Procedia CIRP

Information retrieval for AI-supported product repair and re-design

--Manuscript Draft--

Manuscript Number:	PROCIR-D-24-01127R2
Article Type:	VSI: PROCIRP_CIRP DESIGN 2025
Section/Category:	VSI: PROCIRP_CIRP DESIGN 2025
Corresponding Author:	Benjamin Schleich, Dr.-Ing. Technical University of Darmstadt Darmstadt, GERMANY
First Author:	Lars Arnemann
Order of Authors:	Lars Arnemann
	Sven Winter
	Ilyass Afkir
	Benjamin Schleich, Prof. Dr.-Ing.
Abstract:	Repair data of former product generations has enormous potential for supporting the detection of product weaknesses but is often not available as useful information in product design. In this paper, a novel method for intelligent information retrieval from repair data is presented. Based on unsupervised machine learning techniques, this method identifies failure-relevant patterns and dependencies of interconnected parts and assemblies. These gained insights give a valuable contribution to the systematic re-design of next-generation products. Moreover, the approach supports efficient repair and sustainable operational processes. The validation of the artificial intelligence-based assistance system is shown for welding equipment.

Declaration of interests

☐The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Benjamin Schleich reports financial support was provided by Federal Ministry of Education and Research Bonn Office. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 00 (2025) 000–000

www.elsevier.com/locate/procedia

35th CIRP Design 2025

Information retrieval for AI-supported product repair and re-design

Lars Arnemann[a], Sven Winter[a], Ilyass Afkir[a], Benjamin Schleich[a,]*

[a]*Product Life Cycle Management, Technical University of Darmstadt, Otto-Berndt-Str. 2, 64287 Darmstadt, Germany*

* Corresponding author. Tel.: +49 6151 16-21791; fax: +49 6151 16-21793. *E-mail address:* schleich@plcm.tu-darmstadt.de

Abstract

Repair data of former product generations has enormous potential for supporting the detection of product weaknesses but is often not available as useful information in product design. In this paper, a novel method for intelligent information retrieval from repair data is presented. Based on unsupervised machine learning techniques, this method identifies failure-relevant patterns and dependencies of interconnected parts and assemblies. These gained insights give a valuable contribution to the systematic re-design of next-generation products. Moreover, the approach supports efficient repair and sustainable operational processes. The validation of the artificial intelligence-based assistance system is shown for welding equipment.

Keywords: Design; Machine Learning; Repair

1. Introduction and motivation

Driven by fierce international competition and high customer requirements regarding quality, cost, and sustainability, the fast and systematic re-design of products is a major challenge for many companies. In order to cope with this challenge, the digitalization of product development processes, the targeted moderation of information flows throughout the product development cycle, and careful data analysis and knowledge retrieval using artificial intelligence (AI) are crucial [1]. In this regard, repair data of former product generations has enormous potential for supporting the detection of product weaknesses but is often not available as useful information in product design. Particularly, data from repair processes, which have been necessary due to product failures or malfunctions, can be used to gain insights about which design features should be re-designed in future product generations. However, even though various assistance systems support product (re-) design activities [2], there is a lack of a concept for the targeted analysis of repair data, which provides the opportunity to identify critical design features and speed up product development.

Motivated by this, the paper presents a novel method for intelligent information retrieval from repair data. This method is based on unsupervised machine learning techniques and enables the identification of relevant patterns and dependencies between interconnected assemblies. The gained insights give a valuable contribution to the systematic re-design of next-generation products and assemblies. Moreover, the approach supports efficient repair and sustainable operational processes.

In the following sections, the state of the art concerning the application of assistance systems and artificial intelligence in product design is discussed. After that, the overall method behind the AI-based assistance system is highlighted and the validation of the method on an industrial case study is explained. Finally, we critically discuss our approach and provide an outlook on future research.

2212-8271

2. Background and state of the art

Before highlighting the overall approach to the AI-based repair data analysis, the reader is equipped with background on unsupervised machine learning and the state of the art concerning the application of AI and assistance systems in product development is discussed.

2.1. Background on unsupervised machine learning

While supervised machine learning approaches require human-labeled input data, unsupervised machine learning is critical for extracting patterns from large, unlabeled datasets. Significant techniques involve dimensionality reduction, clustering, and association rule mining. To achieve this goal, the K-means clustering algorithm and the apriori association mining algorithm have exceptional potential to reveal such patterns.

More specifically, the goal of K-means clustering is to minimize the following objective function J:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2$$

$$r_{nk} = \begin{cases} 1, if\ k = argmin_j \left\|x_n - \mu_j\right\|^2 \\ 0, \qquad\qquad otherwise \end{cases}$$

where r_{nk} is an indicator variable that checks if μ_k is the nearest cluster center to the point x_n [3].

The apriori algorithm discovers frequently occurring item sets in a dataset by utilizing the principle that all subsets of a frequent itemset should also be frequent. It searches through a set of items, $I = \{i_1, i_2, \ldots, i_n\}$, to identify those combinations that surpass a predetermined support threshold [4]. Subsequently, the algorithm creates rules from these item sets and evaluates their validity using a confidence metric [4]. This procedure reveals vital associations, providing understanding into the fundamental structure of intricate systems.

2.2. AI and assistance systems in product development

In order to design products, which fulfil fierce customer requirements during operation as well as requirements from manufacturing, assembly, and end-of-life, the processing and reuse of knowledge is crucial in engineering [5]. In this context, digital assistance systems that leverage AI algorithms for knowledge discovery in the database can enhance decision-making processes, provide predictive insights, and streamline design workflows [2]. The resulting symbiotic relationship between AI and digital assistants leads to more efficient, accurate, and innovative approaches to product development [1].

While the application of artificial intelligence machine learning in manufacturing and production engineering has shown great potential [6], such as specifically for the realization of self-repair capabilities of production systems [7] or the energy load forecasting of production systems [8], particularly, the application of artificial intelligence has proven to be effective in supporting various design tasks, such as, among others, the customer segmentation [9], the design automation of mechanical components [10], or the generation of synthetic part representatives in tolerancing [11]. Moreover, integrating artificial intelligence and advanced digital assistance systems in the maintenance industry constitutes a significant advancement in industrial operations [12]. In such settings, AI employs intricate algorithms to examine various data sources, including repair histories, detailed performance metrics, and real-time sensor data. This comprehensive data analysis enables the prompt detection of possible malfunctions and simplifies maintenance planning.

3. Data model and assistance system conceptualization

Motivated by the increasing possibilities of AI in engineering design and the potentials of exploiting repair data for product improvement, this paper proposes a novel framework for the knowledge retrieval from repair data. In doing so, the main research questions, which are answered in the paper, refer to the design of a suitable data model for repair data as well as the conceptualization of an AI-based design assistance system.

