



## OPEN ACCESS

## EDITED BY

Fengqi You,  
Cornell University, United States

## REVIEWED BY

Zhengbing Yan,  
Wenzhou University, China  
José Ezequiel Santibañez-Aguilar,  
Tecnologico de Monterrey, Mexico  
Q. Peter He,  
Auburn University, United States

## \*CORRESPONDENCE

Helen H. Lou,  
✉ Helen.lou@lamar.edu

RECEIVED 05 January 2023

ACCEPTED 04 July 2023

PUBLISHED 12 July 2023

## CITATION

Cai T, Fang J, Daida S and Lou HH (2023),  
Review of synergy between machine  
learning and first principles models for  
asset integrity management.  
*Front. Chem. Eng.* 5:1138283.  
doi: 10.3389/fceng.2023.1138283

## COPYRIGHT

© 2023 Cai, Fang, Daida and Lou. This is  
an open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Review of synergy between machine learning and first principles models for asset integrity management

Tianxing Cai, Jian Fang, Sharath Daida and Helen H. Lou\*

Dan F. Smith Department of Chemical and Biomolecular Engineering, Lamar University, Beaumont, TX, United States

The chemical process industry (CPI) accumulated a rich data asset through industrial 4.0. There is a strong drive to develop and utilize effective approaches for process performance prediction and improvement, process control, sensor development, asset management, etc. The synergy between machine learning and first principles models can bring new insights and add tremendous value to the CPI. This paper reviews various applications of the synergies towards asset integrity management. An overview of some related commercial software packages are also provided.

## KEYWORDS

machine learning, first principles, preventive maintenance, digital twin, asset integrity management

## Highlights

- Motivation and recent resolutions for asset integrity management are discussed.
- Methodologies of first principles models, machine learning models, and hybrid models for asset integrity management are introduced and reviewed.
- Recent advances in software development and potential applications in the chemical industry are given.

## 1 Introduction to asset integrity management

The term asset integrity or asset integrity management (AIM) refers to the ability of an asset to perform accurately and effectively, as well as to ensure its safety and wellbeing (Chandima Ratnayake and Market, 2012). AIM applies to an asset's operation, from its design phase to its decommissioning and replacement (Oil & Gas IQ Editor, 2019).

Many critical industries, such as utilities, energy, and chemical process industries (CPI), are always searching solutions to achieve reliable operation, reduce maintenance costs, improve the performance of those assets and elongate their service life. The accelerant of information analysis will help the asset owners make better decisions, and gain a competitive advantage (Chandima Ratnayake and Market, 2012).

Traditional AIM processes, for instance, routine inspection and maintenance, are based on the fixed time interval's planning. This traditional approach ignores the asset's unique characteristics, condition, or inherent risks when using time-based methods, comparing with risk-based methods (Maguire and Fish, 2019). Capital resources may be wasted by owner-operators on unnecessary inspections (Maguire and Fish, 2019). However, many dramatic

**TABLE 1** Summary of the synergy between machine learning and first principles models.

Synergy	Explanation
Improved Prediction Accuracy	Machine learning models can be used to improve the prediction accuracy of first principles models. Vice versa is also true
Reduced Computational Costs	First principles models can be computationally expensive to run, especially for large systems or complex processes. Machine learning models can be used to reduce computational costs by serving as an efficient surrogate model, allowing for faster simulations of the same system or process
Handling Complex Systems	First principles models are limited by their ability to accurately capture all the complexities of a given system or process. Machine learning models can be used to handle the complexities that are difficult to capture using first principles models, such as non-linear behavior, non-stationary systems, and interactions between different parts of the system
Data-Driven Model Development	Machine learning models can be used to develop first principles models by learning patterns from data. This can help identify key features and relationships that can be incorporated into a first principles model
Improved Understanding of System Behavior	Machine learning models can be used to identify relationships and patterns in data that are not apparent from a first principles model alone. By combining machine learning models with first principles models, it is possible to gain a deeper understanding of the behavior of a system
Robustness and Flexibility	Combining machine learning models with first principles models can improve the robustness and flexibility of the overall model. By incorporating machine learning models, it is possible to handle cases where the first principles model may not be sufficient or where there the model parameters are highly uncertain

cases that the management of asset integrity failed, such as Deepwater Horizon, 2010 (Klimas, 2010) and Chevron Refinery Pipe Rupture and Fire, 2012 (Hodgson, 2013).

Moreover, the market still lacks out-of-the-box catch-all solutions, even though there are more AIM systems than ever before (Kusumawardhani et al, 2016; Bokrantz, Skoogh, Berlin, Wuest, & Stahre, 2020; Beneroso & Robinson, 2022). No inspection plan or database can handle all the issues that may arise in AIM. In addition to the lack of resources, suppliers, and training support (Kusumawardhani et al, 2016; Bokrantz, Skoogh, Berlin, Wuest, & Stahre, 2020), there are technical challenges like aging equipment, corrosion, degradation, and cracking (Zha, Lan, Lin, & Meng, 2023), human error challenges (Elidolu, Akyuz, Arslan, & Arslanoğlu, 2022), and engineering design challenges (Beneroso & Robinson, 2022). In the current process industries, material failures, degradation, and corrosion pose a significant challenge. Wood et al (Wood et al, 2013) reported that approximately 20% of significant refinery accidents since 1984 were attributed to the corrosion-related accidents. In the oil and gas industry, much of the infrastructure is approaching or aging well beyond its operational life span. For many facilities, AIM is widely used as a strategic plan due to the prohibitively high cost of replacing assets and the subsequent turnaround time.

New approaches, including hybrid AIM approaches, predictive maintenance, and even prescription maintenance has gained wide adoption in the industry as a result of the rapid development of data analytics and data visualization technologies. Using digital twins, asset owners are able to improve the design, production, and management of their assets. First principle models and machine learning models can be complementary in developing digital twins and hybrid AIM approaches. In the following sections, the comparison of different types of models will be presented.

## 2 First principles models vs. machine learning models

Imagination, recognizing patterns, creation, memorization and understanding, making choices, adapting to change, and learning

from experience are all aspects of intelligence (Oksanen, 2018). Artificial intelligence aims to make computers behave like humans in a more human manner in a shorter time than a human and in a more efficient manner in comparison to human behavior (Poole et al, 1998).

