and outputs. Sutthithatip et al. (2021) elaborated on the transparency characteristic of the machine learning algorithm (training transparency) and that of a model and its components. The authors considered interpretability and explainability of a model as well as its input and output. In addition, a science-centric and a human-centric perspective were incorporated.

The XAI perspectives included in Roscher et al. (2020) and Sutthithatip et al. (2021) are illustrated in Figure 1.

The capabilities, limitations, characteristics, and risks of XAI tools were discussed in Fiok et al. (2022). Expectations from the XAI tools from two perspectives, training and education and data science, were analysed. Nakhle and Harfouche (2021) published a tutorial on fundamentals of artificial intelligence and its applications in image analysis. Though the focus of the paper was on open-source platforms and libraries, a guide for implementation of XAI algorithms was provided.

The literature on explainable artificial intelligence (XAI) has contributed new approaches. The search for new solutions is expanding across different domains. Examples of the recently developed representative XAI solutions are discussed next.

A visual analytics system, explAIner, for interactive and explainable machine learning was presented in Spinner et al. (2020). The system enhances understanding of the data-derived models, explains the model limitations, and optimises the models. The utility of explAIner was validated in a user study. The paper compared 21 different XAI methods. A modular approach, called eXplainable Mapping Analytical Process (XMAP), supporting interpretability across all phases of the model development was introduced in Nguyen and Tran (2022). The XMAP includes algorithms capturing the data structure and its context. Buczak et al. (2022) developed a model for predicting disruptive events around the world. A Shapley additive explanation approach was used to explain the predictions. The proposed approach can be applied to deep reinforcement learning in other domains, e.g. autonomous cars and unmanned aircrafts. Bin Iqbal, Muqeet, and Bae (2022) visualised the high-impact image regions of the outcome predicted by a deep neural network. The adjacent layers were screened for contributions among the connected structures and assigned scores were used to identify discernible neurons. A visualisation map was constructed. He, Aouf, and Song (2021) applied a feature attribution approach to a deep reinforcement learning model for path planning of unmanned aerial vehicles. The behaviours of interest were explained with the text and visualisation. The use of conceptual knowledge in training explainable models was discussed by Holzinger et al. (2021). The graph neural networks were applied for interactive explainability with the goal of the development of human-AI interfaces. Jia et al. (2022) introduced a visual explainable active learning approach for zero-shot classification involving disjoint training and test classes. Four actions, i.e. ask, explain, recommend, and respond were used by an analyst to understand misclassifications. The proposed approach improved the efficiency of building zero-shot classification models.

A review of the recently published papers on the design of explainable systems in health-care applications was authored by Markus, Kors, and Rijnbeek (2021). A framework for class selection of explainable approach was included. Dey et al. (2022) focused on expainabilty in the healthcare domain. Besides reviewing the literature and classification of XAI methods, the authors made a claim that the XAI methods are not sufficient for the implementation of AI solutions in healthcare. Rather, multi-layered cybersecurity solutions, including the concepts of AI trustworthiness and AI fairness would need to be implemented. The paper touched on an important aspect of knowledge transfer for explainability.

The research on explainable artificial intelligence in manufacturing is gaining momentum. The papers illustrating representative applications in manufacturing are discussed in Section 2.2.

2.2. Explainable AI in manufacturing

Inspired by the developments in Industry 4.0, Ahmed, Jeon, and Piccialli (2022) surveyed applications of artificial intelligence and XAI methods in industry. Future research directions of AI applications were outlined. The recently published papers on XAI in manufacturing grouped in six categories, ranging from manufacturing systems to cybersecurity, are discussed next.

2.2.1. Manufacturing systems

Rožanec et al. (2022) presented a human-centric architecture involving explainable AI in a broader context of manufacturing evolution towards Industry 5.0. The proposed architecture is synergistic with the Big Data Value Association Reference Architecture Model. Taj and Jhanjhi (2022) assessed challenges and opportunities of XAI solutions in Industry 5.0 applications. The key attributes of Industry 5.0 such as productivity, human-machine collaboration, data transmission, interoperability, security, and privacy were emphasised.

2.2.2. Manufacturing processes

Kuhnle et al. (2022) investigated explainable reinforcement learning in production control. Control strategies in the form understandable to a user were illustrated in a semiconductor case study. The paper by Goldman et al. (2023) aimed at the development of trustworthy AI solutions for applications in manufacturing. Class activation maps were applied for prediction of quality and variability reduction in manufacturing. Robustness of the classifiers was evaluated with the contrastive gradient-based saliency maps. Kotriwala et al. (2021) discussed challenges facing applications of artificial intelligence in the process industry. A few successful applications of XAI were reviewed.

2.2.3. Condition monitoring

A deep neural network was considered by Keleko et al. (2023) for condition monitoring of hydraulic systems using data from multiple sensors. A deep Shapley additive approach was used to explain the importance of data provided by the sensors and the results generated by the neural network. Jakubowski, Stanisz, and Nalepa

(2022) modelled the degradation the cold-rolling process with a physics-informed autoencoder. An XAI solution was deployed to explain the prediction results. A data-derived model for classification of tool wear using acoustic emission sensors was discussed in Schmetz et al. (2021). The interpretability of the predicted results was enhanced with the feature importance analysis. Hrnjica and Softic (2020) provided an example illustrating the concept of explainable AI versed in the gradient-boosting decision tree in a predictive maintenance application.

2.2.4. Fault prediction

A data-driven approach for fault diagnosis of 3-D printers was proposed by Chowdhury, Sinha, and Das (2023). The Shapley additive explanation approach was applied for the interpretation of the prediction results. Lee, Jeon, and Lee (2022) discussed explainable AI to the deep-learning image model classifying defects of TFT–LCD panels. The interpretability of the results was enhanced with visualisation based on a layer-wise relevance propagation and a decision tree. Cheng et al. (2022) discussed a rule-based approach for explanation of the local defect patterns classified by the machine-learning algorithm. The proposed XAI approach provided association between the root causes and the defects. Defect detection with a deep neural network was discussed by Lorentz et al. (2021). Human-friendly saliency maps highlighting the image areas impacting the predictions were offered. A metric for quantification of the saliency maps was proposed.