3.1. Data model for repair data

An optimized repair data model is essential for organizing and standardizing the unique data elements associated with repair processes. To ensure precise and consistent data collection, a stringent standardization process is rigorously followed. For each step of the repair process outlined in Fig. 1, technicians are required to manually collect data and enter it into an enterprise resource planning (ERP) system.

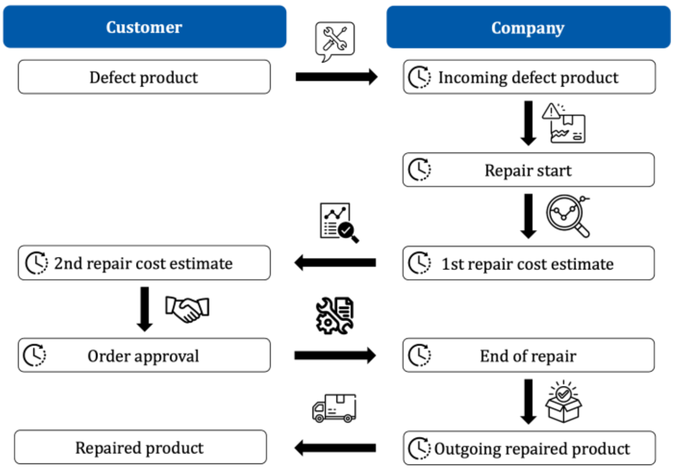

Figure 1 - Typical industrial repair process

This system records and archives the data in a SQL database. The database serves as the foundational layer for the data warehouse. Key data elements, also known as features in the realm of data science, that are collected and recorded include ServiceOrderID, ProductID, RepairTimestamp_{Step}, Customer-ID, TechnicianID, PartID, and PartQuantity.

A unique "ServiceOrderID" identifies each repair instance, with "ProductID" indicating the specific item undergoing repair. The repair workflow is recorded sequentially using "Repair-Timestamp_{Step}" entries, where "{Step}" denotes different stages in the repair process, as illustrated in Fig. 1. To keep track of customers and technicians, their associations are maintained through "CustomerID" and "TechnicianID," correspondingly. Meanwhile, the "PartID" and "PartQuantity" are used to monitor the used parts for repair.

3.2. Assistance system conceptualization and implementation

As a starting point for the conceptualization of repair data use in product development, the Reference Architectural Model for Industry 4.0 (RAMI 4.0) is utilized. This three-dimensional framework aligns operational assets and information technology (IT) components within a structured life cycle and layer model, simplifying the complexities of industrial digital transformation by providing a roadmap for leveraging i. a. repair data at different hierarchical levels [13].

Repair data is employed primarily within the “Instance” phase of the RAMI 4.0 lifecycle value stream, which focuses on addressing immediate technical issues at lower hierarchical tiers, including the product and field device levels. However, for product design, repair data is instrumental in informing the “Type” phase of the life cycle value stream framework. This integration is crucial as it significantly influences the enhancement processes at more advanced hierarchical echelons, notably at the enterprise and connected world levels. Unlike the confined scope of repair data within the repair process, its application in product design yields substantially more value but poses intricate challenges requiring detailed analysis and gradual implementation of improvements.

Our proposed assistance system aims to optimize and automate the use of repair data for product design by utilizing three primary components: data acquisition, data pre-processing, and data analysis. This approach eliminates complexities and is depicted in Fig. 2. The implementation of these components follows a bespoke adaptation of the Data Mining Methodology for Engineering Applications (DMME), which expands upon the principles of the Cross Industry Standard Process for Data Mining (CRISP-DM) [14]. The components are aligned with the stages of DMME, as outlined in Table 1, ensuring a coherent methodology and workflow.

Table 1 - Data Assistance System and DMME Components

Data Assistance Components	DMME Components
Data Acquisition	Business Understanding, Technical Understanding, Technical Realization
Data Preprocessing	Data Understanding, Data Preparation
Data Analysis	Modeling, Evaluation, Technical Implementation, Deployment

Data acquisition. The data acquisition component serves as the basis of the entire assistance system. The objective is to automatically extract two primary repair datasets from the data warehouse using a network management system. The initial dataset furnishes an outline of the repair process, detailing step-by-step timelines and information about defective products involved. The second dataset concentrates on specific defective parts, including their respective types, quantities, and defect characteristics.

Data preprocessing. In the development of our digital assistance system, a systematic approach is taken to the data pre-processing component to ensure the accuracy and usefulness of the repair data. This process consists of several sequential steps, as displayed in Fig. 2. The first phase involves meticulous data cleaning of the two extracted data sets to identify and remove errors and inconsistencies from repair data that are often found, mainly due to human input errors. Such inaccuracies may potentially lead to erroneous conclusions during product re-design. The approach to handling this in big datasets requires the complete elimination or interpolation of such data points to ensure a high quality of the dataset [15].

The next step is to merge the two datasets into one comprehensive dataset. The “ServiceOrderID” feature serves as the principal integration point. After merging, the data is arranged in chronological order, with recent outgoing repaired products timestamps (“RepairTimestep_Outgoing”). A further pre-processing step involves the creation of new features, specifically the calculation of the exact duration between repair process steps as shown in Fig. 1, and the inclusion of a “unit_of_work” feature, which aims to measure the time a technician spends on repair activities, excluding ancillary tasks. The final stage of our data pre-processing component is critical data filtering. This process removes redundant and non-critical components from the repair data set. The focus is on commonplace sundry supplies used frequently during repairs and exceedingly rare or specialized parts. These have minimal impact on product re-design and development decisions. The filtering process by selectively eliminating such data emphasizes more critical and influential repair parts. The output of this data preprocessing component is a fully pre-processed dataset, ready for further in-depth analysis.

Data analysis. A goal-driven development process is key to efficiently and successfully realizing our data analysis component. The implemented data analysis methods highlighted in Fig. 2 are classified chronologically into three distinct categories: descriptive, diagnostic, and predictive analyses, as listed in Table 2. A similar classification was proposed in the framework presented in [16] for designing and specifying data analytics projects.