To develop first principles (FP) models, a fundamental understanding of physico-chemical phenomena is essential. These include symmetries, conservation laws (mass, momentum, energy, etc.), thermodynamics, asymptotic behavior, rate equations, reaction kinetics, etc. Mathematical formulations, such as algebraic equations, ordinary differential equations, and partial differential equations, can provide valuable insights into the system's operating behavior. In most cases, FP can simulate multiple unit operations with various constraints, such as monotonicity, convexity, and non-negativity of variables (Oil & Gas IQ Editor, 2019; Rajulapati et al, 2022).

The correlations used in science and engineering, including chemical engineering, are explicit and well-known. Moreover, they can also be utilized to connect multiple unit operations in a process flowsheet using mass and heat balance, which are based on direct relationships within a chemical process (Werner, 2021). The FP models, however, are dependent on many factors, including the properties of chemicals and mixtures, chemical kinetics, thermodynamics, and even equipment manufacturers, catalyst suppliers, process licensees, and chemical operations. While some of the necessary data can be found in readily available chemical engineering literature, more specific data must be obtained through experiments designed specifically for the described process. Additionally, accessing the information needed is still challenging. There can be significant inaccuracies in the final result due to unknowns or uncertainties, which makes it difficult to capture the full impact of all factors, both online and in real time (Werner, 2021).

As a subset of artificial intelligence (AI), the term "machine learning (ML)" refers to the process of creating models from data (Haghighatdari and Hachmann, 2019) instead of FP models. The use of ML is a powerful tool for generating hidden or complex correlations, patterns, insights, rules, and guidance from data sets

in a mathematically coherent manner. An ML algorithm builds a mathematical model on the basis of sample data, referred to as training data (Bishop, 2006; Khanzode and Sarode, 2020). ML methods have been applied to various modeling applications, including a wide range of materials and endpoints (Winkler, 2020).

ML methods can be classified as supervised, semi-supervised, unsupervised, and reinforcement learning (RL) methods (Sarker, 2021). A key difference between the supervised, semi-supervised, unsupervised, and reinforcement learning methods is the amount of information (i.e., labeling, context) that can be obtained for the target variable for training an ML algorithm (Haghighatlari and Hachmann, 2019). Traditionally, most chemical engineering related research has relied on the supervised learning approach, even though it may result in major uncertainties and prediction errors due to simplistic and unrealistic assumptions (Haghighatlari and Hachmann, 2019).

Each of these machine learning approaches has its unique characteristics and applications, and the pick of a particular approach depends on the nature of the problem and the available data.

Supervised learning is a machine learning technique where the model is trained on labeled data (Caruana & Niculescu-Mizil, 2006). Labeled data consists of input features (also called independent variables) and their corresponding output labels (also called dependent variables or targets). The goal of supervised learning is to learn a mapping function that can predict the correct output label for new, unseen inputs (Liu & Liu, 2011). The model learns from the labeled examples provided during training and generalizes its knowledge to make predictions on new, unlabeled data. Common algorithms used in supervised learning include decision trees, support vector machines (SVM), and neural networks (Liu & Liu, 2011; Zhou, 2018; Sommers, Menkovski, & Fahland, 2023).

Semi-supervised learning is a combination of supervised and unsupervised learning. In this approach, the model is trained on a dataset that contains both labeled and unlabeled examples (Hady & Schwenker, 2013). The labeled examples are used to guide the learning process, similar to supervised learning, while the unlabeled examples help the model to capture additional patterns or structures in the data. The objective is to leverage the unlabeled data to improve the model's performance or generalization ability. Semi-supervised learning can be useful when obtaining labeled data is expensive or time-consuming. Algorithms such as self-training, co-training, and generative models like autoencoders are commonly used in semi-supervised learning (Reddy, Viswanath, & Reddy, 2018; Zhao, Zhou, Duan, Wang, Qi, & Shi, 2022).

Unsupervised learning is a type of machine learning where the model learns from unlabeled data, without any predefined output labels (Berg-Kirkpatrick, Bouchard-Côté, DeNero, & Klein, 2010). The goal is to discover patterns, structures, or relationships within the data. Unlike supervised learning, the model doesn't receive explicit feedback on the correctness of its predictions. Unsupervised learning algorithms aim to find hidden patterns or clusters in the data, dimensionality reduction, or discover the underlying probability distribution of the data. Common techniques used in unsupervised learning include clustering algorithms (e.g., k-means, hierarchical clustering), dimensionality reduction (e.g., principal component analysis), and generative

models (e.g., Gaussian mixture models) (Dike, Zhou, Deveerasetty, & Wu, 2018; Cabannes, Bietti, & Balestrero, 2023).

Reinforcement learning is a machine learning paradigm that focuses on learning how to make decisions or take actions in an environment to maximize a cumulative reward (Kaelbling, Littman, & Moore, 1996). It is based on the idea of an agent interacting with an environment, receiving feedback in the form of rewards or penalties for its actions. The agent's objective is to learn a policy that guides its decision-making process to achieve the highest possible reward over time. Reinforcement learning involves a trial-and-error learning approach, where the agent explores the environment, takes actions, observes the outcomes, and adjusts its policy accordingly. Algorithms like Q-learning and deep reinforcement learning (using neural networks) are often used in reinforcement learning (Barto & Sutton, 1995; Brunke, Greeff, Hall, Yuan, Zhou, Panerati, & Schoellig, 2022; Ladosz, Weng, Kim, & Oh, 2022).

These four learning approaches continue to be actively researched and applied in various domains due to their effectiveness and potential for solving complex problems. A brief comparison of their applications are provided below.

Supervised learning remains in tasks such as image classification, speech recognition, natural language processing, and many other applications. As more labeled data becomes available and computing power increases, supervised learning techniques can be further refined to improve accuracy and performance (Aminian, Abroshan, Khalili, Toni, & Rodrigues, 2022; Zhou, Liu, Pourpanah, Zeng, & Wang, 2022; Iman, Arabia, & Rasheed, 2023).

Due to the more availability of large amounts of unlabeled data in many domains, semi-supervised Learning offers the advantage of leveraging both labeled and unlabeled data, making it attractive when obtaining labeled data is expensive or time-consuming (Yang, Song, King, & Xu, 2022). By effectively utilizing unlabeled data, semi-supervised learning has the potential to improve the performance of models, especially when labeled data is limited (Tang & Jia, 2022).

Unsupervised learning is recognized by researchers and practitioners because of its potential for discovering hidden patterns and structures in large, unlabeled datasets (Liu, Yoo, Xing, Oh, El Fakhri, Kang, & Woo, 2022). With the increasing availability of massive amounts of unlabeled data, unsupervised learning techniques can help in tasks such as clustering, anomaly detection, and data exploration. Unsupervised learning is also valuable in pre-training deep learning models, which can then be fine-tuned using supervised learning techniques (Wang & Biljecki, 2022).

