Contents lists available at SciVerse ScienceDirect

CIRP Annals Manufacturing Technology

Journal homepage: www.elsevier.com/locate/cirp

Big data analytics for smart factories of the future

Robert X. Gao (1)[a], Lihui Wang (1)[b], Moneer Helu (3)[c] and Roberto Teti (1)[d]

[a] Department of Mechanical and Aerospace Engineering, Case Western Reserve University, Cleveland, OH, USA
[b] Department of Production Engineering, KTH Royal Institute of Technology, Stockholm, Sweden
[c] Engineering Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA
[d] Department of Chemical, Materials and Industrial Production Engineering, University of Naples Federico II, Naples, Italy

Continued advancement of sensors has led to an ever-increasing amount of data of various physical nature to be acquired from production lines. As rich information relevant to the machines and processes are embedded within these "big data", how to effectively and efficiently discover patterns in the big data to enhance productivity and economy has become both a challenge and an opportunity. This paper discusses essential elements of and promising solutions enabled by data science that are critical to processing data of high volume, velocity, variety, and low veracity, towards the creation of added-value in smart factories of the future.

Keywords: Digital Manufacturing System, Information, Learning

1. Introduction

The first modern use of the word "data" refers to "transmissible and storable computer information" [161]. As data have gradually permeated throughout all aspects of the modern society, the meaning of the word has evolved to "information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful" [139]. This shift reflects how data has been transformed from a passive information carrier to an active value enabler.

1.1. From data to big data

The proliferation of computers, Internet, sensors, mobile devices, and smartphones has fundamentally changed the way data are generated, collected, transmitted, and stored. In terms of data volume, approximately 3 Exabytes ($3x10^{18}$ bytes) of data existed globally in 1986, whereas by 2011 over 300 Exabytes of data were stored [82]. The pace of data generation and collection has been accelerating drastically especially in the last decade [18, 82]. Data are not just "big" in terms of volume and the rate at which they are collected and stored, but have also become "big" in terms of their diversity and richness, enabling a more comprehensive and descriptive digital reflection of the physical world that can generate significant value to support policy making [44].

The term "big data" generally describes data that may be of high *volume* or *variety* or that may be collected at high *velocity* with potentially high or low *veracity* such that increasingly specific analytical technologies are needed to transform it into *valuable* information [36, 136]. For example, one challenge of big data is that a traditional centralized data storage approach may no longer satisfy the significant increase in data volume, which can create an urgent need for distributed data handling, storage, and management techniques. Similarly, the increased rate of data generation continuously challenges data transmission standards and techniques, making it difficult to leverage collected data for timely decision making. Furthermore, collected data have become increasingly diverse, including structured (e.g., data recoded in spreadsheet), semi-structured (e.g.,, data recorded in markup languages such as extensible markup language, or XML) and unstructured (e.g. text, audio, image, and video) data. Processing these forms of data requires specific techniques to be effective and efficient, which refers to how well the outcome of the performed data processing task meets the expectation, and if the time or computational load expended to perform the task is optimized, respectively. Finally, distinguishing between reliable and unreliable data has become more difficult due to the lack of tools to quantify the uncertainty involved [95].

Despite these challenges, the rich information embedded in data has led to the proclamation that data is the most valuable resource of the world today [44]. Effective extraction and use of data have become the next frontier to drive innovation, competitiveness, and economic growth in many industries including retail, finance, healthcare, transportation, and manufacturing [136]. Indeed, the advancement of computational infrastructure and innovation in data analysis techniques has allowed industry to begin to harness the insight embedded in big data to improve value creation [82].

1.2. Data as a co-product of manufacturing

When data were manually recorded, the amount of data was low, the quality was inconsistent, and the associated value was little to support improvement of the manufacturing processes [41, 145]. As digital sensors have increasingly replaced manual data recording, and sensor-rich machines become commonplace on the factory floors, the availability of large amount of high-quality, high-value data has fundamentally shifted the role of data, making it an insepareable co-product of modern manufacturing (Fig. 1).