2.2.5. Decision support

Cochran et al. (2022) discussed the use of information models in support of XAI. An information model allows to capture changing conditions of the production environment. An approach based on the Kano model was applied by Joung and Kim (2022) to offer insights into neural networks. The utility and efficiency of the proposed approach were validated in a case study involving three Fitbit models.

2.2.6. Cybersecurity

Makridis et al. (2022) applied three XAI methods to defend against gradient evasion attacks in the classification of manufacturing images: (i) local interpretable model-agnostic explanations, (ii) saliency maps, and (iii) gradient-weighted class activation mapping. Performance of the three methods was evaluated, with the first method outperforming the remaining two. The need for XAI solutions in anomaly detection in industrial control systems was discussed by Ha et al. (2022). The XAI proposed approach involving the long short-term memory-based autoencoder model was tested with a public data set.

The discussion above supported by the recently published papers has demonstrated that XAI is a complex topic (e.g. Ahmed, Jeon, and Piccialli 2022). In fact, a claim could be made that the rate of deployment of AI solutions in industrial and other applications is conditioned on the progress in XAI (Kusiak 2023). Though the volume of XAI literature is growing, no universal solution has been developed, rather the techniques developed to date are primary science- and user-centric (see Figure 1) data or algorithm centric.

The research published to date covers many facets of XAI. The focus of this paper is on the user-centric domain of XAI (see Figure 1) providing insights into the models and the predicted outcomes. This view is of great interest to digital manufacturing where predictive models may cover diverse phenomena encompassing multiple domains.

From a user perspective, the common concerns of XAI mentioned in the literature (e.g. Barredo Arrieta et al. 2020; Doran, Schulz, and Besold 2017; Sutthithatip et al. 2021) are:

✓ Complexity;
✓ Privacy; and
✓ Trust.

The federated explainable approach (fXAI) approach presented later in this paper (see Section 5) supports the first two concerns, complexity and privacy, and it enhances trust in the results produced by the model. The decision trees generated from the subsets of model parameters, offer user-focused insights into the model. The complexity of each user view can be controlled by the parameters selected for model building. The information and knowledge shared among different users can be managed for privacy protection. As trust involves many aspects of XAI, the different explicit views of the model contribute to the trust enhancement.

3. Rule-based explainabilty

The explicit learning algorithms such as decision tree, decision rules, association rules, and clustering algorithms deserve research attention as the results they produce are comprehensible to users.

Decision trees, in particular, are a popular choice as they explicitly demonstrate how different parameters contribute towards the predicted outcome. Cao, Sarlin, and Jung (2020) extracted decision trees in the sparse

Table 1. The data set used to demonstrate the XRule algorithm.

No.	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	Decision
1	0	R	0	*P*	T	B	X	y	0	s	0	12.1	0.4	B
2	0	R	0	*P*	t1	black	X	y	0	s	0	0	0	C
3	0	R	7.2	pink	T	B	X	y	0	s	0	0	0	B
4	0	R	0	*P*	T	B	X	y	0	b_3	4.6	0	0	C
5	0	R	7.8	pink	T	B	X	y	0	s	0	0	0	B
6	1.8	red	0	*P*	T	B	X	y	0	s	0	0	0	A
7	0	R	0	*P*	t1	black	X	y	0	s	0	0	0	C
8	0	R	0	*P*	T	B	X	y	0	s	0	13.3	0.2	B
9	0	R	0	*P*	T	B	b_2	a_4	5.1	s	0	0	0	A
10	0	R	8.1	pink	T	B	X	y	0	s	0	0	0	B
11	0	R	0	*P*	T	B	b_2	a_4	6	s	0	0	0	A
12	0	R	0	*P*	t1	black	X	y	0	s	0	0	0	C
13	0	R	0	*P*	T	B	b_2	a_4	5.5	s	0	0	0	A
14	2	red	0	*P*	T	B	X	y	0	s	0	0	0	A
15	0	R	0	*P*	T	B	X	y	0	b_3	4.2	0	0	C

Note: A data set with 15 columns and 16 rows. The columns and labelled, 'No.', 'P_1' through 'P_{13}', and 'Decision'. Fifteen rows of data are included in the data set.

k-conjunctive normal form. Offline and online methods offering different trade-offs between accuracy and computational complexity were developed. The accuracy of the rules learned by this approach was discussed.

It is obvious that data sets are key to the data science research. Such data falls in three main categories: public data repositories, applications collected data, and synthetic data. Each category of data serves its purpose, e.g. model development in an application of interest. For the best delivery of the research results discussed in this paper, a synthetic data set presented in Table 1 has been constructed. This data set is used to demonstrate the XRule algorithm introduced in this paper. In addition, the XRule algorithm is illustrated with three industrial applications.

The data set in Table 1 illustrates the proposed XRule algorithm for extraction of user-accepted rules. The data set contains 13 input parameters $P_1, \ldots, P_{13}$ and the output labeled 'Decision' with three values, A, B, and C.

The decision tree extracted by the C&RT (classification and regression tree) algorithm from the data set in Table 1 is shown next. Here, the C&RT algorithm of Statistica (a commercial software tool) was used. All statistica-generated decision trees used in this paper, including the one in Figure 2, have been redrawn for better visualisation.

The rule explaining the outcome, here Decision = A is short, IF $P_9 > 2.55$ THEN Decision = A. However, the rules explaining decisions B and C are longer. The length of the rules is one of the constraints incorporated in the XRule algorithm. Some users may prefer that decisions generated from data be explained with rules that are short. Rule redundancy and confidence make additional constraints. Redundant rules enhance user's confidence in the decision-making model. Such rules may also represent different perspectives. Other users may want to extract rules that include specific parameters

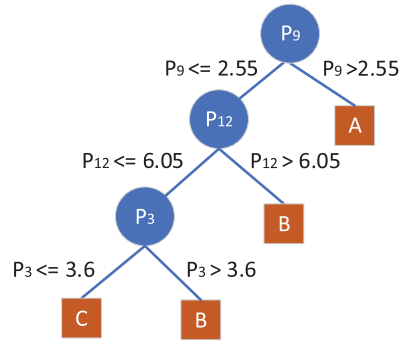

Figure 2. The initial decision tree.

and could be long. The XRule algorithm extracts rules accommodating preferences of the users with different backgrounds.