Reinforcement learning allows agents to learn optimal behaviors through interactions with their environments, particularly in the fields of autonomous systems (Moos, Hansel, Abdulsamad, Stark, Clever, & Peters, 2022) since the agents in RL are trained to take the optimal actions in order to maximize a cumulative reward signal. Recent advancements in deep reinforcement learning, where deep neural networks are combined with reinforcement learning algorithms, have led to impressive achievements in complex tasks such as the control of engineering systems, and resource allocation optimization (Zhang, Seal, Wu, Bouffard, & Boulet, 2022).

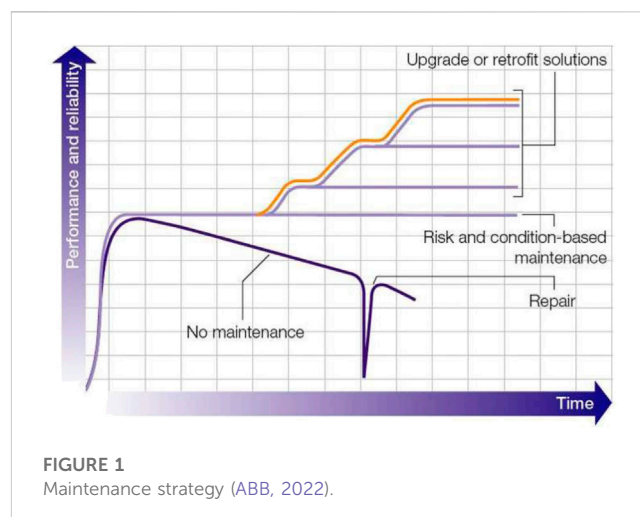
Researchers and practitioners are exploring ways to combine different learning paradigms, develop more efficient algorithms, and address challenges such as data scarcity, scalability, and interpretability. Hybrid models that combine the strengths of ML models with first principles model can provide solutions to overcome the disadvantages of a stand-alone approach and provide a wide range of advantages. In data-driven ML algorithms, available data is utilized to the fullest, while models based on first principles can be used to verify whether the data is reasonable and explore the unknown operating ranges. Through this synergy, AIM predictions become as accurate as the prediction model employed and high prediction accuracy can be achieved.

Overall, the combination of machine learning and first principles models can provide a powerful approach for understanding and predicting complex systems and processes. The following Table 1 summarizes the synergy between machine learning and first principle models:

### 3 Application of hybrid approach for AIM

Compare et al (Compare et al, 2020) studied oil and gas industry gas turbine maintenance. They developed a sequential decision-making approach involving maintenance part flow management and proposed RL as a solution. A case study derived from real industrial applications was used to test the RL method, with the hybrid approach of machine learning and first principle model, and it proved that the hybrid approach was more effective than a rule-based experience. In the context of gas turbine maintenance, RL can be used to optimize part flow management by determining the most efficient way to move parts through the maintenance process. To apply RL to part flow management, a first principle model of the gas turbine maintenance process can be developed. This model should capture the physical and operational constraints of the system, such as the time required for each maintenance step, the availability of resources, and the flow of parts through the system. Once the model is developed, an RL agent can be trained to learn the optimal sequence of actions to take in order to minimize maintenance time and maximize resource utilization. The agent should be able to learn from experience by interacting with the model and receiving feedback in the form of a reward signal that reflects the efficiency of the maintenance process. The agent can be trained using simulation data generated by the first principle model, and its performance can be evaluated by testing it on new, unseen scenarios. The use of RL in optimizing part flow management for gas turbine maintenance can lead to significant improvements in efficiency and cost-effectiveness, by identifying the most efficient way to move parts through the maintenance process while minimizing resource utilization.

Ratnayake (2012) proposed an analytical hierarchy process (AHP) approach as a method for identifying how personnel in an asset-intensive organization can make decisions on AIM-related factors related to “triple bottom line” demands, translating these factors into numerical values based on their experience, intuitions, and intentions. The sustainable performance of AIM can be assessed using a hierarchical model based on the hierarchy of AIM-related factors (elements). It is important to recognize that AHP has significant effects on mitigating the occurrence of catastrophic events.



However, the AIM-related challenges can also involve interactions and feedback dependencies between factors (elements) belonging to different levels of a hierarchy as well as those affiliated at the same level.

Iqbal et al (2021) developed a risk-based assessment approach that focuses on the logical connections between AIM components and their contributions to safety culture maturity (SCM). These attributes are integrated into the processes of SCM and AIM assessment. The evaluation of 88 oil and gas pipeline companies operating in British Columbia was assessed using the risk-based assessment application, which analyzed the overall performance of the companies measured by AIM and SCM. Comparison among the performance of small, medium, and large companies showed that large companies with more resources (financial and technical) performed better on AIMs than small companies. In addition, the large companies tended to focus more on the technical aspects of AIMs. However, it was evident that most companies failed to manage their AIM effectively, which is manifested by low levels of internal audits, competency training, management reviews, and non-conformance controls.

### 4 Predictive/Prescription maintenance

The modern chemical industry is transforming from reactive trouble-shooting to proactive and predictive maintenance in order to enhance operational availability and efficiency of industrial asset and infrastructure platforms, which is illustrated in Figure 1 (ABB, 2022). The chemical plants go through different stages (Corrective Maintenance, Preventive Maintenance, Risk-based Maintenance, Condition-based Maintenance) of maintenance strategy with the industrial and technology revolution. In the initial stage of Corrective Maintenance, once the abnormal situation is detected or the malfunction is identified, the maintenance job will be conducted in order to return to normal operation. This strategy may be effective, but it cannot help the plant to prevent or reduce the occurrence of catastrophic failures. For Preventive Maintenance, the plant will follow the time intervals defined by the requirements, such as industry standards and customer requirements to have regular maintenance activities so that the risk of performance failure can be



reduced, and the equipment degradation can be minimized. With the help of data analytics and sensor technology, chemical plants can move to the next stage of Risk-based Maintenance. In this approach, routine preventive maintenance can rely on sensor measurement, regular testing, and diagnostic analysis to transform the collected information of the operation and process condition into asset health and risk assessment suggestions to optimize the maintenance program. In addition, if the chemical plants want to achieve reliability, safety, and efficiency in the long term, they need to upgrade their maintenance strategy to Condition-based Maintenance so that the online monitoring of critical equipment and process parameters will address the indicators or signals of warning for advanced asset management.