The 1st industrial revolution greatly expanded production capability due to the emergence of the steam engine, but it had little impact on data collection and use. With the increasing demands for mass production, the 2nd industrial revolution has highlighted the need for controlling production quality [86]. As a result, quality variables associated with production began to be measured, and scientific methods were developed to process the collected data [144]. Monitoring charts were introduced to track quality variables for defect detection, representing the beginning

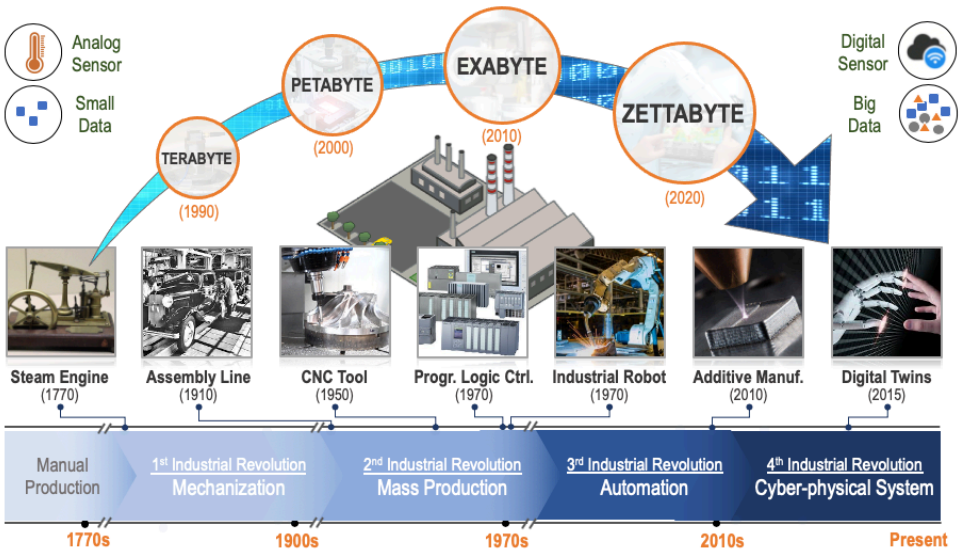

Fig. 1. Evolution of manufacturing systems and data as co-product, adapted from [56]

of statistical process monitoring (SPM) [73, 114]. Fundamental work on topics such as design of experiments (DoE) and response surface methodology (RSM) were conducted to systematically investigate the causal relationship between variations in process parameters and the conformance to product quality for improved process control and optimization [144]. These techniques assume that the collected data are samples generated from models that can provide insight to process status and product quality (e.g., whether a process parameter has significant influence on quality) [19, 144]. These methods have contributed significantly to reducing variability and improving quality and have revealed an increasing awareness of the purpose and value of data [144]. As explained by W. Edwards Deming, "the ultimate purpose of taking data is to provide a basis or a recommendation for action" [37].

The 3rd industrial revolution witnessed the shift from manual production to digital technology-enabled automation with the adoption and proliferation of computers and sensors. A great variety of sensors and machine controllers have been deployed for process monitoring and fault detection [105, 111, 207]. Information systems (e.g., enterprise resource planning, or ERP) started to be deployed to facilitate the management of production information (e.g., orders, materials supply, and production capacity). Enabled by numerical simulations (e.g., computer-aided design, or CAD, and finite element analysis, or FEA), manufacturing processes were decomposed into specific steps and reconstructed as virtual models for analysis, verification, and improvement [180]. As a result, the diversity of data in manufacturing has expanded from single measurements of quality to a mixture of data from transactions, simulations, scheduling, production, and maintenance each of which offers enormous potential as sources of new knowledge generation.

The accelerated availability of a large amount of data has led to questions about how to make effective use of data, as it has became obvious that statistical models are increasingly limited by the complexity of and uncertainty associated with data to gain real insight [73]. In addition, traditional analysis tools have been limited to a small selection of models and methods that have been unable to tackle the heterogeneity of manufacturing tasks that are increasingly distributed across different levels of a manufacturing system including system-, machine/process- and material-level [19]. Driven by the need for data mining to discover information underlying production systems for purpose of operation monitoring, fault diagnosis, and performance prognosis, data-driven techniques such as machine learning (ML) have been investigated, starting in the 1960's, to complement physics-based analysis and numerical simulations [90, 168]. These techniques can inductively learn relevant patterns from the data processed, and relate them to product quality in the form of quality prediction. Effective applications of ML techniques such as artificial neural networks (ANNs), random forest (RF), and support vector machines (SVM) have been reported in numerous manufacturing applications [67, 71, 121, 183]. The ever increasing availability of manufacturing related data has made the need for improving data quality more critical. As data grow in terms of volume, velocity and variety, data quality or veracity, such as erroneous, missing, or contradictory data, uncertainty, and information redundancy, have taken on increased significance. Developing proper techniques to mitigate data quality issues have become a major focus for reliable big data analytics [25].