The following terms are defined for use in the XRule algorithm:

- ✓ *Current parameter set* includes all parameters used to build a decision tree at the corresponding node. The initial current parameter set of the data in Figure 1 includes all input parameters, i.e. All = $\{P_1, \ldots, P_{13}\}$.
- ✓ *Current data set* is the data set for the current parameter set.
- ✓ *Current parent* is the parent node of a decision tree generated from the data set corresponding to this node. For example, P_{12} in Figure 2 is the current parent in the decision tree generated based on the initial parameter set All = $\{P_1, \ldots, P_{13}\}$.
- ✓ *Current parent parameter* is a parameter branching out of the current parent node.
- ✓ *Current parent parameter value* is the value of a parameter branching out of the current parent node. For example, the value of $P_{12} > 6.05$ in Figure 2 is defined based on the current parent node P_{12}.

✓ *XRule* is a decision rule that contributes to explaining the corresponding outcome and meets the user-imposed constraints.
✓ *XRule_Set* is the set of decision rules (XRules) generated. Once all XRules have been generated, the set becomes complete.

The XRule algorithm presented next utilises the C&RT (classification and regression tree) algorithm of Statistica, a commercial software platform. The C&RT algorithm can be replaced with any other decision-tree or decision-rule algorithm. The novelty of the XRule algorithm is in multiple extractions of decision rules in the presence of user-defined constraints that may change over time.

3.1. The XRule algorithm

Step 1: Apply the C&RT (classification and regression tree) algorithm to the current data set. The initial current parameter set includes all input parameters, $\{P_1, \ldots, P_{13}\}$.

Step 2: Check if the current data set has produced an XRule for any of the decisions. If yes, store it in the XRule_Set and go to Step 3; otherwise go to Step 4.

Step 3: If the XRule_Set is not complete, remove from the current data set the parameters involved in the condition of the XRule and go to Step 1; otherwise go to Step 5.

Step 4: If the XRule_Set is not complete, remove from the current parameter set a parameter included in the XRule and go to Step 1; otherwise go to Step 5.

Step 5: Stop when all XRules have been generated.

The data set of Figure 1 is used to illustrate the XRule algorithm.

3.2. Illustrative example

The following two user preferences are considered: (1) the maximum number of parameters included in the condition of each XRule is 1, and (2) the minimum redundancy for each XRule is 2.

The iterations of the XRule algorithm are listed next.

3.2.1. Iteration 1

Step 1: Applying the C&RT algorithm to the original data set has produced the decision tree in Figure 2. The initial current parameter set is All $= \{P_1, \ldots, P_{13}\}$.

Step 2: The current data set has produced the XRule, IF $P_9 > 2.55$ THEN Decision $=$ A, which is added to the XRule_Set. Go to Step 3. Note that this XRule is the only one meeting the first user preference, and therefore the other two branches ending with the nodes P_3 and P_{12} of the tree in Figure 3 are fathomed. Each of the two nodes

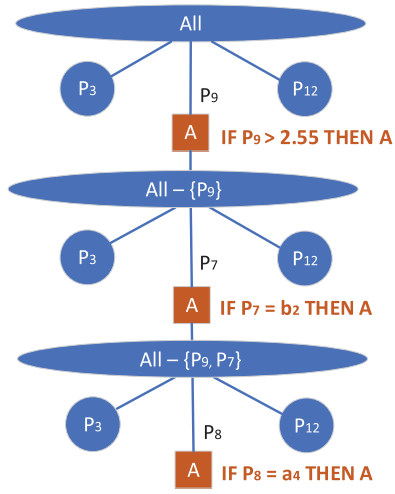

Figure 3. Illustration of the process of deriving the XRule, IF 'One_Condition' THEN Decision = A.

leads to a longer rule (constraint (1) violation), the current parameter set is updated by the removal of parameter P_9. A new current parent node is established with the data corresponding to the current parameter set.

Step 3: Since the XRule_Set is not complete, parameter P_9 is removed from the current parameter set. The current parameter set becomes, All – $\{P_9\} = \{P_1, \ldots, P_8, P_{10}, \ldots, P_{13}\}$.

3.2.2. Iteration 2

Step 1: Applying the C&RT algorithm to the data set with the parameters, All – $\{P_9\}$, is illustrated in Figure 3.

Step 2: The current parent parameter set has produced the XRule, IF $P_7 = b_2$ THEN Decision $=$ A, which is added to the XRule_Set. Go to Step 3.

Step 3: Since the XRule_Set is not complete, parameter P_9 is removed from the current parameter set. The current set of parameter becomes All – $\{P_9\}$

3.2.3. Teration 3

Step 1: Applying the C&RT algorithm to the data set with parameters All – $\{P_9, P_7\}$ is illustrated in Figure 3.

Step 2: The current parent parameter set has produced the XRule, IF $P_8 = a_4$ THEN Decision $=$ A, which is added to the XRule_Set. Go to Step 3.

Step 3: Since the XRule_Set is not complete, parameter P_8 is removed from the current parameter set. The current f parameter set is, All – $\{P_9, P_7, P_8\}$.

3.2.4. Iteration 4

Step 1: Applying the C&RT algorithm to the data set with parameters, All – $\{P_7, P_8, P_9\}$, is illustrated in Figure 4.