Equipment and critical facilities are the backbones of the energy and chemical industry. It has become essential to offer suitable models for proactive and predictive maintenance. This research field has attracted the attention and effort of many research groups in the world. Big Data Analytics (BDA) is becoming increasingly important in the chemical industry, as it enables the discovery of insights from large volumes of data that can be used to optimize manufacturing processes, reduce costs, and improve product quality. One approach to implementing BDA in the chemical industry is to use cloud manufacturing, which involves the integration of cloud computing technologies with manufacturing processes. [Sebbar et al \(2022\)](#) studied a major African Phosphates Company and proposed a BDA-enabler architecture based on cloud manufacturing to identify digital opportunities and key benefits regarding performance management, production control, and maintenance. To develop a BDA-enabler architecture based on cloud manufacturing for the chemical industry, a first principle model can be developed to capture the physical and operational characteristics of the chemical manufacturing process. This model should include variables such as temperature, pressure, flow rates, and chemical composition, and should be capable of predicting the behavior of the system under different operating conditions. Once the first principle model is developed, machine learning techniques can be used to develop data-driven models that can be used to optimize the manufacturing process. For example, supervised learning techniques such as regression or classification can be used to develop models that predict the quality of the final product based on input variables such as temperature and pressure. Unsupervised learning techniques such as clustering or anomaly detection can also be used to identify patterns in the data that may be indicative of process inefficiencies or quality issues. Reinforcement learning can be used to optimize the manufacturing process by identifying the best sequence of actions to take to achieve a desired outcome, such as minimizing waste or maximizing yield. The BDA-enabler architecture should be designed to integrate these machine learning models with the first principle model. This will enable real-time monitoring of the manufacturing process and the generation of insights that can be used to optimize the process and improve product quality.

In the rubber industry, proactive maintenance can be critical to ensuring the reliability of equipment and preventing unexpected downtime. One approach to implementing proactive maintenance is to use RL algorithms to develop predictive models that can be used to identify potential equipment failures before they occur. A research group in India, [Senthil and Pandian \(2022\)](#), conducted an investigation into the enhancement of availability of a curing

machine deployed in the rubber company in India. Critical component identification in curing machines is necessary to prevent rapid failure followed by subsequent repairs that extend curing machine downtime. They adopted RL algorithm to prevent frequent downtime by improving the availability of the curing machine at the time when unscheduled long-term maintenance would interfere with operation, due to the occurrence of unspecified failure of critical components. The agent can then use this knowledge to predict when equipment failures are likely to occur and recommend maintenance actions that can be taken to prevent them. The RL agent can be trained using different algorithms and methods, depending on the complexity of the model and the desired performance. The agent's performance can be evaluated by testing it on new, unseen scenarios and comparing its predictions to actual equipment failures. The use of RL algorithms to develop proactive maintenance models in the rubber industry can lead to significant improvements in equipment reliability and production efficiency. By identifying potential equipment failures before they occur and recommending maintenance actions that can prevent them, the proactive maintenance model can help reduce downtime, increase throughput, and improve product quality.

In addition, the software packages, such as AspenTech's Asset Performance Management Suite (including Aspen Fidelis, Aspen Mtell, Aspen ProMV, Aspen Process Pulse, and Aspen Unscrambler) ([Aspen Technology, 2022b](#)), Emerson's DeltaV Preventive Maintenance ([EMERSON, 2022a](#)) and AVEVA's PRISM Predictive Asset Analytics ([AVEVA, 2022](#)), have been created to enhance the capability and capacity of preventive maintenance and risk-based inspection by the application of hybrid models for asset integrity management in the recent years.

AspenTech, Emerson and AVEVA all use first principles based process simulation software packages for industrial application. Given a process design and an appropriate selection of thermodynamic models, these software packages use mathematical models of material balance and energy balance to handle very complex chemical reaction and unit operation processes. Aspen Technology has developed Aspen Fidelis to enable quantitative risk assessment and derive data-driven decisions that line up plant performance with operational and economic goals ([Aspen Technology, 2022a](#)). Machine learning technology is applied to predict equipment failures weeks in advance so that the user knows how and when to perform the repair to maximize productivity and minimize risk. It also enables plant managers to identify the optimum approach for improving decision agility and throughput at the lowest cost. The approach behind Fidelis is a risk management/decision making tool to help engineers and decision makers figuring out how many spare parts might be needed, and enables plant managers to identify the optimum approach for improving throughput at the lowest cost. The applications of Aspen Fidelis Reliability Software extend over the plant life cycle including the phases of plant design, plant operation, and maintenance. Emerson, AVEVA and AspenTech all have similar solutions (Aspen Fidelis, Aspen Mtell, Aspen ProMV, Aspen Process Pulse, Aspen Unscrambler, DeltaV Preventive Maintenance, and AVEVA's PRISM Predictive Asset Analytics) on preventive maintenance and asset management ([EMERSON, 2022a](#); [AVEVA, 2022](#); [Aspen Technology, 2022b](#)). For example, Emerson's DeltaV Preventive Maintenance integrate

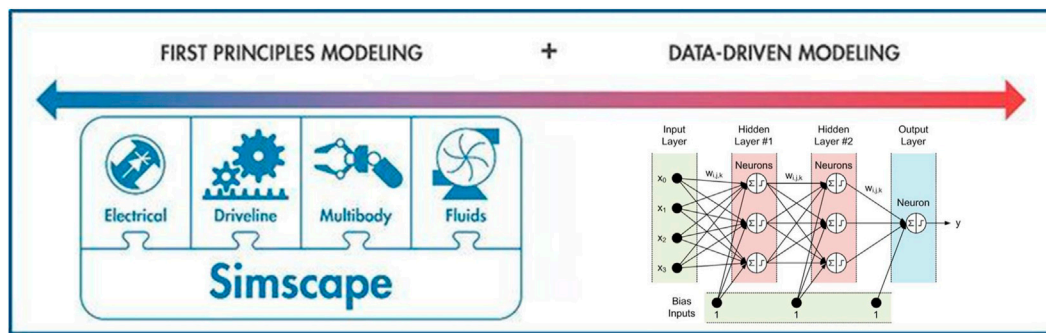


FIGURE 2  
Methods for digital twins: first principles model and data-driven model (Valentine and Hahn, 2021; Mathworks, 2022).

multiple modules including System Updates and Hotfix, Controllers, Cabinet, Workstations, I/O Subsystems, Network, DeltaV SIS Maintenance, Backup and Recovery, Automated Patch Management, and Virtualization (EMERSON, 2022b).