As the need for quality, flexibility, and efficiency in manufacturing continue to grow and new paradigms such as mass personalization emerge [86], a deeper understanding of production machines and process as part of the cyber-physical systems (CPS) paradigm has become a central topic of the 4th industrial revolution (Fig. 1). CPS refers to "physical and engineered systems whose operations are monitored, controlled, coordinated and integrated by a computing and communicating core" [143]. This definition provides the vision for the "smart factories of the future", which are characterized by the timely acquisition, distribution, and utilization of information from machines and processes on manufacturing shop floors where big data analytics will play a critical role in dynamically linking all the operations within the factories and retrieving knowledge from the data to enable in-situ monitoring, operation optimization, informed decision-making and adaptive control, with humans in the center of the loop. More events on the physical shop floor are being recorded, communicated, analysed, and used for continuous improvement enabled by sensing, connectivity, computing, and learning techniques [190]. Factory floors have evolved into fully-connected, digitalized, and intelligent data acquisition platforms.

The digital transformation of manufacturing has drastically expanded the horizon of data generation throughout production and has provided unprecedented data availability and diversity. Although such trends have been observed since the beginning of the 3rd industrial revolution, recent progress across several technological frontiers suggests that the 4th industrial revolution is different because of:

- The development of sensors and the maturation of wireless technologies have allowed data collection and communication unconstrained by limitations typically encountered in manufacturing plants [54];
- Advancements in computational infrastructure, represented by cloud and edge computing, have made the management of big data feasible to support tasks of different temporal requirements, from process control that requires m-second-

level response, to production scheduling that requires second-level adjustment [22, 188, 220];

- Breakthrough technolgies, such as deep learning (DL), have enabled more powerful pattern recognition and provide the computational backbone for learning from a large amount of manufacturing data to improve decision making [117, 217];
- Digital twin has emerged as a new tool that can optimize and validate process control in the data-constructed virtual space, leading to improvement on the physical shop floor [3, 202];
- The advent of distributed manufacturing and Manufacturing-as-a-Service (MaaS) has transformed traditional centralized factories to more service-oriented and personalized manufacturing resources [115].

This paper discusses the essential elements and promising solutions that are critical to advanced manufacturing in the 21st century, which explore big data and data analytics as an enabling tool, with a focus on the data generated by and collected from machines and processes within the factory. Data associated with the business and supply chain's aspects are considered out of the scope of the paper, thus are not included in the discussion. As illustrated in Fig. 2, the paper uses the data lifecycle from data generation to transmission, processing, storage, and learning as an underlying structure to understand how to build smart factories of the future. Throughout this paper, the 5Vs that characterize the big data paradigm (i.e., volume, velocity, variety, varacity, and value) are reflected in the discussion. Data security is also discussed, followed by the examples of successful implementations with industry. Finally, topics for future research are highlighted to summarize this paper.

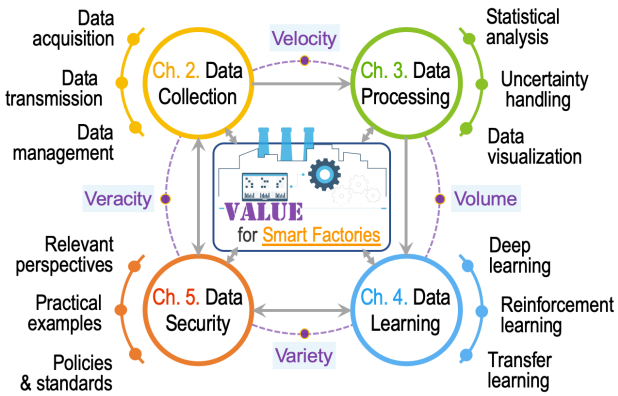

Fig. 2. Schematic relationship among key elements in big data analytics and smart factories

2. Data Collection

The availability and accessibility of data across the entire spectrum of manufacturing has grown at an unprecedented rate. These data can be broadly classified to five categories: 1) Management data from information systems, such as those related to production planning and inventory management; 2) Process data from sensors, e.g., on real-time machine performance; 3) User data from the written logs and online behaviors related to web browsing, purchasing, and review history; 4) Product data from its lifecycle related to performance and the context of use; 5) Public data from regulatory institutions through open databases, such as regulations and industrial standards [201].