Table 2 – Classification of data analysis methods

Data Analysis Method	Classification
Time Analysis	Descriptive
Component Analysis	Descriptive
Dependency Analysis	Diagnostic
Clustering	Diagnostic
Demand Forecasting	Predictive

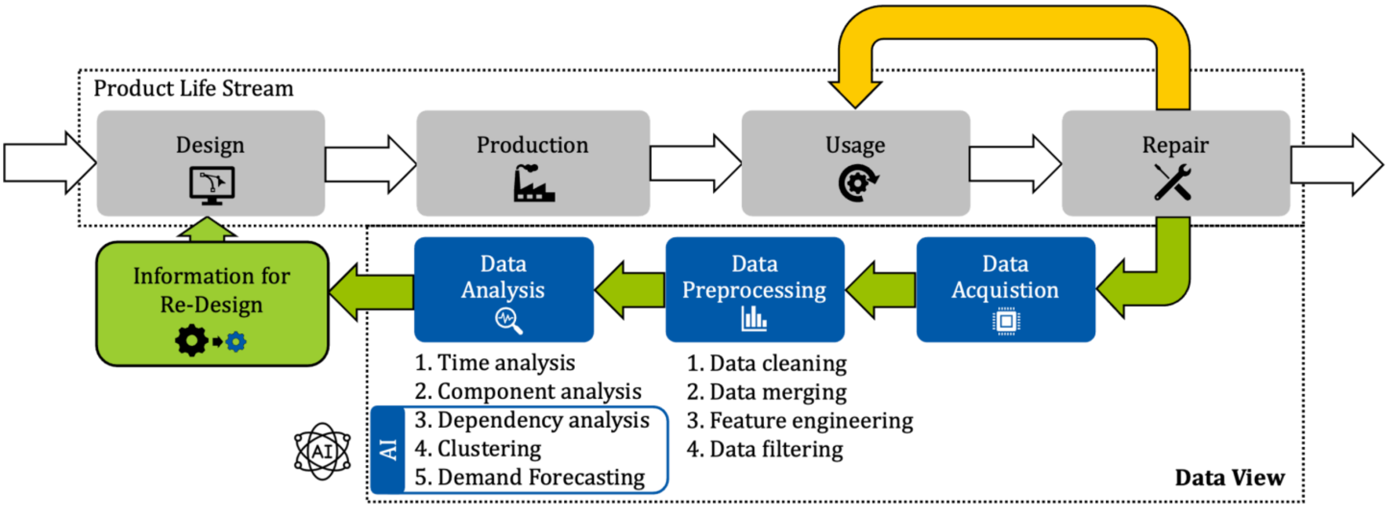

Figure 2 - Integration of the proposed AI-based assistance system (Data View) into the Product Life Stream

The initial step shared by all data analysis methods outlined in Table 2 is to enhance the interpretability and ease of reasoning for product re-design through a contextualization process [15]. This step involves filtering the pre-processed dataset and selectively analyzing data within a certain time frame, focusing on the "RepairTimestep_Outgoing" feature and targeting a specific product using the "ProductID" feature. The culmination of every data analysis method is the formation of a structured results dataset, which can be optionally converted into visual representations.

In line with these common steps, our data analysis procedure starts with a descriptive analysis. This phase encompasses time and component analyses to provide an initial, comprehensive view and preliminary insights into the repair dataset. The time analysis aims to understand the proficiency and effectiveness of the overall repair process. By calculating the average number of workdays per repair instance for each repair process step, using the “RepairTimestamps_{Step}” feature, potential areas for improvement, such as the speed and quality of executing these steps, can be pinpointed.

The component analysis aims to provide a clear insight of the repaired products and their defective parts. This is instrumental in understanding recurring trends and common issues associated with specific parts, as well as in evaluating the workload linked to each repaired product. To achieve this several key metrics are calculated as listed in Table 3.

Table 3 - Key metrics of the components analysis

Key Metrics	Feature
Service order quantity	“ServiceOrderID”
Work units	“unit_of_work”
Avg. work units per repair instance	“unit_of_work”, “SerivceOrderID”
Service order cycle	“ServiceOrderID”, “RepairTimestep_Outgoing”
Part quantity	“PartQuantity”, “PartID”
Avg. part quantity per repair instance	“ServiceOrderID”, “PartQuantity”, “PartID”

The next phase involves diagnostic analytics, which aims to uncover the underlying patterns that account for the events observed in the previous descriptive analytics phase. Unsupervised machine learning techniques, specifically K-means clustering and apriori for association rule mining, are used to delve deeper into the repair data. The main objective of association rule mining is to reveal the relationships and dependencies among various components in a given dataset. This analytical technique is particularly significant for pinpointing failure points that could lead to cascading effects in interlinked systems. Such insights are crucial during product re-design phases as they enable engineers to fortify critical dependencies or restructure components to minimize the likelihood of systemic failures. This study employs the apriori algorithm to analyze data. Firstly, the algorithm iterates through each "ServiceOrderID" for a given product to compile a list of the parts ("PartID") utilized in each order. This compilation then forms a list of lists, with each sub list representing the individual parts used in a specific service order. The subsequent step is to identify frequent item sets comprising components commonly appearing across service orders. To establish what qualifies as 'frequent,' a minimum support threshold is established. In this context, support refers to the proportion of service orders featuring a particular combination of parts. A higher support value indicates that the components are more commonly found in the dataset. Once these frequent item sets are identified, the algorithm generates association rules, which predict the probability of certain components appearing together. The strength of these rules is quantified through confidence, which evaluates the frequency with which a specific relationship holds across all service orders. For example, a high confidence value in a rule incorporating components X and Y suggests that the likelihood of component Y being defective is high when com-ponent X is defective. This guarantees that only the most statistically significant patterns are considered, providing invaluable insights into component relationships and trends and facilitating strategic decision-making for product re-design and maintenance.

The objective of utilizing clustering to categorize components by their criticality is a crucial element in optimizing repair processes and guiding product re-design. This classification enables the identification of parts more susceptible to failure, necessitating greater focus on maintenance and design improvement. We use the K-means algorithm and certain features in our dataset to achieve this classification: "PartQuantity" and the frequency of every part's appearance in service orders, as identified by the "ServiceOrderID." The final stage of our data analytics process incorporates predictive analytics, which is focused on forecasting potential faulty product components. We use historical repair data for features such as "PartQuantity" and "PartID" in this stage. This information is the basis for our predictive models, which accurately predict which product components will likely be faulty in the next month.

4. Assistance system application and validation

The AI-based assistance system described in section 3 has been implemented for the repair data analysis in the industrial case of plastic welding equipment and has significantly enhanced repair procedures and product re-design. The system used unsupervised machine learning techniques to analyze four years of historical repair data. The dataset consisted of 33,727 data points and 32 features after pre-processing. Table 4 describes one of the analyzed products, which could be improved by clustering its critical parts, as shown exemplary for the top 5 critical parts in Fig. 3.