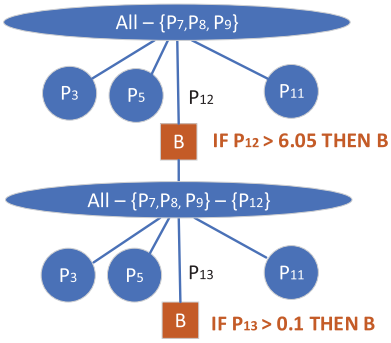

Figure 4. Illustration of the process of deriving the XRule, IF 'One_Parameter' THEN Decision = B.

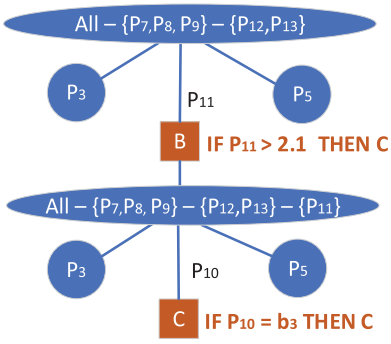

Figure 5. Illustration of the process of deriving the XRule, IF 'One_Parameter' THEN Decision = C.

Step 2: The current parent parameter set has produced the XRule, IF $P_{12} > 6.05$ THEN Decision = B, which is added to the XRule_Set. Go to Step 3.

Step 3: Since the XRule_Set is not complete, parameter P_8 is removed from the current parameter set. The current parameter set becomes, All – {P_7, P_8, P_9, P_{12}}.

3.2.5. Iteration 5

For the current parameter set All – {P_7, P_8, P_9, P_{12}}, the XRule, IF $P_{13} > 0.1$ THEN Decision = B, is generated (see Figure 4).

3.2.6. Iteration 6

For the current parameter set All – {P_7, P_8, P_9, P_{12}, P_{13}}, the XRule, IF $P_{11} > 2.1$ THEN Decision = C, is generated (see Figure 5).

3.2.7. Iteration 7

For the current parameter set All – {P_7, P_8, P_9, P_{11}, P_{12}, P_{13}}, the XRule, IF $P_{10} = b_3$ THEN Decision = C, is generated (see Figure 5).

Since the XRule_Set is complete, the algorithm terminates in Step 5.

The XRule algorithm can be applied by multiple users to derive rules of interest. The model itself to which the XRule algorithm is applied, can be built by one participant or multiple participants. The latter mode of machine learning, known as federated learning, is discussed next.

4. Federated learning

Federated learning (FL) aims at building models with machine learning algorithms by multiple participants without sharing data. It is an emerging distributed machine learning framework primarily intended for privacy protection. Federated learning was introduced by Google in 2015 in collaborative model development (Konečny et al. 2016). Banabilah et al. (2022) defined a process of federated learning as well as reviewed and categorised the literature on federated learning across various application domains, from blockchain and Internet of Things to autonomous driving and industry. Research opportunities and challenges were identified. The literature on federated learning frameworks and technologies was surveyed by Ghimire and Rawat (2022). Security and performance issues in the IoT applications of federated learning were discussed. The authors characterised and compared centralised, distributed, and federated learning. The survey paper by Boobalan et al. (2022) reviewed the published literature on privacy issues of federated learning in various industrial IoT applications, including automotive, energy, and healthcare. Research directions, potential challenges, and different ways of handling big heterogeneous data were outlined.

The performance of models developed in federated learning is usually inferior to those trained in the standard learning mode, in particular, in the presence of non-independent and non-identically distributed data (non-IID). The paper by Zhu et al. (2021) analysed the impact of non-IID data on the quality of machine learning models. Challenges of horizontal and vertical federated learning with non-IID data were discussed. Wang et al. (2022) proposed a federated transfer-learning framework involving a central server and smart devices for smart manufacturing applications with limited training data and high data privacy expectations. The framework was designed for model sharing between the central server and smart devices without exposing the training data. The Internet of Things (IoT) was featured as an enabling technology for connectivity between manufacturing equipment and control systems with business processes and information systems. An architecture named, FedeX, was developed to address the challenges of anomaly detection systems such as detection accuracy, training data and time, and computing resource requirements. The performance of the developed system was compared with the fourteen existing anomaly detection solutions.

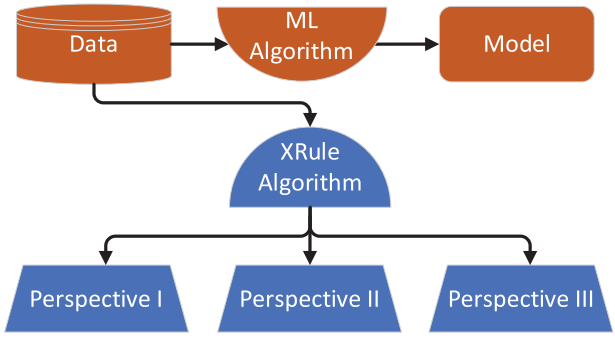

Figure 6. The concept of fXAI based on the XRule algorithm.

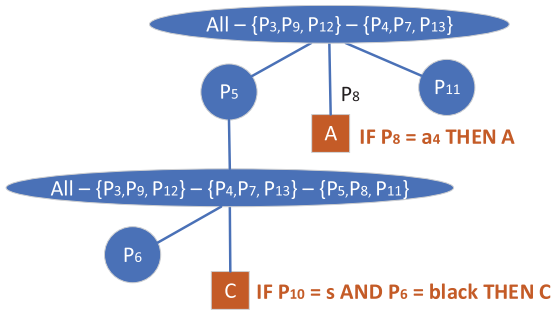

Figure 7. The XRules generated from the data set in Table 1 in a federated mode.

In this paper, a new concept of federated explainable artificial intelligence (fXAI) is proposed. It is analogous to federated learning in machine learning. The proposed fXAI concept is discussed next.

5. Federated explainable artificial intelligence (fXAI)

In analogy to federated learning, explainable AI can be performed in a federated environment, called here federated explainable artificial intelligence (fXAI). While federated learning (FL) focuses on the construction of models by multiple participants for privacy protection, federated XAI (fXAI) supports the extraction of explainable models meeting expectation of different users. Sharing data or information among the users is not a concern in fXAI.