Using this data presents challenges that are unique to manufacturing due to the large range of temporal scales over which analysis and decision making must occur, as shown in Fig. 3 [215]. Supporting such diverse temporal scales requires manufacturers to address traditional questions, such as which variable to measure and how to measure it, as well as how the acquired data should be transmitted, stored, contextualized, and computed to ensure that subsequent data analysis can be efficiently conduted in a timely manner. The state-of-the-art for data collection is summarized in Table 1.

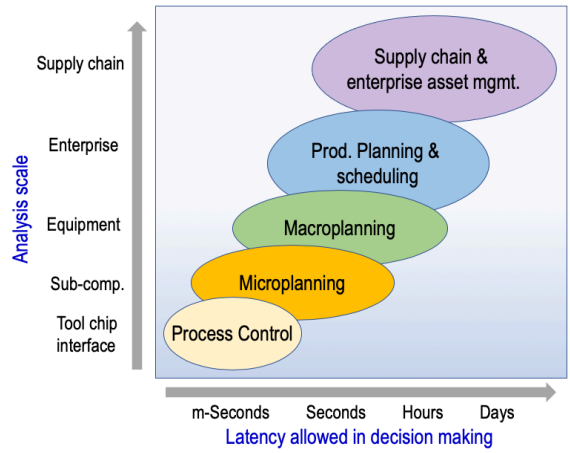

Fig. 3. Temporal decision scale in manufacturing, adapted from [215]

2.1. Acquisition

The Sensors and Sensor Networks program of the US National Science Foundation described the convergence of the Internet and communication and information technologies with techniques for miniaturization as having "placed sensor technology at the threshold of a period of major growth" [158]. The subsequent decade has witnessed the rapid development of sensing and data acquisition technologies. Advances in sensor design and realization have created new ways for acquiring increasingly diverse and high quality data at high speed, which has greatly enhanced the observability of manufacturing processes [28].

2.1.1. Process-embedded sensing

While manufacturers have the domain knowledge needed to identify the desired process data to measure, the physical constaints and adverse operating conditions may limit measurement, and limited commercially available sensors often prevent acquisition. The past decade has witnessed new, process-embedded sensing designs that have helped to overcome these limitations and have enabled the acquisition of critical variables [28, 49, 51, 101]. For example, Smolenicki *et al.* described a spring-loaded pin design to measure the interface friction coefficient of an orthogonal turning process at cutting speed up to 300 m/min [192]. Groche *et al.* developed a sensoric fastener for structural

Table 1. State-of-the-art for data collection

Big data V's	Data acquisition	Data transmission	Data management
Volume	Process-embedded sensing design [49, 51, 61, 192]	Compressive sensing [127, 130, 241]	Distributed storage: NoSQL database [21, 31, 85, 91, 105]
Velocity	High speed data acquisition: X-Ray imaging [164, 175]	Edge computing [125, 169]	
Variety	Multi-variate sensing design [101]	Acoustic wireless transmission [54]	Semantic indexing [39, 103, 232]
Veracity	Trust-incorporated sensing network [6]	Redundancy reduction via compressive sensing [127, 130, 241]	Data cleansing [135, 236]

joint load measurement in stamping processes by embedding strain gages inside the fastener and a thermocouple for temperature compensation [61]. Fujishima *et al.* presented a capacitance-type coolant level sensor for milling processes where the coolant level corresponds to a different voltage output pattern of nine electrode pairs geometrically spread out along the sensing probe [51]. A capacitive pressure sensor was reported in [49] that measures the pressure distribution across the interface between the rotating roll and metal foils, which enables online surface texture quantification during the microrolling process. Kazmer *et al.* report on a multivariate sensor with acoustic wave-based wireless data transmission that enables the simultaneous measurement of temperature, pressure, melt flow velocity, and viscosity within the cavity of an injection mold using only one sensor package [101].

Machines are also increasingly equipped with commercial sensors, and this trend has in turn driven interest in data acquisition systems capable of managing big data. For example, Fujishima *et al.* describe a machining center equipped with 24 additional sensors beyond its base configuration [51], which include six accelerometers for spindle and table vibration measurement, two coolant level sensors, eight load cell sensors embedded in the adjustable legs, one current and one voltage sensor to measure energy consumption of the entire machine tool, three temperature sensors embedded in spindle and work table to compensate for thermal displacement, and three cutting force sensors. As a result, a customized platform was developed (see Fig. 4) with interface boards to allow manufacturers to leverage data acquired in-situ beyond the limitation associated with the PLC.