## 5 Digital twin

Digital twin technology is a powerful tool for asset integrity management, as it allows for the creation of a virtual replica of a physical asset. Machine learning and first principles models can be used in synergy with digital twin technology to improve asset integrity management.

### 5.1 Fundamentals of digital twin

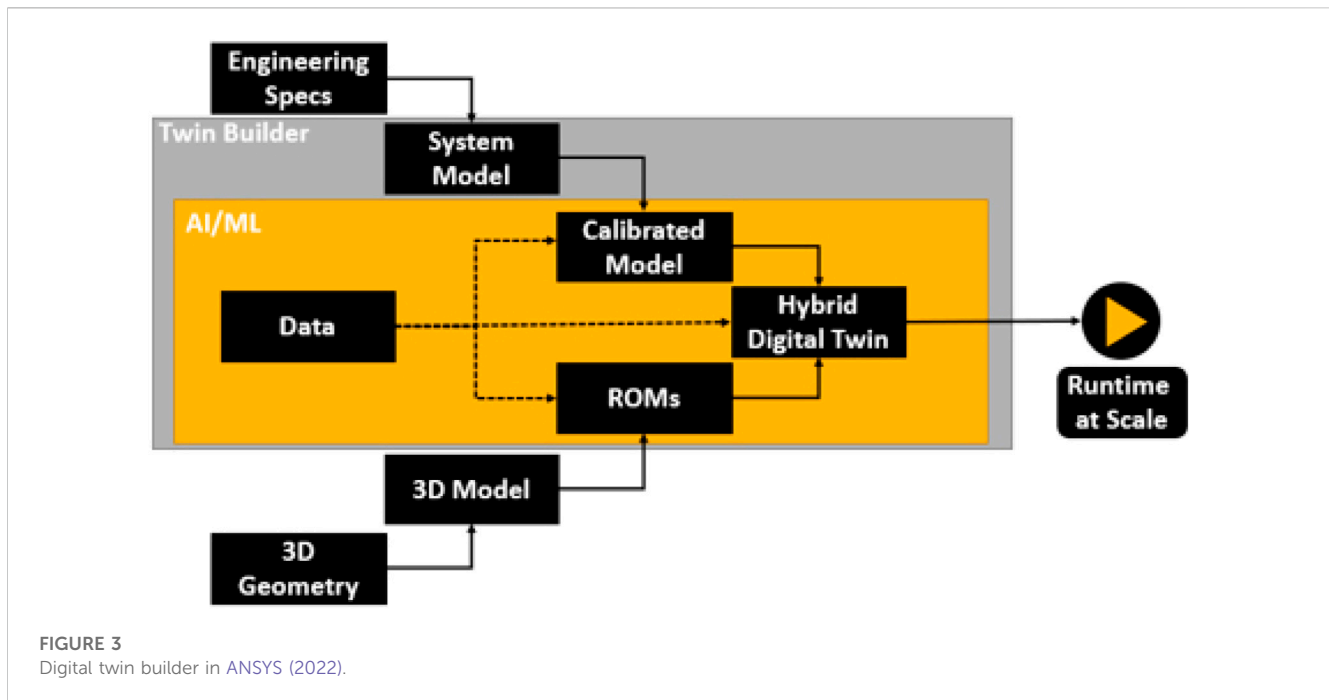
A digital twin (Zhou et al, 2014; Gao et al, 2022) allows the diagnosis and prediction of the physical twin by relying on a combination of physics-based (physics from first principles) models and data-driven analytics (Ma et al, 2022; Tao et al, 2022). The digital twin is built upon the integration of hardware (the sensor, measurement technologies, industrial Internet of Things), and software (simulation and modeling, and machine learning). Using simulations, and other decision-making technologies, digital twins can help predict the physical system's performance under the impact of abnormal situations and help engineers to come up with a good asset management strategy.

### 5.2 Types of digital twin

Physics-based models or First Principle Models are derived from the first principles of physics (material balance and energy balance). Due to the limited knowledge, a single process model or even comprehensive model system may only be able to simulate part of the overall plant operation. Multiple models may be needed to run in parallel to reflect the various aspects of a physical process. The first principles models can also take inputs from sensor measurement and sensory to make the models adaptable to the dynamic environment.

There are different types of digital twins focusing on different aspects of the engineering system. For example, MATLAB can mimic the industrial control systems. ANSYS Digital Twin cannot mimic the control systems, but it can build 2-D or 3-D geometry. Amazon Digital Twin (AWS IoT TwinMaker) cannot create 3-D models but it can provide the platform to allow the users to combine their own existing 3D model with real world data to create the digital twin model (Viola and Chen, 2020; Amazon, 2022; Beer and Hahne, 2022).

With MATLAB, a model can be defined by using data from the connected asset. Matlab can be used to create a physics-based model by using multi-domain modeling tools. Figure 2 shows that the understanding of the system's physics can be used to derive a mathematical representation in MATLAB through Simscape, which has the function modules of Fluids, Multibody, Driveline and Electronics, and Power System. The online first principles simulation models are the fundamentals to develop simulation-based digital twins (Valentine and Hahn, 2021). They can be used for numerous important applications with many advantages in the industrial adoption of first principles simulation models. For example, a PID or PLC controller model can be built and implemented in the mathematical package of MATLAB. The structure and timing diagrams can be used to describe its operation. The model may be valuable to decision makers or multiple stakeholders working in the field of industrial control systems (Brylina et al, 2020). Then, the system test data can be used to generate data-driven modeling. In MATLAB, the available tools for first principles modeling are Simulink, Simscape, Partial Differential Equation, FEM, and Symbolic Math while the available tools for data driven modeling include System Identification, Machine Learning, and Test & Measurement Tools (Valentine and Hahn, 2021; Mathworks, 2022). Both data-driven and physics-based models can be tuned with data from the operating asset. These digital twins can be utilized for prediction, what-if analysis, abnormal situation detection and management, fault identification, and more. Data-driven methods available with MATLAB include Machine Learning (ML), Deep Learning (DL), Neural Networks (NN), and System Identification (SI). A set of data can be used to train or extract a model with the rest of data to validate the models. Digital twins and Artificial Intelligence are



connected to optimize the yield increase and enhance PID control with the adoption of hybrid control strategy (Andrade et al, 2021).