Table 4 - Characteristics of analysed plastic welding product

Characteristic	Metric
Max. Welding capacity	1.8 kg/h, Ø 4 mm; 1.1 kg/h, Ø 3 mm
Welding material	PP / PE / PVDF
Filler metal	Round cord Ø 3/4 mm
Application range	Wall thicknesses 4 - 15 mm
Weight	5.5 kg
Length	470 mm
Electronic speed control	Yes
Heating tape	400 W
With power box	No
Preheating fan	2300 W

Particularly, it can be seen from Fig. 3 that the use of carbon brushes in electric motors highlights their importance as crucial components and their vulnerability to failure, which could be improved by re-design of future product generations. Moreover, their replacement establishes a significant association, leading to concurrent changes by replacing the deep groove ball bearing and its oil as a lubricant with a confidence level of 0.766. This highlights a critical interdependency identified through association rules, which improves the repair process and increases the mean time between repair.

Moreover, the assistance system's user-friendliness is reinforced by a graphical user interface (GUI) for the time and component analysis. This assists technicians and designers in efficiently identifying critical areas for repair and re-design.

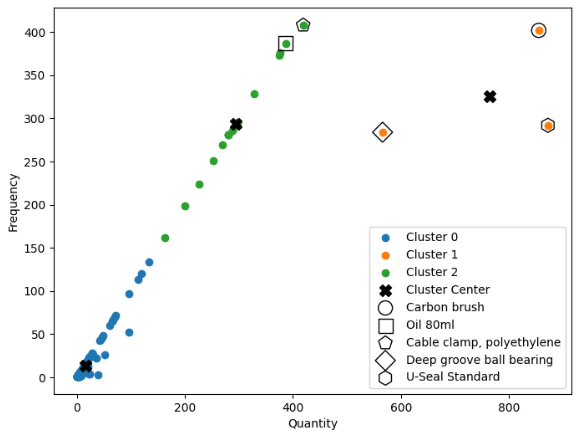

Figure 3 - Clustering of critical parts

5. Critical discussion and outlook

Analyzing repair data allows for accurate and efficient repair procedures and enhanced product re-design. The application of sophisticated machine learning methods such as clustering and association rule mining enables the detection of complex patterns and interdependencies, which is pivotal for informed and efficient product re-design. This method to anticipate potential malfunctions aids in the durability and dependability of products. Motivated by this, the paper describes a data model and novel AI-based assistance system for the repair data analysis as well as the successful application of this system to plastic welding equipment.

However, the approach carries some limitations. Reliance on extensive and high-quality data could hinder the system's effectiveness due to data gaps or biases. Additionally, integrating such a sophisticated AI system into existing workflows presents challenges like complexity and significant resource investment.

Future research should concentrate on improving the system's data collection and processing capabilities to minimize biases and improve predictive accuracy. Efforts should also be made to streamline the integration of this AI system into diverse operational environments, improving its usability and adoption. Furthermore, expanding the system’s adaptability to include a wider range of product types and repair scenarios would greatly increase its usefulness and effectiveness.

Acknowledgements

This research was funded in the DiNaPro-project by the German Federal Ministry of Education and Research (BMBF) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL in order to support the proper language editing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Roucoules, L. & Anwer, N. (2021). Coevolution of digitalisation, organisations and Product Development Cycle. CIRP Annals, 70(2), 519-542. https://doi.org/10.1016/j.cirp.2021.05.003

[2] Stark, R., Brandenburg, E., & Lindow, K. (2021). Characterization and application of assistance systems in digital engineering. CIRP Annals, 70(1), 131-134. https://doi.org/10.1016/j.cirp.2021.04.061

[3] Arora, P., & Varshney, S. (2016). Analysis of k-means and k-medoids algorithm for big data. Procedia Computer Science, 78, 507-512. https://doi.org/10.1016/j.procs.2016.02.095

[4] Schmitt, R. H., Hüttemann, G., & Münker, S. (2021). A priori performance assement of line-less mobile assembly systems. CIRP Annals, 70(1), 389-392. https://doi.org/10.1016/j.cirp.2021.04.059

[5] Kügler, P., Dworschak, F., Schleich, B., & Wartzack, S. (2023). The evolution of knowledge-based engineering from a design research perspective: Literature review 2012-2021. Advanced Engineering Informatics, 55, 101892. https://doi.org/10.1016/j.aei.2023.101892

[6] Gao R X, Krüger J, Merklein M, Möhring H-C, Váncza J. (2024). Artificial Intelligence in manufacturing: State of the art, perspectives, and future directions. CIRP Annals 2024, 73:2, 723-749, https://doi.org/10.1016/j.cirp.2024.04.101

[7] Epureanu B I, Li X, Nassehi A, Koren Y. (2020). Self-repair of smart manufacturing systems by deep reinforcement learning. CIRP Annals 2020, 69:1, 421-424. https://doi.org/10.1016/j.cirp.2020.04.008

[8] Zink R, Ioshchikhes B, Weigold M. (2024). Concept drift monitoring for industrial load forecasting with artificial neural networks. Procedia CIRP 2024. 130, 120-125. https://doi.org/10.1016/j.procir.2024.10.065

[9] Hu, X., Liu, A., Li, X., Dai, Y., & Nakao, M. (2023). Explainable AI for customer segmentation in product development. CIRP Annals, 72(1), 89-92. https://doi.org/10.1016/j.cirp.2023.03.004

[10] Dworschak, F., Dietze, S., Wittmann, M., Schleich, B., & Wartzack, S. (2022). Reinforcement Learning for Engineering Design Automation. Advanced Engineering Informatics, 52, 101612. https://doi.org/10.1016/j.aei.2022.101612

[11] Schleich, B., Qie, Y., Wartzack, S., & Anwer, N. (2022). Generative adversarial networks for tolerance analysis. CIRP Annals, 71(1), 133-136. https://doi.org/10.1016/j.cirp.2022.03.021

[12] Gay, A., Voisin, A., Iung, B., Do, P., Bonidal, R., Khelassi, A. (2022). Data Augmentation-based Prognostics for Predictive Maintenance of Industrial System, CIRP Annals, Volume 71(1):409-412. https://doi.org/10.1016/j.cirp.2022.04.005

[13] Baptista, L. F., & Barata, J. (2021). Piloting Industry 4.0 in SMEs with RAMI 4.0: an enterprise architecture approach. Procedia Computer Science, 192, 2826-2835. https://doi.org/10.1016/j.procs.2021.09.053

[14] Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications–a holistic extension to the CRISP-DM model. Procedia CIRP, 79, 403-408. https://doi.org/10.1016/j.procir.2019.02.106