The fXAI concept is implemented with the XRule algorithm as shown in Figure 6. The data is used by both, the machine learning (ML) algorithm and the XRule algorithm.

The ML algorithm produces a model that is explained with the rules produced by the XRule algorithm, here in three perspectives, I, II, and III. Note that the predictive model can be built in a classical or a federated machine learning mode.

The following basic scenarios can be realised while implementing fXAI with the XRule algorithm:

(a) The XRule algorithm is applied to the entire data set by multiple users at the same time with no information exchange between the users. Thus, the XRules for use in different domains are generated. Discussion and sharing of parameters among different users is envisioned.
(b) The XRule algorithm is applied in a sequence, one user at a time. This way the use of parameters to be used in data analysis are controlled.
(c) The XRule algorithm is applied to partial data sets by multiple users. Users would discuss the parameters and the rules as needed. For example, for a subset of data in Figure 1 one of the users could generate the two XRules illustrated in Figure 7.

The left node branching of the bottom node 'All – $\{P_3, P_9, P_{12}\}$ – $\{P_4, P_7, P_{13}\}$ – $\{P_5, P_8, P_{11}\}$' is labelled P_6. The second arc of the bottom node ends with square C and 'IF $P_{10} = s_3$ AND P_6 = black THEN C' next to it.

In addition, users may impose on the rules their own preferences that could emerge while applying the XRule algorithm, e.g. some users may prefer that XRules include parameters that have not being envisioned a priori.

6. Federated XAI in digital manufacturing

Though federated explainable artificial intelligence app lies to many domains, digital manufacturing offers its own characteristics (Kusiak 2022). Factors such as workforce and its training, processes, models, and the manufacturing software and hardware deployed, make a unique signature of digital manufacturing. The manufacturing workforce is trained in processes and methods with physics, mathematics, and computer science prevailing. The diversity of manufacturing processes calls for training in areas such as materials (from aluminium and steel to ceramics and metal powders), processes (from casting and metal forming to injection moulding, metal removal and additive manufacturing), and information technology. The manufacturing environment is largely structured, with the equipment and software, constituting its core. This structure is further enhanced with the manufacturing standards that are usually well received. Change is a general characteristic that deserves attention in digital manufacturing due to competitive pressure. These characteristics impact the nature of the data used in modelling, included federated XAI.

Though there is no uniformity, the science-trained workforce tends to prefer explanations supported by constructs from physics, diagrams, matrices, and charts in

various forms. The explanation methods are expected to be process specific due to the prevailing specialisation in manufacturing.

A survey of the literature indicates that research in federated learning (FL) in manufacturing is scarce. Ge et al. (2022) reported the results of a federated learning (FL) study on failure prediction in manufacturing. Two machine learning algorithms for horizontal and vertical FL scenarios were introduced. Performance of the failure prediction models generated in FL was comparable to the ones of traditional learning.

New applications of FL in manufacturing are likely to emerge. The proposed concept of federated explainable artificial intelligence (fXAI) can be applied to the models built from data in a traditional or a federated mode.

Three applications of federated explainable artificial intelligence (fXAI) in manufacturing are presented next.

Application 1: Printed Circuit Assembly

Quality issues have emerged in a company assembling printed circuit boards (PCBs). To reduce the number of quality inspections, a neural network model has been developed to predict quality (Acceptable, Not_acceptable) of each circuit board. The accuracy of the neural network model was high so that the items predicted as Not_acceptable were inspected. This was the inspection effort significantly reduced. The data used by the neural network model involved parameters related to the components used in the assembly, characteristics of the printed circuits boards, assembly equipment, and operators. This diverse set of parameters implies a wide range of expertise needed to understand the predictive model. The company management and operators insisted on seeing justifications on the predicted quality outcomes. The application itself was first researched by Kusiak and Kurasek (2001).

Each of the three example rules presented next is intended for a specific group of professionals interested in the assessment of the predicted outcomes. The rules predict the quality of each assembly based on the parameters characterising the components to be assembled and the assembly process.

Rule 1. IF Component = Hand_placed AND Vacuum = ON THEN Quality = Acceptable

Rule 2. IF Assembly_line = 2 AND Position = H32 AND THEN Quality = Acceptable

Rule 3. IF Designator = R AND Position = A12 THEN Quality = Not_acceptable

Both predictive outcomes, Acceptable and Not_acceptable, were carefully analysed. In many instances, alternative rules (e.g. using a set of parameters imposed by a specific user) were extracted for a better understanding of the predictive outcomes. In some cases, the predicted result was Quality = Unknown, which implied that the model did not have enough knowledge to predict Quality = Acceptable or Quality = Not_acceptable. Such cases were scrutinised for additional cues that could lead to one of the two preferred outcomes, Acceptable or Not_Acceptable. In some cases, process modifications were made, prior to the PCB production launch.

The two rules in Application 2 presented next explain the outcomes of a neural network model in a semiconductor industry.

Application 2: Semiconductor Industry

Manufacturing of integrated circuits is preceded by wafer production and processing. A wafer is sliced into discs that undergo different manufacturing operations, including polishing of the disc surfaces. The quality of the polishing process is impacted by different parameters ranging from the wafer chemical composition through equipment type to process parameters, all included in the neural network model developed using the previously collected data. This application was originally investigated by Kusiak (2001).

The two rules presented next illustrate the model's predictive outcomes. The outcome is determined based on the material and process parameters.

Rule 1. IF Temperature in [100, 120] AND Pressure in [156, 175] THEN Category = C1

Rule 2. IF Material = AN271Q AND Operator = 060 AND Paste = 8CD807 THEN Category = C2

The neural network model and the XRule algorithm are run ahead of polishing the discs. This provides the production management team with an opportunity to review the predicted outcomes. If the predictive results do not meet the expectations, changes to the process are made. The neural network and the XRule algorithm are usually executed several times.

The manufacturing and service industry are under pressure from the customers to deliver personalised products and services. To meet the customer demands, data science models are needed. The fXAI concept in personalisation of products is illustrated in Application 3.