ANSYS is a multi-dimensional engineering simulation software which can be used to create digital twins. Digital twins from ANSYS can help to schedule predictive maintenance and optimize the asset performance. ANSYS has the module of Hybrid Analytics which can combine machine learning (ML)-technique with a physics-based methodology to achieve the desired accuracy, as shown in Figure 3, this new version makes the design of many engineering systems easier (ANSYS, 2022). A machine learning model can be merged with physical modeling from 3D Geometry and 3D Model with the guidance of Engineering Specs and System Model to generate gray box models (Calibrated Model and ROMs) and improve the quality of real time simulation.

### 5.3 Digital twin for asset integrity management

Machine learning and first principles models can be used in synergy with digital twin technology for asset integrity management. Some benefits are listed below.

- Improved predictive maintenance: By combining machine learning models with digital twin technology, it is possible to predict the probability of failure and the remaining useful life of an asset. This can help in scheduling maintenance activities before a failure occurs, reducing downtime and increasing asset availability (You, Chen, Hu, Liu, & Ji, 2022).
- Reduced downtime and costs: First principles models can be used in combination with digital twin technology to optimize maintenance schedules and reduce downtime. By simulating different maintenance scenarios and predicting their outcomes, it is possible to identify the optimal maintenance

schedule for an asset, reducing downtime and costs (Wright & Davidson, 2020).

- Improved anomaly detection: Machine learning models can be used to detect anomalies in asset data, such as vibration or temperature readings, that may indicate the potential for failure. By combining machine learning with digital twin technology, it is possible to simulate the behavior of an asset under different conditions, identifying potential anomalies before they become a problem (Trauer, Pflingstl, Finsterer, & Zimmermann, 2021).
- Improved design and optimization: First principles models can be used in combination with digital twin technology to optimize the design of an asset and its components. By simulating different designs and predicting their performance, it is possible to identify the optimal design for an asset, reducing costs and improving its overall performance (Lim, Zheng, Chen, & Huang, 2020).
- Improved risk assessment: Machine learning models can be used in combination with digital twin technology to predict the likelihood of a failure occurring and its potential consequences. This can help in identifying high-risk assets and prioritizing maintenance activities accordingly (Melesse, Di Pasquale, & Riemma, 2021).

Overall, the synergy between machine learning and first principles models with digital twin technology can provide a powerful approach for asset integrity management, improving the accuracy of predictions and reducing downtime and costs.

## 6 Conclusion

This paper reviewed the recent contributions towards various applications that merges the physics-based and data driven

knowledge for asset integrity management. The efficient integration between machine learning and the first principles models is essential to achieve asset health and sustainability. In this paper, the method and application of these hybridization techniques are reviewed. Synergies between machine learning and the first principle models in the software platform (AspenTech's Asset Performance Management Suite, Emerson's DeltaV Preventive Maintenance, AVEVA's PRISM Predictive Asset Analytics, MathWorks's MATLAB, ANSYS' Twin Builder) is also reviewed and compared. All these efforts led to the continuous improvement of preventive maintenance for the asset integrity management. Anifowose et al, 2012; Matsuo et al, 2022; Sattari et al, 2022.

## Author contributions

HL conceived the main conceptual ideas and devised the framework of this manuscript with the proof of outline. Sections of Highlight, Abstract and Conclusion were also written by HL. Sections of Introduction to Asset Integrity Management, and First Principles Models vs. Machine Learning Models, and Hybrid