[15] Gao, R. X., Wang, L., Helu, M., & Teti, R. (2020). Big data analytics for smart factories of the future. CIRP Annals, 69(2), 668-692. https://doi.org/10.1016/j.cirp.2020.05.002

[16] Kühn A, Joppen R, Reinhart F, Röltgen D, von Enzberg S, Dumitrescu R. (2018). Analytics Canvas - A Framework for the Design and Specification of Data Analytics Projects. Procedia CIRP 2018, 70, 162-167. https://doi.org/10.1016/j.procir.2018.02.031

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 00 (2025) 000–000

www.elsevier.com/locate/procedia

35th CIRP Design 2025

Information retrieval for AI-supported product repair and re-design

Lars Arnemann[a], Sven Winter[a], Ilyass Afkir[a], Benjamin Schleich[a,*]

[a]*Product Life Cycle Management, Technical University of Darmstadt, Otto-Berndt-Str. 2, 64287 Darmstadt, Germany*

* Corresponding author. Tel.: +49 6151 16-21791; fax: +49 6151 16-21793. *E-mail address:* schleich@plcm.tu-darmstadt.de

Abstract

Repair data of former product generations has enormous potential for supporting the detection of product weaknesses but is often not available as useful information in product design. In this paper, a novel method for intelligent information retrieval from repair data is presented. Based on unsupervised machine learning techniques, this method identifies failure-relevant patterns and dependencies of interconnected parts and assemblies. These gained insights give a valuable contribution to the systematic re-design of next-generation products. Moreover, the approach supports efficient repair and sustainable operational processes. The validation of the artificial intelligence-based assistance system is shown for welding equipment.

Peer-review under responsibility of the scientific committee of the 35th CIRP Design 2025

Keywords: Design; Machine Learning; Repair

1. Introduction and motivation

Driven by fierce international competition and high customer requirements regarding quality, cost, and sustainability, the fast and systematic re-design of products is a major challenge for many companies. In order to cope with this challenge, the digitalization of product development processes, the targeted moderation of information flows throughout the product development cycle, and careful data analysis and knowledge retrieval using artificial intelligence (AI) are crucial [1]. In this regard, repair data of former product generations has enormous potential for supporting the detection of product weaknesses but is often not available as useful information in product design. Particularly, data from repair processes, which have been necessary due to product failures or malfunctions, can be used to gain insights about which design features should be re-designed in future product generations. However, even though various assistance systems support product (re-) design activities [2], there is a lack of a concept for the targeted analysis of repair data, which provides the opportunity to identify critical design features and speed up product development.

Motivated by this, the paper presents a novel method for intelligent information retrieval from repair data. This method is based on unsupervised machine learning techniques and enables the identification of relevant patterns and dependencies between interconnected assemblies. The gained insights give a valuable contribution to the systematic re-design of next-generation products and assemblies. Moreover, the approach supports efficient repair and sustainable operational processes.

In the following sections, the state of the art concerning the application of assistance systems and artificial intelligence in product design is discussed. After that, the overall method behind the AI-based assistance system is highlighted and the validation of the method on an industrial case study is explained. Finally, we critically discuss our approach and provide an outlook on future research.

2212-8271
Peer-review under responsibility of the scientific committee of the 35th CIRP Design 2025

2. Background and state of the art

Before highlighting the overall approach to the AI-based repair data analysis, the reader is equipped with background on unsupervised machine learning and the state of the art concerning the application of AI and assistance systems in product development is discussed.

2.1. Background on unsupervised machine learning

While supervised machine learning approaches require human-labeled input data, unsupervised machine learning is critical for extracting patterns from large, unlabeled datasets. Significant techniques involve dimensionality reduction, clustering, and association rule mining. To achieve this goal, the K-means clustering algorithm and the apriori association mining algorithm have exceptional potential to reveal such patterns.

More specifically, the goal of K-means clustering is to minimize the following objective function J:

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|x_n - \mu_k\|^2$$

$$r_{nk} = \begin{cases} 1, if\ k = argmin_j \left\|x_n - \mu_j\right\|^2 \\ 0, \quad otherwise \end{cases}$$

where r_{nk} is an indicator variable that checks if μ_k is the nearest cluster center to the point x_n [3].

The apriori algorithm discovers frequently occurring item sets in a dataset by utilizing the principle that all subsets of a frequent itemset should also be frequent. It searches through a set of items, $I = \{i_1, i_2, \dots, i_n\}$, to identify those combinations that surpass a predetermined support threshold [4]. Subsequently, the algorithm creates rules from these item sets and evaluates their validity using a confidence metric [4]. This procedure reveals vital associations, providing understanding into the fundamental structure of intricate systems.

2.2. AI and assistance systems in product development

In order to design products, which fulfil fierce customer requirements during operation as well as requirements from manufacturing, assembly, and end-of-life, the processing and reuse of knowledge is crucial in engineering [5]. In this context, digital assistance systems that leverage AI algorithms for knowledge discovery in the database can enhance decision-making processes, provide predictive insights, and streamline design workflows [2]. The resulting symbiotic relationship between AI and digital assistants leads to more efficient, accurate, and innovative approaches to product development [1].

While the application of artificial intelligence machine learning in manufacturing and production engineering has shown great potential [6], such as specifically for the realization of self-repair capabilities of production systems [7] or the energy load forecasting of production systems [8], particularly, the application of artificial intelligence has proven to be effective in supporting various design tasks, such as, among others, the customer segmentation [9], the design automation of mechanical components [10], or the generation of synthetic part representatives in tolerancing [11]. Moreover, integrating artificial intelligence and advanced digital assistance systems in the maintenance industry constitutes a significant advancement in industrial operations [12]. In such settings, AI employs intricate algorithms to examine various data sources, including repair histories, detailed performance metrics, and real-time sensor data. This comprehensive data analysis enables the prompt detection of possible malfunctions and simplifies maintenance planning.

3. Data model and assistance system conceptualization

Motivated by the increasing possibilities of AI in engineering design and the potentials of exploiting repair data for product improvement, this paper proposes a novel framework for the knowledge retrieval from repair data. In doing so, the main research questions, which are answered in the paper, refer to the design of a suitable data model for repair data as well as the conceptualization of an AI-based design assistance system.

3.1. Data model for repair data

An optimized repair data model is essential for organizing and standardizing the unique data elements associated with repair processes. To ensure precise and consistent data collection, a stringent standardization process is rigorously followed. For each step of the repair process outlined in Fig. 1, technicians are required to manually collect data and enter it into an enterprise resource planning (ERP) system.