Application 3: Mass Customisation

Personalisation of products (a customer perspective) is realised with mass customisation in manufacturing. Data-derived models cover various aspects of mass customisation, including predicting the type and quantity of components and assembles needed to meet the customer demand and the manufacturing cost, inventory level, and delivery time objectives. This domain is known for the availability of large volumes of high-quality data. Details of mass customisation applications are presented in Kusiak, Smith, and Song (2007) and Song and Kusiak (2009).

The two rules presented next illustrate the predicted outcomes in mass customisation. These rules explore commonality among components and assembles making products.

Rule 1. IF Option_3 = Yes AND Option_12 = No THEN Option_9 = Yes

Rule 2. IF Option_6 = Yes AND Option_5 = Yes THEN Subassembly = S8

The goal of the predictive model is to meet the customer specifications while minimising the total number of different product variants produced, which leads to meeting the cost, inventory, and delivery time objectives. The XRules are key to accomplishing this goal, as they provide insights into the logic behind forming subassemblies and the final product options. They may also trigger ideas for product and manufacturing changes, some of which would be difficult to discover in a usual manufacturing decision-making environment.

6.1. Research outlook

Given the growing coverage of the XAI topic in the literature, new solutions are likely to emerge. The undergoing digitisation of manufacturing, including the development of digital twins, will enhance visibility of manufacturing processes. As digitisation calls for in-depth understanding of the manufacturing data, its origin, and the flow, digital models may become a backbone of predictive models and XAI solutions. This may create an environment for the development of XAI systems tightly integrated with the digital models. As manufacturing tends to be distributed, federated explainable artificial intelligence (fXAI) is likely to be an asset.

7. Conclusion

The scope and intensity of research and development activities in digital manufacturing are growing. The speed and the degree of monetisation of manufacturing data have become a measurable component of the success. Data science has become an important tool in the value conversion process. It allows to model phenomena that are large in scope are complex with accuracy determined by the data available for modelling. As most of data-derived models are non-explicit, a justification of the decisions generated by the model is expected. The large scale and complexity of such models call for a greater clarity of the decisions to users with diverse expertise, e.g. manufacturing processes, materials, supply chains, management, or information technology. The literature in explainable artificial intelligence (XAI) has focused the model development, methodologies, and tools in support of informed decision-making. The XRule algorithm proposed in the paper is versed in the existing explicit data science algorithms in support of explainable artificial intelligence. The decision-tree algorithm implemented in the XRule algorithm can be replaced with a decision-rule or an association rule algorithm. The ease of incorporation of user-defined requirements (e.g. specific parameters, rule length, or rule redundancy) enhances the usability of the XRule algorithm. The XAI concept was expanded to distributed decision-making scenario in the form of a federated explainable artificial intelligence (fXAI). The concepts introduced in the paper were illustrated with an example and three industrial applications.

Acknowledgement

In memory of Dr Jean-Marie Proth, INRIA, Metz, France – a long-time friend and a research collaborator.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Data availability statement

Data sharing is not applicable to this article as all data is included in the manuscript.

Notes on contributors

Andrew Kusiak is a Professor in the Department of Industrial and Systems Engineering at The University of Iowa, Iowa City and Director of Intelligent Systems Laboratory. He has chaired two different departments, Industrial Engineering, and Mechanical and Industrial Engineering. His current research focuses on applications of computational intelligence in manufacturing, renewable energy, automation, sustainability, and healthcare. He is the author or coauthor of numerous books and hundreds of technical papers published in journals sponsored by professional societies, such as AAAI, ASME, IISE, IEEE, INFORMS, and other societies. He is a frequent speaker at international meetings, conducts professional seminars, and consults for industrial corporations. Dr Kusiak has served in several elected professional society positions as well as editorial boards of over fifty journals, including the Editor position of five different IEEE Transactions. Professor Kusiak is a Fellow of the Institute of Industrial and Systems Engineers and the Editor-in-Chief of the Journal of Intelligent Manufacturing (Springer Nature).

References

Ahmed, I., G. Jeon, and F. Piccialli. 2022. "From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, how, and Where." *IEEE Transactions on Industrial Informatics* 18 (8): 5031–5042. https://doi.org/10.1109/TII.2022.3146552.

Banabilah, S., M. Aloqaily, E. Alsayed, N. Malik, and Y. Jararweh. 2022. "Federated Learning Review: Fundamentals, Enabling Technologies, and Future Applications." *Information Processing and Management* 59 (6): 103061. https://doi.org/10.1016/j.ipm.2022.103061.

Barredo Arrieta, A., N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, et al. 2020. "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI." *Information Fusion* 58: 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

Bin Iqbal, Md. T., A. Muqeet, and S.-H. Bae. 2022. "Visual Interpretation of CNN Prediction Through Layerwise Sequential Selection of Discernible Neurons." *IEEE Access* 10: 81988–82002. https://doi.org/10.1109/ACCESS.2022.3188394.

Boobalan, P., S. P. Ramu, Q.-V. Pham, K. Dev, S. Pandya, P. K. R. Maddikunta, T. R. Gadekallu, and T. Huynh-The. 2022. "Fusion of Federated Learning and Industrial Internet of Things: A Survey." *Computer Networks* 212: 109048. https://doi.org/10.1016/j.comnet.2022.109048.

Buczak, A. L., B. D. Baugher, A. J. Berlier, K. E. Scharfstein, and C. S. Martin. 2022. "Explainable Forecasts of Disruptive Events Using Recurrent Neural Networks." 2022 IEEE international conference on assured autonomy (ICAA), pp. 64–73, https://doi.org/10.1109/ICAA52185.2022.00017.

Cao, H. E. C., R. Sarlin, and A. Jung. 2020. "Learning Explainable Decision Rules via Maximum Satisfiability." *IEEE Access* 8: 218180–218185. https://doi.org/10.1109/ACCESS.2020.3041040.