Approach for AIM were written by JF. Sections of Predictive/ Prescription Maintenance and Digital Twin were written by TC. List of References was prepared by SD. All authors contributed to the article and approved the submitted version.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- ABB (2022). 4 types of maintenance strategy, which one to choose? ABB. Available at: <https://new.abb.com/medium-voltage/service/maintenance/feature-articles/4-types-of-maintenance-strategy-which-one-to-choose>.
- Amazon (2022). Amazon IoT TwinMaker Amazon. Available at: <https://aws.amazon.com/iot-twinmaker/>.
- Aminian, G., Abroshan, M., Khalili, M. M., Toni, L., and Rodrigues, M. (2022). "An information-theoretical approach to semi-supervised learning under covariate-shift," in International Conference on Artificial Intelligence and Statistics, Italy, May 2022 (PMLR), 7433–7449.
- Andrade, M. A. N. D., Lepikson, H. A., and Machado, C. A. T. (2021). A model proposal for digital twin development: Essential oil extraction perspective. *Res. Sq.* doi:10.21203/rs.3.rs-300217/v1
- Anifowose, B., Lawler, D. M., van der Horst, D., and Chapman, L. (2012). Attacks on oil transport pipelines in Nigeria: A quantitative exploration and possible explanation of observed patterns. *Appl. Geogr.* 32, 636–651. doi:10.1016/j.apgeog.2011.07.012
- ANSYS (2022). Digital twins ANSYS. Available at: <https://www.ansys.com/products/digital-twin>.
- Aspen Technology, I. (2022a). Aspen Fidelis reliability software aspen technol. Inc. Available at: <https://www.aspentech.com/en> (Accessed December 11, 2022).
- Aspen Technology, I. (2022b). Asset performance management aspen technol. Inc. Available at: <https://www.aspentech.com/en/solutions/asset-performance-management> (Accessed October 11, 2022).
- AVEVA (2022). AVEVA predictive analytics. Available at: <https://www.aveva.com/en/products/predictive-analytics/> (Accessed November 11, 2022).
- Barto, A. G., and Sutton, R. S. (1995). "Reinforcement learning," in *Handbook of brain theory and neural networks* (Massachusetts, United States: MIT Press), 804–809.
- Beer, J. D., and Hahne, M. (2022). Connecting the simulation model to the digital twin to help drive sustainability. *Comput. Aided Chem. Eng.* 49, 991–996. doi:10.1016/B978-0-323-85159-6.50165-2
- Beneroso, D., and Robinson, J. (2022). Online project-based learning in engineering design: Supporting the acquisition of design skills. *Educ. Chem. Eng.* 38, 38–47. doi:10.1016/j.ece.2021.09.002
- Berg-Kirkpatrick, T., Bouchard-Côté, A., DeNero, J., and Klein, D. (2010). "Painless unsupervised learning with features," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles California, June 2010, 582–590.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer.
- Bokrantz, J., Skoogh, A., Berlin, C., Wuest, T., and Stahre, J. (2020). Smart maintenance: A research agenda for industrial maintenance management. *Int. J. Prod. Econ.* 224, 107547. doi:10.1016/j.ijpe.2019.107547
- Brunke, L., Greeff, M., Hall, A. W., Yuan, Z., Zhou, S., Panerati, J., et al. (2022). Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annu. Rev. Control, Robotics, Aut. Syst.* 5, 411–444. doi:10.1146/annurev-control-042920-020211
- Brylina, O. G., Kuzmina, N. N., and Osintsev, K. V. (2020). "Modeling as the foundation of digital twins," in Proc. - 2020 Glob. Smart Ind. Conf. GloSIC 2020, Chelyabinsk, Russia, 17–19 November 2020, 276–280. doi:10.1109/GloSIC50886.2020.9267812
- Cabannes, V., Bietti, A., and Balestrieri, R. (2023). "June). On minimal variations for unsupervised representation learning," in InICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, June 4 to June 10, 2023 (IEEE), 1–5.
- Caruana, R., and Niculescu-Mizil, A. (2006). "An empirical comparison of supervised learning algorithms," in Proceedings of the 23rd international conference on Machine learning, Pittsburgh, Pennsylvania, June 2006, 161–168.
- Chandima Ratnayake, R. M., and Markeset, T. (2012). Asset integrity management for sustainable industrial operations: Measuring the performance. *Int. J. Sustain. Eng.* 5, 145–158. doi:10.1080/19397038.2011.581391
- Compare, M., Bellani, L., Cobelli, E., Zio, E., Annunziata, F., Carlevaro, F., et al. (2020). A reinforcement learning approach to optimal part flow management for gas turbine maintenance. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* 234, 52–62. doi:10.1177/1748006X19869750
- Dike, H. U., Zhou, Y., Deveerasetty, K. K., and Wu, Q. (2018). "Unsupervised learning based on artificial neural network: A review," in 2018 IEEE International Conference on Cyborg and Bionic Systems (CBS), Shenzhen, China, October 2018 (IEEE), 322–327.
- Elidolu, G., Akyuz, E., Arslan, O., and Arslanoğlu, Y. (2022). Quantitative failure analysis for static electricity-related explosion and fire accidents on tanker vessels under fuzzy bow-tie CREAM approach. *Eng. Fail. Anal.* 131, 105917. doi:10.1016/j.engfailanal.2021.105917
- Emerson (2022a). DeltaV lifecycle services - DeltaV preventive maintenance EMERSON. Available at: <https://www.emerson.com/en-us/catalog/deltav-preventive-maintenance> (Accessed October 11, 2022).
- Emerson (2022b). DeltaV preventive maintenance EMERSON. Available at: <https://www.emerson.com/en-us> (Accessed December 11, 2022).
- Gao, L., Jia, M., and Liu, D. (2022). Process digital twin and its application in petrochemical industry. *J. Softw. Eng. Appl.* 15, 308–324. doi:10.4236/jsea.2022.158018
- Hady, M. F. A., and Schwenker, F. (2013). Semi-supervised learning. *Handb. Neural Inf. Process.*, 215–239.
- Haghighatlari, M., and Hachmann, J. (2019). Advances of machine learning in molecular modeling and simulation. *Curr. Opin. Chem. Eng.* 23, 51–57. doi:10.1016/j.coche.2019.02.009