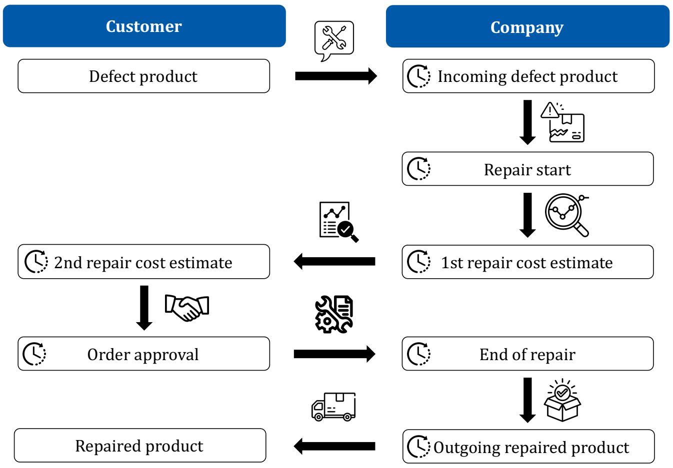

Figure 1 - Typical industrial repair process

This system records and archives the data in a SQL database. The database serves as the foundational layer for the data warehouse. Key data elements, also known as features in the realm of data science, that are collected and recorded include ServiceOrderID, ProductID, RepairTimestamp_{Step}, Customer-ID, TechnicianID, PartID, and PartQuantity.

A unique "ServiceOrderID" identifies each repair instance, with "ProductID" indicating the specific item undergoing repair. The repair workflow is recorded sequentially using "Repair-Timestamp_{Step}" entries, where "{Step}" denotes different stages in the repair process, as illustrated in Fig. 1. To keep track of customers and technicians, their associations are maintained through "CustomerID" and "TechnicianID," correspondingly. Meanwhile, the "PartID" and "PartQuantity" are used to monitor the used parts for repair.

3.2. Assistance system conceptualization and implementation

As a starting point for the conceptualization of repair data use in product development, the Reference Architectural Model for Industry 4.0 (RAMI 4.0) is utilized. This three-dimensional framework aligns operational assets and information technology (IT) components within a structured life cycle and layer model, simplifying the complexities of industrial digital transformation by providing a roadmap for leveraging i. a. repair data at different hierarchical levels [13].

Repair data is employed primarily within the “Instance” phase of the RAMI 4.0 lifecycle value stream, which focuses on addressing immediate technical issues at lower hierarchical tiers, including the product and field device levels. However, for product design, repair data is instrumental in informing the “Type” phase of the life cycle value stream framework. This integration is crucial as it significantly influences the enhancement processes at more advanced hierarchical echelons, notably at the enterprise and connected world levels. Unlike the confined scope of repair data within the repair process, its application in product design yields substantially more value but poses intricate challenges requiring detailed analysis and gradual implementation of improvements.

Our proposed assistance system aims to optimize and automate the use of repair data for product design by utilizing three primary components: data acquisition, data pre-processing, and data analysis. This approach eliminates complexities and is depicted in Fig. 2. The implementation of these components follows a bespoke adaptation of the Data Mining Methodology for Engineering Applications (DMME), which expands upon the principles of the Cross Industry Standard Process for Data Mining (CRISP-DM) [14]. The components are aligned with the stages of DMME, as outlined in Table 1, ensuring a coherent methodology and workflow.

Table 1 - Data Assistance System and DMME Components

Data Assistance Components	DMME Components
Data Acquisition	Business Understanding, Technical Understanding, Technical Realization
Data Preprocessing	Data Understanding, Data Preparation
Data Analysis	Modeling, Evaluation, Technical Implementation, Deployment

Data acquisition. The data acquisition component serves as the basis of the entire assistance system. The objective is to automatically extract two primary repair datasets from the data warehouse using a network management system. The initial dataset furnishes an outline of the repair process, detailing step-by-step timelines and information about defective products involved. The second dataset concentrates on specific defective parts, including their respective types, quantities, and defect characteristics.

Data preprocessing. In the development of our digital assistance system, a systematic approach is taken to the data pre-processing component to ensure the accuracy and usefulness of the repair data. This process consists of several sequential steps, as displayed in Fig. 2. The first phase involves meticulous data cleaning of the two extracted data sets to identify and remove errors and inconsistencies from repair data that are often found, mainly due to human input errors. Such inaccuracies may potentially lead to erroneous conclusions during product re-design. The approach to handling this in big datasets requires the complete elimination or interpolation of such data points to ensure a high quality of the dataset [15].

The next step is to merge the two datasets into one comprehensive dataset. The “ServiceOrderID” feature serves as the principal integration point. After merging, the data is arranged in chronological order, with recent outgoing repaired products timestamps (“RepairTimestep_Outgoing”). A further pre-processing step involves the creation of new features, specifically the calculation of the exact duration between repair process steps as shown in Fig. 1, and the inclusion of a “unit_of_work” feature, which aims to measure the time a technician spends on repair activities, excluding ancillary tasks. The final stage of our data pre-processing component is critical data filtering. This process removes redundant and non-critical components from the repair data set. The focus is on commonplace sundry supplies used frequently during repairs and exceedingly rare or specialized parts. These have minimal impact on product re-design and development decisions. The filtering process by selectively eliminating such data emphasizes more critical and influential repair parts. The output of this data preprocessing component is a fully pre-processed dataset, ready for further in-depth analysis.

Data analysis. A goal-driven development process is key to efficiently and successfully realizing our data analysis component. The implemented data analysis methods highlighted in Fig. 2 are classified chronologically into three distinct categories: descriptive, diagnostic, and predictive analyses, as listed in Table 2. A similar classification was proposed in the framework presented in [16] for designing and specifying data analytics projects.