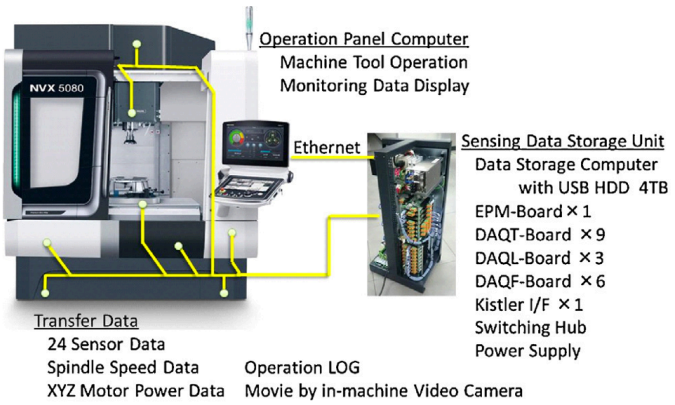

Fig. 4. Acquisition and storage system for machining center [51]

2.1.2. From time series to image data

Data generated from manufacturing processes has historically been limited to 1-D time series data, such as vibration and pressure. These provide temporal information for monitoring processes, such as metal forming, and machine components, such as induction motors [28, 129]. In comparison, image data, including 2-D images and 3-D videos, provide spatial information that is desirable for many applications, such as surface quality inspection. Images can also be captured in a contactless manner, which makes it particularly well suited for operations that requires non-intrusive measurement. The effective use of image data, however, requires novel signal processing capability and high computational power.

Recent developments in signal processing and computational hardware, such as graphical processing units (GPUs), have significantly reduced time required to process image data [117]. This has led to the rapid increase of image acquisition across manufacturing. Also, emerging manufacturing technologies, such as additive manufacturing (AM), have motivated the development and maturation of image-based techniques, which has in turn improved these methods [47]. In addition, novel applications enabled by image data have emerged, such as the use of speckle photography for in-process strain measurement at the machined boundary zone for grinding (Fig. 5) [203]. Such measurement was not feasible previously due to the harsh process conditions that prevent contact-based sensing.

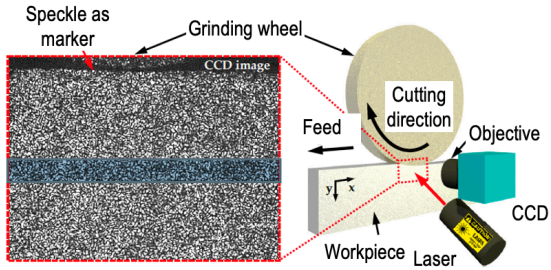

Fig. 5. Strain measurement during grinding by speckle photography, adapted from [203] (CCD: charge-coupled device)

2.1.3. Towards higher acquisition rates

Technological advances have also enabled the measurement of data at higher sampling rates than previously available. As an example, high-speed sensing has allowed processes to be probed with unprecedented temporal resolution [249].

Another example is online AM process monitoring. While widely considered as an indicator of process condition and part quality, in-process characterization of melt pool remains challenging due to its transient nature that requires micro-second sensing capability to track its evolution [47]. Such high-frequency sensing has been only achieved previously by temperature (e.g., two-color pyrometer) and visible-light or thermal (e.g. infra-red light) imaging, which are limited to surface level monitoring [52]. More recently, the development of high-speed X-ray makes it possible to capture the internal structure around the melt pool and consequently its full dynamics [249]. The high-speed X-ray imagining system developed in [175] (see Fig. 6) has achieved sampling rate of 50 kHz, which allows the melt pool surface wave movement in Selective Laser Melting (SLM) process to be quantified for process improvement and numerical modeling. Similarly, an ultrafast X-ray imaging system is described in [164], which has achieved a sampling rate of 6.5 MHz to capture the SLM phenomena such as vapor depression, melt-pool dynamics, and powder-spatter ejection.