- Hodgson, M. (2013). Chevron Richmond refinery crude unit incident. *Inst. Chem. Eng.* 234, 12–16.
- Iman, M., Arabnia, H. R., and Rasheed, K. (2023). A review of deep transfer learning and recent advancements. *Technologies* 11 (2), 40. doi:10.3390/technologies11020040
- Iqbal, H., Haider, H., Waheed, B., Tesfamariam, S., and Sadiq, R. (2021). Benchmarking of oil and gas pipeline companies in British Columbia: Integrating integrity management program and safety culture using a risk-based approach. *Eng. Manag. J.* 00, 526–542. doi:10.1080/10429247.2021.1954818
- Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: A survey. *J. Artif. Intell. Res.* 4, 237–285. doi:10.1613/jair.301
- Khanzode, K. C. A., and Sarode, R. D. (2020). Advantages and disadvantages of artificial intelligence and machine learning: A literature review. *Int. J. Libr. Inf. Sci.* 9, 30–36.
- Klemas, V. (2010). Tracking oil slicks and predicting their trajectories using remote sensors and models: Case studies of the sea princess and deepwater Horizon oil spills. *J. Coast. Res.* 265, 789–797. doi:10.2112/10a-00012.1
- Kusumawardhani, M., Kumar, R., and Tore, M. (2016). Asset integrity management: Offshore installations challenges. *J. Qual. Maint. Eng.* 22, 238–251. doi:10.1108/JQME-06-2015-0023
- Ladosz, P., Weng, L., Kim, M., and Oh, H. (2022). Exploration in deep reinforcement learning: A survey. *Inf. Fusion* 85, 1–22. doi:10.1016/j.inffus.2022.03.003
- Lim, K. Y. H., Zheng, P., Chen, C. H., and Huang, L. (2020). A digital twin-enhanced system for engineering product family design and optimization. *J. Manuf. Syst.* 57, 82–93. doi:10.1016/j.jmsy.2020.08.011
- Liu, B., and Liu, B. (2011). *Supervised learning*. Heidelberg: Springer Berlin, 63–132.
- Liu, X., Yoo, C., Xing, F., Oh, H., El Fakhri, G., Kang, J. W., et al. (2022). Deep unsupervised domain adaptation: A review of recent advances and perspectives. *APSIPA Trans. Signal Inf. Process.* 11 (1). doi:10.1561/116.00000192
- Ma, S., Ding, W., Liu, Y., Ren, S., and Yang, H. (2022). Digital twin and big data-driven sustainable smart manufacturing based on information management systems for energy-intensive industries. *Appl. Energy* 326, 119986. doi:10.1016/j.apenergy.2022.119986
- Maguire, D., and Fish, R. (2019). Leverage digitalization for asset integrity management: Select the right management technologies for modernization. *Plant Eng.* 73, 51–53.
- Mathworks (2022). Digital twins Mathworks. Available at: <https://www.mathworks.com/discovery/digital-twin.html>.
- Matsuo, Y., LeCun, Y., Sahani, M., Precup, D., Silver, D., Sugiyama, M., et al. (2022). Deep learning, reinforcement learning, and world models. *Neural Netw.* 152, 267–275. doi:10.1016/j.neunet.2022.03.037
- Moos, J., Hansel, K., Abdulsamad, H., Stark, S., Clever, D., and Peters, J. (2022). Robust reinforcement learning: A review of foundations and recent advances. *Mach. Learn. Knowl. Extr.* 4 (1), 276–315. doi:10.3390/make4010013
- Oil & Gas IQ Editor (2019). “What does asset integrity mean for the oil and gas industry? Oil gas IQ.” Available at: <https://www.oilandgasiq.com/oil-gas/news/what-is-asset-integrity> (Accessed November 1, 2022).
- Oksanen, R. (2018). *New technology-based recruitment methods CORE View metadata, citation and similar papers at core*. Tampere, Finland: University Of Tampere.
- Poole, D., Mackworth, A., and Goebel, R. (1998). *Computational intelligence: A logical approach*. New York: Oxford University Press, Inc. doi:10.5860/choice.35-5701
- Rajulapati, L., Chinta, S., Shyamala, B., and Rengaswamy, R. (2022). Integration of machine learning and first principles models. *AIChE J.* 68, 1–17. doi:10.1002/aic.17715
- Ratnayake, R. M. C. (2012). Modelling of asset integrity management process: A case study for computing operational integrity preference weights. *Int. J. Comput. Syst. Eng.* 1, 3. doi:10.1504/ijcsyse.2012.044738
- Reddy, Y. C. A. P., Viswanath, P., and Reddy, B. E. (2018). Semi-supervised learning: A brief review. *Int. J. Eng. Technol.* 7 (1.8), 81. doi:10.14419/ijet.v7i1.8.9977
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Comput. Sci.* 2 (3), 160. doi:10.1007/s42979-021-00592-x
- Sattari, F., Lefsrud, L., Kurian, D., and Macciotta, R. (2022). A theoretical framework for data-driven artificial intelligence decision making for enhancing the asset integrity management system in the oil & gas sector. *J. Loss Prev. Process Ind.* 74, 104648. doi:10.1016/j.jlp.2021.104648
- Sebbar, A., Zkik, K., Belhadi, A., Benghalia, A., Boulmalf, M., and Ech-Cherif El Kettani, M. D. (2022). BDA-Enabler architecture based on cloud manufacturing: The case of chemical industry. *Int. J. Supply Oper. Manag.* 9, 251–263. doi:10.22034/ijssom.2021.109150.2208
- Senthil, C., and Pandian, R. S. (2022). Proactive maintenance model using reinforcement learning algorithm in rubber industry. *Processes* 10, 1–18. doi:10.3390/pr10020371
- Sommers, D., Menkovski, V., and Fahland, D. (2023). Supervised learning of process discovery techniques using graph neural networks. *Inf. Syst.* 115, 102209. doi:10.1016/j.is.2023.102209
- Tang, H., and Jia, K. (2022). “Towards discovering the effectiveness of moderately confident samples for semi-supervised learning,” in In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, June 2022, 14658–14667.
- Tao, F., Xiao, B., Qi, Q., Cheng, J., and Ji, P. (2022). Digital twin modeling. *J. Manuf. Syst.* 64, 372–389. doi:10.1016/j.jmsy.2022.06.015
- Trauer, J., Pflingstl, S., Finsterer, M., and Zimmermann, M. (2021). Improving production efficiency with a digital twin based on anomaly detection. *Sustainability* 13 (18), 10155. doi:10.3390/su131810155
- Valentine, D. T., and Hahn, B. D. (2021). “Essential MATLAB for engineers and scientists,” in *Elsevier katey birtcher*. 8 (Massachusetts, United States: Academic Press).
- Viola, J., and Chen, Y. Q. (2020). “Digital twin enabled smart control engineering as an industrial AI: A new framework and case study,” in 2nd Int. Conf. Ind. Artif. Intell. IAI 2020, Shenyang, China, October 2020. doi:10.1109/IAI50351.2020.9262203
- Wang, J., and Biljecki, F. (2022). Unsupervised machine learning in urban studies: A systematic review of applications. *Cities* 129, 103925. doi:10.1016/j.cities.2022.103925
- Werner, S. (2021). Chemical process models – from first principle to hybrid models Navigance. Available at: <https://blog.navigance.com/chemical-process-models#First-principle-models> (Accessed February 11, 2022).
- Winkler, D. A. (2020). Role of artificial intelligence and machine learning in nanosafety. *Small* 16, 1–13. doi:10.1002/smll.202001883
- Wright, L., and Davidson, S. (2020). How to tell the difference between a model and a digital twin. *Adv. Model. Simul. Eng. Sci.* 7 (1), 13. doi:10.1186/s40323-020-00147-4
- Yang, X., Song, Z., King, I., and Xu, Z. (2022). A survey on deep semi-supervised learning. *IEEE Trans. Knowl. Data Eng.*, 1–20. doi:10.1109/tkde.2022.3220219
- You, Y., Chen, C., Hu, F., Liu, Y., and Ji, Z. (2022). Advances of digital twins for predictive maintenance. *Procedia Comput. Sci.* 200, 1471–1480. doi:10.1016/j.procs.2022.01.348
- Zha, S., Lan, H. Q., Lin, N., and Meng, T. (2023). Degradation and characterization methods for polyethylene gas pipes after natural and accelerated aging. *Polym. Degrad. Stab.* 208, 110247. doi:10.1016/j.polymdegradstab.2022.110247
- Zhang, H., Seal, S., Wu, D., Bouffard, F., and Boulet, B. (2022). Building energy management with reinforcement learning and model predictive control: A survey. *IEEE Access* 10, 27853–27862. doi:10.1109/access.2022.3156581
- Zhao, Z., Zhou, L., Duan, Y., Wang, L., Qi, L., and Shi, Y. (2022). “Dc-ssl: Addressing mismatched class distribution in semi-supervised learning,” in In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, June 25 2021, 9757–9765.
- Zhou, X., Liu, H., Pourpanah, F., Zeng, T., and Wang, X. (2022). A survey on epistemic (model) uncertainty in supervised learning: Recent advances and applications. *Neurocomputing* 489, 449–465. doi:10.1016/j.neucom.2021.10.119
- Zhou, Z. H. (2018). A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* 5 (1), 44–53. doi:10.1093/nsr/nwx106
- Zhou, Z., Lu, X., Zhuang, M., Hu, L., and Xia, G. (2014). Design and development of a device management platform for EAST cryogenic system. *Fusion Eng. Des.* 89, 595–599. doi:10.1016/j.fusengdes.2014.03.023