Cheng, K. C.-C., K. S.-M. Li, S. J. Wang, A. Y.-A. Huang, C.-S. Lee, L. L.-Y. Chen, P. Y.-Y. Liao, and N. C.-Y. Tsai. 2022. "Wafer Defect Pattern Classification with Explainable-Decision Tree Technique." Proceedings of the IEEE international test conference, Anaheim, CA, pp. 549–553, 185603.

Chowdhury, D., A. Sinha, and D. Das. 2023. "XAI-3DP: Diagnosis and Understanding Faults of 3-D Printer with Explainable Ensemble AI." *IEEE Sensors Letters* 7 (1): 1–41. https://doi.org/10.1109/LSENS.2022.3228327.

Cochran, D., J. Smith, B. G. Mark, and E. Rauch. 2022. Information model to advance explainable AI-based decision support systems in manufacturing system design, 1st International Symposium on Industrial Engineering and Automation, ISIEA 2022, LNNS 525, pp. 49-60.

Dey, S., P. Chakraborty, B. C. Kwon, A. Dhurandhar, M. Ghalwash, F. J. Suarez Saiz, K. Ng, D. Sow, K. R. Varshney, and P. Meyer. 2022. "Human-centered Explainability for Life Sciences, Healthcare, and Medical Informatics." *Patterns* 3 (5): 100493. https://doi.org/10.1016/j.patter.2022.100493.

Doran, D., S. Schulz, and T. R. Besold. 2017. What does explainable AI really mean? A New conceptualization of perspectives, arXiv:1710.00794v1 [cs.AI].

Fiok, K., F. V. Farahani, W. Karwowski, and T. Ahram. 2022. "Explainable Artificial Intelligence for Education and Training." *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology* 19 (2): 133–144. https://doi.org/10.1177/15485129211028651.

Ge, N., G. Li, L. Zhang, and Y. Liu. 2022. "Failure Prediction in Production Line Based on Federated Learning: An Empirical Study." *Journal of Intelligent Manufacturing* 33 (8): 2277–2294. https://doi.org/10.1007/s10845-021-01775-2.

Ghimire, B., and D. B. Rawat. 2022. "Recent Advances on Federated Learning for Cybersecurity and Cybersecurity for Federated Learning for Internet of Things." *IEEE Internet of Things Journal* 9 (11): 8229–8249. https://doi.org/10.1109/JIOT.2022.3150363.

Goldman, C. V., M. Baltaxe, D. Chakraborty, J. Arinez, and C. E. Diaz. 2023. "Interpreting Learning Models in Manufacturing Processes: Towards Explainable AI Methods to Improve Trust in Classifier Predictions." *Journal of Industrial Information Integration* 33: 100439. https://doi.org/10.1016/j.jii.2023.100439.

Ha, D. T., N. X. Hoang, N. V. Hoang, N. H. Du, T. T. Huong, and K. P. Tran. 2022. "Explainable Anomaly Detection for Industrial Control System Cybersecurity." *IFAC-PapersOnLine* 55 (10): 1183–1188. https://doi.org/10.1016/j.ifacol.2022.09.550.

He, L., N. Aouf, and B. Song. 2021. "Explainable Deep Reinforcement Learning for UAV Autonomous Path Planning." *Aerospace Science and Technology* 118: 107052. https://doi.org/10.1016/j.ast.2021.107052.

Holzinger, A., B. Malle, A. Saranti, and B. Pfeifer. 2021. "Towards Multi-Modal Causability with Graph Neural Networks Enabling Information Fusion for Explainable AI." *Information Fusion* 71: 28–37. https://doi.org/10.1016/j.inffus.2021.01.008.

Hrnjica, B., and S. Softic. 2020. "Explainable AI in Manufacturing: A Predictive Maintenance Case Study." *IFIP Advances in Information and Communication Technology* 592: 66–73. https://doi.org/10.1007/978-3-030-57997-5_8.

Jakubowski, J. P., S. Bobek Stanisz, and G. Nalepa. 2022. "Roll Wear Prediction in Strip Cold Rolling with Physics-Informed Autoencoder and Counterfactual Explanations." Proceedings of the IEEE 9th international conference on data science and advanced analytics, DSAA 2022, Shenzhen, China, 186596.

Jia, S., Z. Li, N. Chen, and J. Zhang. 2022. "Towards Visual Explainable Active Learning for Zero-Shot Classification." *IEEE Transactions on Visualization and Computer Graphics* 28 (1): 791–800. https://doi.org/10.1109/TVCG.2021.3114793.

Joung, J., and H. M. Kim. 2022. "Explainable Neural Network-Based Approach to Kano Categorisation of Product Features from Online Reviews." *International Journal of Production Research* 60 (23): 7053–7073. https://doi.org/10.1080/00207543.2021.2000656.

Keleko, A. T., K.-F. Bernard, R. H. Ngouna, and A. Tongne. 2023. "Health Condition Monitoring of a Complex Hydrau lic System Using Deep Neural Network and DeepSHAP Explainable XAI." *Advances in Engineering Software* 175: 103339. https://doi.org/10.1016/j.advengsoft.2022.103339.

Konečny, J., H. B. McMahan, D. Ramage, and P. Richtárik. 2016. Federated optimization: Distributed machine learning for on-device intelligence, arXiv:1610.02527.

Kotriwala, A., B. Kloepper, M. Dix, G. Gopalakrishnan, D. Ziobro, and A. Potschka. 2021. XAI for operations in the process industry - Applications, theses, and research directions, Proceedings of the 2021 AAAI Spring Symposium on Combining Machine Learning and Knowledge Engineering, AAAI-MAKE 2021, Palo Alto, CA, 168287.

Kuhnle, A., M. C. May, L. Schäfer, and G. Lanza. 2022. "Explainable Reinforcement Learning in Production Control of job

Shop Manufacturing System." *International Journal of Production Research* 60 (19): 5812–5834. https://doi.org/10.1080/00207543.2021.1972179.