Table 2 – Classification of data analysis methods

Data Analysis Method	Classification
Time Analysis	Descriptive
Component Analysis	Descriptive
Dependency Analysis	Diagnostic
Clustering	Diagnostic
Demand Forecasting	Predictive

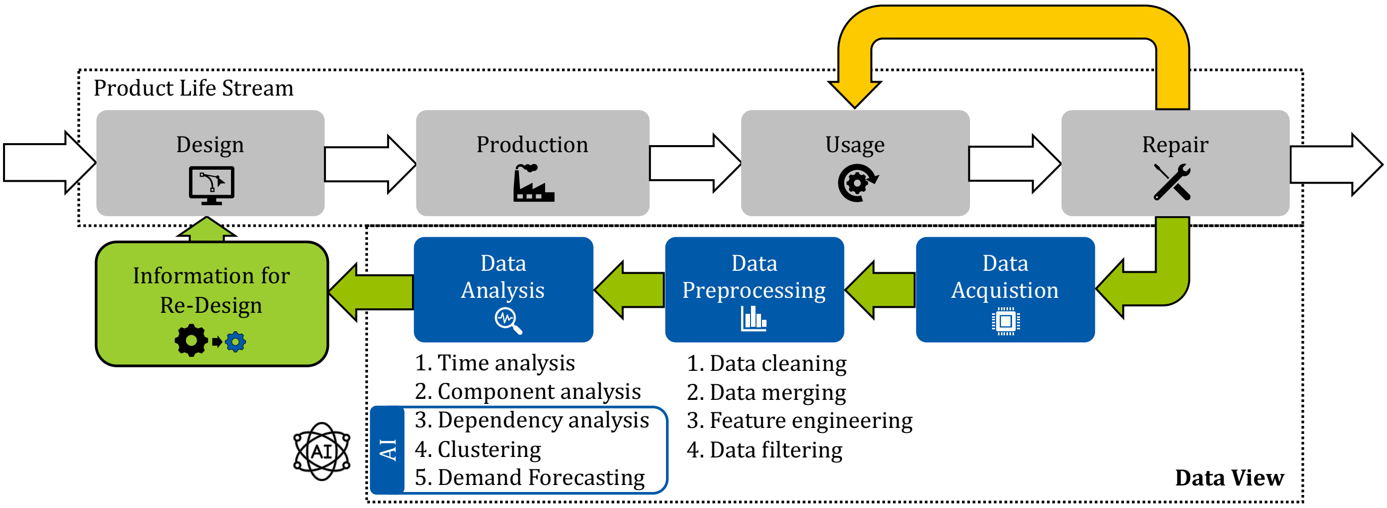

Figure 2 - Integration of the proposed AI-based assistance system (Data View) into the Product Life Stream

The initial step shared by all data analysis methods outlined in Table 2 is to enhance the interpretability and ease of reasoning for product re-design through a contextualization process [15]. This step involves filtering the pre-processed dataset and selectively analyzing data within a certain time frame, focusing on the "RepairTimestep_Outgoing" feature and targeting a specific product using the "ProductID" feature. The culmination of every data analysis method is the formation of a structured results dataset, which can be optionally converted into visual representations.

In line with these common steps, our data analysis procedure starts with a descriptive analysis. This phase encompasses time and component analyses to provide an initial, comprehensive view and preliminary insights into the repair dataset. The time analysis aims to understand the proficiency and effectiveness of the overall repair process. By calculating the average number of workdays per repair instance for each repair process step, using the “RepairTimestamps_{Step}” feature, potential areas for improvement, such as the speed and quality of executing these steps, can be pinpointed.

The component analysis aims to provide a clear insight of the repaired products and their defective parts. This is instrumental in understanding recurring trends and common issues associated with specific parts, as well as in evaluating the workload linked to each repaired product. To achieve this several key metrics are calculated as listed in Table 3.

Table 3 - Key metrics of the components analysis

Key Metrics	Feature
Service order quantity	“ServiceOrderID”
Work units	“unit_of_work”
Avg. work units per repair instance	“unit_of_work”, “SerivceOrderID”
Service order cycle	“ServiceOrderID”, “RepairTimestep_Outgoing”
Part quantity	“PartQuantity”, “PartID”
Avg. part quantity per repair instance	“ServiceOrderID”, “PartQuantity”, “PartID”

The next phase involves diagnostic analytics, which aims to uncover the underlying patterns that account for the events observed in the previous descriptive analytics phase. Unsupervised machine learning techniques, specifically K-means clustering and apriori for association rule mining, are used to delve deeper into the repair data. The main objective of association rule mining is to reveal the relationships and dependencies among various components in a given dataset. This analytical technique is particularly significant for pinpointing failure points that could lead to cascading effects in interlinked systems. Such insights are crucial during product re-design phases as they enable engineers to fortify critical dependencies or restructure components to minimize the likelihood of systemic failures. This study employs the apriori algorithm to analyze data. Firstly, the algorithm iterates through each "ServiceOrderID" for a given product to compile a list of the parts ("PartID") utilized in each order. This compilation then forms a list of lists, with each sub list representing the individual parts used in a specific service order. The subsequent step is to identify frequent item sets comprising components commonly appearing across service orders. To establish what qualifies as 'frequent,' a minimum support threshold is established. In this context, support refers to the proportion of service orders featuring a particular combination of parts. A higher support value indicates that the components are more commonly found in the dataset. Once these frequent item sets are identified, the algorithm generates association rules, which predict the probability of certain components appearing together. The strength of these rules is quantified through confidence, which evaluates the frequency with which a specific relationship holds across all service orders. For example, a high confidence value in a rule incorporating components X and Y suggests that the likelihood of component Y being defective is high when com-ponent X is defective. This guarantees that only the most statistically significant patterns are considered, providing invaluable insights into component relationships and trends and facilitating strategic decision-making for product re-design and maintenance.

The objective of utilizing clustering to categorize components by their criticality is a crucial element in optimizing repair processes and guiding product re-design. This classification enables the identification of parts more susceptible to failure, necessitating greater focus on maintenance and design improvement. We use the K-means algorithm and certain features in our dataset to achieve this classification: "PartQuantity" and the frequency of every part's appearance in service orders, as identified by the "ServiceOrderID." The final stage of our data analytics process incorporates predictive analytics, which is focused on forecasting potential faulty product components. We use historical repair data for features such as "PartQuantity" and "PartID" in this stage. This information is the basis for our predictive models, which accurately predict which product components will likely be faulty in the next month.

4. Assistance system application and validation

The AI-based assistance system described in section 3 has been implemented for the repair data analysis in the industrial case of plastic welding equipment and has significantly enhanced repair procedures and product re-design. The system used unsupervised machine learning techniques to analyze four years of historical repair data. The dataset consisted of 33,727 data points and 32 features after pre-processing. Table 4 describes one of the analyzed products, which could be improved by clustering its critical parts, as shown exemplary for the top 5 critical parts in Fig. 3.

Table 4 - Characteristics of analysed plastic welding product

Characteristic	Metric
Max. Welding capacity	1.8 kg/h, Ø 4 mm; 1.1 kg/h, Ø 3 mm
Welding material	PP / PE / PVDF
Filler metal	Round cord Ø 3/4 mm
Application range	Wall thicknesses 4 - 15 mm
Weight	5.5 kg
Length	470 mm
Electronic speed control	Yes
Heating tape	400 W
With power box	No
Preheating fan	2300 W

Particularly, it can be seen from Fig. 3 that the use of carbon brushes in electric motors highlights their importance as crucial components and their vulnerability to failure, which could be improved by re-design of future product generations. Moreover, their replacement establishes a significant association, leading to concurrent changes by replacing the deep groove ball bearing and its oil as a lubricant with a confidence level of 0.766. This highlights a critical interdependency identified through association rules, which improves the repair process and increases the mean time between repair.