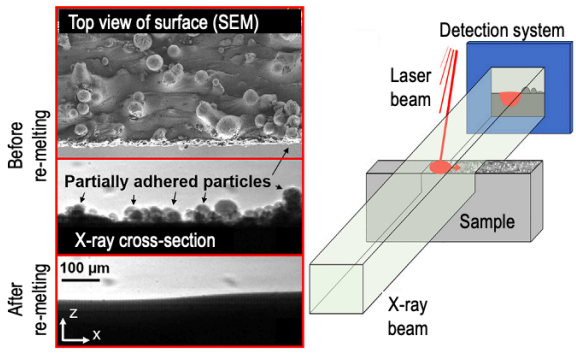

Fig. 6. High-speed X-ray imaging and observed remelting process, adapted from [175]

2.2. Transmission

High velocity data acquisition from a large number of data sources can create a heavy burden on data transmission infrastructure that can limit transmission bandwidth. Varying latency requirements for different data can create additional challenges. Data compression techniques can alleviate the bandwidth limitation in transmitting manufacturing data, and

techniques such as edge computing complement cloud computing in handling data with varying requirements of latency [188].

2.2.1. Bandwidth

Dictated by the Shannon-Nyquist sampling theorem, many process variables are measured at high acquisition rate in order to obtain useful information [127]. For example, vibration signals in rotatry machines are commonly measured at the micro-second scale. The result is a potentially massive amount of data to be transmitted at high velocity and severe usage of bandwidth [241]. The situation is further exacerbated by the addition of image data.

A technique that can potentially alleviate this issue is compressive sensing (CS). The CS theorem states that signal sparsity can be exploited through numerical optimization to allow signal recovery from far fewer data points than required by the Shannon-Nyquist theorem [27]. Much effort has been dedicated to the application of CS in reducing the amount of data needed for different types of analysis. For example, Liu *et al.* applied CS in reducing acoustic emission sensing data for bearing condition monitoring [127]. As shown in Fig. 7, by exploiting the sparsity of signal in the frequency domain, condition-related frequency information can be recovered from the compressed signal that contains only 1/8 of the original data volume. In a similar work, Yuan and Lu extended the CS-based method to bearing fault diagnosis under variable speeds [241]. Lu and Wang studied physics-based CS to monitor the temperature field of AM processes. They have shown that the data volume and number of sensors needed for process monitoring can be significantly reduced by leveraging prior knowledge of the physical quantities to be measured. A compression ratio of two orders of magnitude has been achieved as compared to standard CS methods [130].

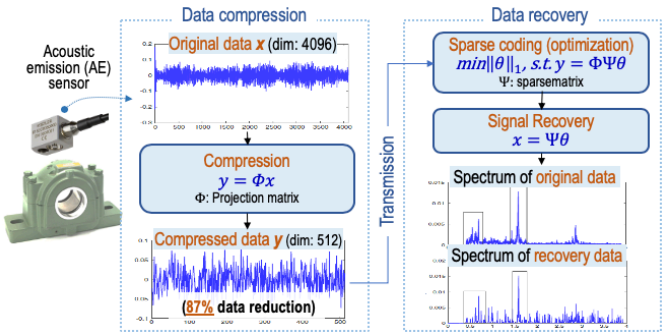

Fig. 7. Signal compression and recovery based on CS, adapted from [127]

2.2.2. Edge computing

Edge computing extends cloud computing to the source of data to address the need for time-sensitive data transmission. It helps to reduce the latency in transmission and free up bandwidth by allowing data to be stored and computed locally [17, 206, 209, 253]. A system consisting of sensors with edge and cloud computing (see Fig. 8) allows data transmission and subsequent storage and analysis to be adaptively allocated based on the requirement of each task with time-tolerant requests that do not require real-time responses transmitted to the cloud. Cloud computing provides scalable data storage and computational capability that can scale up to computation-intensive tasks, such as predictive modeling. In comparison, edge computing responds to the urgent and non-computationally intensive tasks triggered by machines or sensors that cannot be delayed due to the latency involved in transmission to the cloud [214]. Furthermore, edge computing leverages the spare capacity of locally available resources, such as narrow-band Internet of Things (NB-IoT), to achieve further reduced latency and perform computing in a more cost-effective manner [32]. Advances in IoT and edge computing have provided the technological foundation for transmiting high-speed, time-sensitive data and allow distributed manufacturing resources to be effectively integrated [188].