Kusiak, A. 2001. "Rough set Theory: A Data Mining Tool for Semiconductor Manufacturing." *IEEE Transactions on Electronics Packaging Manufacturing* 24 (1): 44–50. https://doi.org/10.1109/6104.924792.

Kusiak, A. 2022. "From Digital to Universal Manufacturing." *International Journal of Production Research* 60 (1): 349–360. https://doi.org/10.1080/00207543.2021.1948137.

Kusiak, A. 2023. "Predictive Models in Digital Manufacturing: Research, Applications, and Future Outlook." *International Journal of Production Research* 61 (17): 6052–6062. https://doi.org/10.1080/00207543.2022.2122620.

Kusiak, A., and C. Kurasek. 2001. "Data Mining of Printed-Circuit Board Defects." *IEEE Transactions on Robotics and Automation* 17 (2): 191–196. https://doi.org/10.1109/70.928564.

Kusiak, A., M. R. Smith, and Z. Song. 2007. "Planning Product Configurations Based on Sales Data." *IEEE Transactions on Systems, Man, and Cybernetics: Part C* 37 (4): 602–609. https://doi.org/10.1109/TSMCC.2007.897503.

Lee, M., J. Jeon, and H. Lee. 2022. "Explainable AI for Domain Experts: A Post hoc Analysis of Deep Learning for Defect Classification of TFT–LCD Panels." *Journal of Intelligent Manufacturing* 33 (6): 1747–1759. https://doi.org/10.1007/s10845-021-01758-3.

Li, X.-H., C. C. Cao, Y. Shi, W. Bai, H. Gao, L. Qiu, C. Wang, et al. 2022. "A Survey of Data-Driven and Knowledge-Aware EXplainable AI." *IEEE Transactions on Knowledge and Data Engineering* 34 (1): 29–49. https://doi.org/10.1109/TKDE.2020.2983930.

Lorentz, J., T. Hartmann, A. Moawad, F. Fouquet, and D. Aouada. 2021. "Explaining Defect Detection with Saliency Maps." Proceedings of the 34th international conference on industrial, engineering and other applications of applied intelligent systems, IEA/AIE 2021, LNAI Vol. 12799. pp. 506–518.

Makridis, G., S. Theodoropoulos, D. Dardanis, I. Makridis, M. M. Separdani, G. Fatouros, D. Kyriazis, and P. Koulouris. 2022. "XAI Enhancing Cyber Defense Against Adversarial Attacks in Industrial Applications." 5th IEEE international image processing, applications and systems conference, IPAS 2022. Genova, Italy, 187000.

Markus, A. F., J. A. Kors, and P. R. Rijnbeek. 2021. "The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies." *Journal of Biomedical Informatics* 113: 103655. https://doi.org/10.1016/j.jbi.2020.103655.

Mohamed, E., K. Sirlantzis, and G. Howells. 2022. "A Review of Visualisation-as-Explanation Techniques forConvolutional Neural Networks and Their Evaluation." *Displays* 73: 102239. https://doi.org/10.1016/j.displa.2022.102239.

Nakhle, F., and A. L. Harfouche. 2021. "Ready, Steady, go AI: A Practical Tutorial on Fundamentals of Artificial Intelligence and its Applications in Phenomics Image Analysis." *Patterns* 2 (9): 100323. https://doi.org/10.1016/j.patter.2021.100323. https://www.cell.com/patterns/fulltext/S2666-3899(21)00171-9.

Nguyen, S., and B. Tran. 2022. "XMAP: EXplainable Mapping Analytical Process." *Complex & Intelligent Systems* 8 (2): 1187–1204. https://doi.org/10.1007/s40747-021-00583-8.

Roscher, R., B. Bohn, M. F. Duarte, and J. Garcke. 2020. "Explainable Machine Learning for Scientific Insights and Discoveries." *IEEE Access* 8: 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199.

Rožanec, J. M., I. Novalija, P. Zajec, K. Kenda, H. Tavakoli Ghinani, S. Suh, E. Veliou, et al. 2022. "Human-centric Artificial Intelligence Architecture for Industry 5.0 Applications." *International Journal of Production Research*. https://doi.org/10.1080/00207543.2022.2138611.

Schmetz, A., C. Vahl, Z. Zhen, D. Reibert, S. Mayer, D. Zontar, J. Garcke, and C. Brecher. 2021. "Decision Support by Interpretable Machine Learning in Acoustic Emission Based Cutting Tool Wear Prediction, IEEE International Conference on Industrial Engineering and Engineering Management." *IEEM* 2021: 629–633. https://doi.org/10.1109/IEEM50564.2021.9673044.

Song, Z., and A. Kusiak. 2009. "Optimization of Product Configurations with a Data-Mining Approach." *International Journal of Production Research* 47 (7): 1733–1751. https://doi.org/10.1080/00207540701644235.

Spinner, T., U. Schlegel, H. Schafer, and M. El-Assady. 2020. "explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning." *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 1064–1074. https://doi.org/10.1109/TVCG.2019.2934629.

Sutthithatip, S., S. Perinpanayagam, S. Aslam, and A. Wileman. 2021. "Explainable AI in Aerospace for Enhanced System Performance." 2021 IEEE/AIAA 40th digital avionics systems conference (DASC), pp. 1–7.

Taj, I., and N. Z. Jhanjhi. 2022. "Towards Industrial Revolution 5.0 and Explainable Artificial Intelligence: Challenges and Opportunities." *International Journal of Computing and Digital Systems* 12 (1): 285–310. https://doi.org/10.12785/ijcds/120124.

Wang, K. I.-K., X. Zhou, W. Liang, Z. Yan, and J. She. 2022. "Federated Transfer Learning Based Cross-Domain Prediction for Smart Manufacturing." *IEEE Transactions on Industrial Informatics* 18 (6): 4088–4096. https://doi.org/10.1109/TII.2021.3088057.

Zhu, H., J. Xu, S. Liu, and Y. Jin. 2021. "Federated Learning on non-IID Data: A Survey." *Neurocomputing* 465: 371–390. https://doi.org/10.1016/j.neucom.2021.07.098.