Moreover, the assistance system's user-friendliness is reinforced by a graphical user interface (GUI) for the time and component analysis. This assists technicians and designers in efficiently identifying critical areas for repair and re-design.

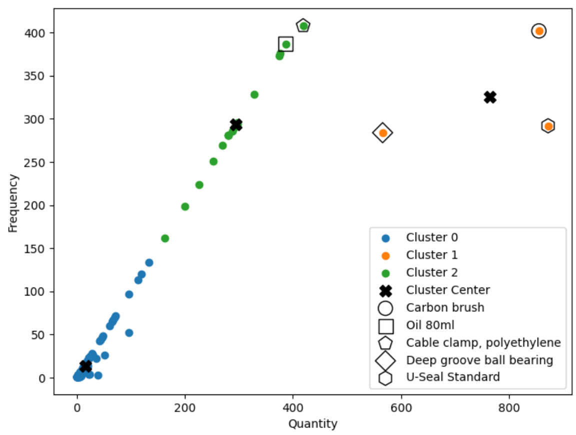

Figure 3 - Clustering of critical parts

5. Critical discussion and outlook

Analyzing repair data allows for accurate and efficient repair procedures and enhanced product re-design. The application of sophisticated machine learning methods such as clustering and association rule mining enables the detection of complex patterns and interdependencies, which is pivotal for informed and efficient product re-design. This method to anticipate potential malfunctions aids in the durability and dependability of products. Motivated by this, the paper describes a data model and novel AI-based assistance system for the repair data analysis as well as the successful application of this system to plastic welding equipment.

However, the approach carries some limitations. Reliance on extensive and high-quality data could hinder the system's effectiveness due to data gaps or biases. Additionally, integrating such a sophisticated AI system into existing workflows presents challenges like complexity and significant resource investment.

Future research should concentrate on improving the system's data collection and processing capabilities to minimize biases and improve predictive accuracy. Efforts should also be made to streamline the integration of this AI system into diverse operational environments, improving its usability and adoption. Furthermore, expanding the system's adaptability to include a wider range of product types and repair scenarios would greatly increase its usefulness and effectiveness.

Acknowledgements

This research was funded in the DiNaPro-project by the German Federal Ministry of Education and Research (BMBF) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used DeepL in order to support the proper language editing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Roucoules, L. & Anwer, N. (2021). Coevolution of digitalisation, organisations and Product Development Cycle. CIRP Annals, 70(2), 519-542. https://doi.org/10.1016/j.cirp.2021.05.003

[2] Stark, R., Brandenburg, E., & Lindow, K. (2021). Characterization and application of assistance systems in digital engineering. CIRP Annals, 70(1), 131-134. https://doi.org/10.1016/j.cirp.2021.04.061

[3] Arora, P., & Varshney, S. (2016). Analysis of k-means and k-medoids algorithm for big data. Procedia Computer Science, 78, 507-512. https://doi.org/10.1016/j.procs.2016.02.095

[4] Schmitt, R. H., Hüttemann, G., & Münker, S. (2021). A priori performance assement of line-less mobile assembly systems. CIRP Annals, 70(1), 389-392. https://doi.org/10.1016/j.cirp.2021.04.059

[5] Kügler, P., Dworschak, F., Schleich, B., & Wartzack, S. (2023). The evolution of knowledge-based engineering from a design research perspective: Literature review 2012-2021. Advanced Engineering Informatics, 55, 101892. https://doi.org/10.1016/j.aei.2023.101892

[6] Gao R X, Krüger J, Merklein M, Möhring H-C, Váncza J. (2024). Artificial Intelligence in manufacturing: State of the art, perspectives, and future directions. CIRP Annals 2024, 73:2, 723-749, https://doi.org/10.1016/j.cirp.2024.04.101

[7] Epureanu B I, Li X, Nassehi A, Koren Y. (2020). Self-repair of smart manufacturing systems by deep reinforcement learning. CIRP Annals 2020, 69:1, 421-424. https://doi.org/10.1016/j.cirp.2020.04.008

[8] Zink R, Ioshchikhes B, Weigold M. (2024). Concept drift monitoring for industrial load forecasting with artificial neural networks. Procedia CIRP 2024. 130, 120-125. https://doi.org/10.1016/j.procir.2024.10.065

[9] Hu, X., Liu, A., Li, X., Dai, Y., & Nakao, M. (2023). Explainable AI for customer segmentation in product development. CIRP Annals, 72(1), 89-92. https://doi.org/10.1016/j.cirp.2023.03.004

[10] Dworschak, F., Dietze, S., Wittmann, M., Schleich, B., & Wartzack, S. (2022). Reinforcement Learning for Engineering Design Automation. Advanced Engineering Informatics, 52, 101612. https://doi.org/10.1016/j.aei.2022.101612

[11] Schleich, B., Qie, Y., Wartzack, S., & Anwer, N. (2022). Generative adversarial networks for tolerance analysis. CIRP Annals, 71(1), 133-136. https://doi.org/10.1016/j.cirp.2022.03.021

[12] Gay, A., Voisin, A., Iung, B., Do, P., Bonidal, R., Khelassi, A. (2022). Data Augmentation-based Prognostics for Predictive Maintenance of Industrial System, CIRP Annals, Volume 71(1):409-412. https://doi.org/10.1016/j.cirp.2022.04.005

[13] Baptista, L. F., & Barata, J. (2021). Piloting Industry 4.0 in SMEs with RAMI 4.0: an enterprise architecture approach. Procedia Computer Science, 192, 2826-2835. https://doi.org/10.1016/j.procs.2021.09.053

[14] Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications–a holistic extension to the CRISP-DM model. Procedia CIRP, 79, 403-408. https://doi.org/10.1016/j.procir.2019.02.106

[15] Gao, R. X., Wang, L., Helu, M., & Teti, R. (2020). Big data analytics for smart factories of the future. CIRP Annals, 69(2), 668-692. https://doi.org/10.1016/j.cirp.2020.05.002

[16] Kühn A, Joppen R, Reinhart F, Röltgen D, von Enzberg S, Dumitrescu R. (2018). Analytics Canvas - A Framework for the Design and Specification of Data Analytics Projects. Procedia CIRP 2018, 70, 162-167. https://doi.org/10.1016/j.procir.2018.02.031