Qian *et al.* provided examples of edge computing in manufacturing such as an edge-based motor fault diagnosis sytem [169]. The motor is monitored by three current probes and a vibration sensor each sampling at 20 kHz. A diagnostic model is built offline and loaded to an edge device to monitor motor conditions in real-time with the fault detection latency being 0.25 s. Similarly, Li *et al.* reported on a vision system for assembly line monitoring where images captured by a mobile robot are analysed by a neural network tranined on a cloud platform and stored in edge devices [125]. To handle tasks that require more computational capability, a cooperative edge computing strategy is developed to dynamically allocate multiple devices for the task.

While edge computing has shown to improve the speed and efficiency of data transmission by allowing localized data analysis for time-critical decision-making, it comes with the trade-off that limits information associated with local data being propagated. This limitation can potentially impacts the degree of global optimality that can be achieved via edge computing. Effective analysis fusion strategy across different edge devices remains an open research topic, which aims to strike a balance between the speed of data analytics and the degree of optimality.

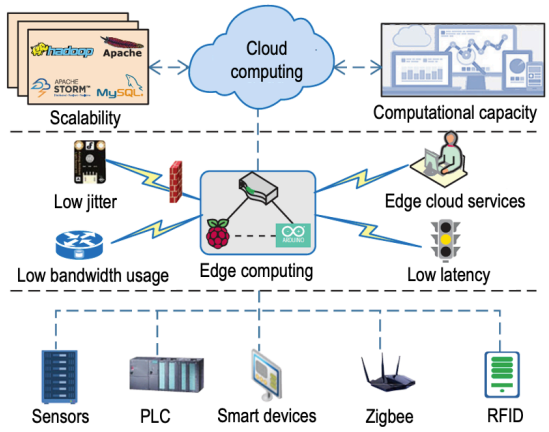

Fig. 8. Edge devices, edge and cloud computing, adapted from [30]

2.3. Data Management

The evolution of big data has presented challenges for data management since traditional data tools, procedures, and infrastructure have not been designed to manage data of high volume and variety, especially if that data have been generated in geographically dispersed silos [26, 100, 167]. Such complexity has motivated big data management as a new discipline with techniques that have been developed to address the storage, contextualization, integration, and access of big data to support subsequent data processing and learning [189]. Four representative considerations that are important to big data management are illustrated below.

2.3.1. Contextualization

Contextualization is the process of identifying the data relevant to an entity (e.g., person, thing, location, or organisation) based on the entity's contextual information [238]. Contextual information is any information about an entity that can be used to reduce the amount of reasoning required (e.g., via filtering, aggregation, or inference) for decision making within the scope of a specific application, and it enhances the processing of data in large-scale, data-intensive IoT applications from various big data aspects, including volume, velocity, and variety [239].

The quality of any knowledge generated from data analysis depends on the appropriateness of the context developed when

collecting and managing the data itself [13]. For example, production and maintenance personnel may require similar data from a piece of manufacturing equipment on the shop floor, but their viewpoint is influenced by different interests since production creates value when the equipment runs, while the opposite is true for maintenance. Given the variety of viewpoints in production, it is critical that data collection and management approaches support multiple viewpoints for different applications by dynamically linking different data, information, and models [76, 81, 172, 176, 180].

To address the need for context, manufacturing data researchers have focused on defining the semantics of data through data interoperability standards, such as MTConnect [148] and the Standard for the Exchange of Product Model Data (STEP or ISO 10303-242:2014) [87]. Two other relevant examples are OPC Unified Architecture (OPC UA), which focuses on syntactic interoperability but also supports semantics through different companion specifications [162], and the Universal Machine Tool Interface (umati), which is a standards branding effort through OPC UA proposed by the German Machine Tool Builders' Association (VDW) and Mechanical Engineering Industry Association (VDMA) [57]. As data analysis in manufacturing has moved beyond correlation towards causation (i.e., the reason that explains an observation), though, data scientists and manufacturing solution providers at the leading edge have recognized a fundamental limitation inherent to the data currently collected from production systems [13]. That is, the semantics and information models used today may not enable the collection of data of sufficient quality to identify causation. In response, researchers have started to explore ways to incorporate more context to data by creating additional links between collected data through alternative concepts, such as graph theory. For example, the idea of a minimum information model that identifies the minimum number of links between different pieces of information to establish a complete product definition has been explored by [178]. Similarly Bajaj and Hedberg presented a preliminary implementation of a linked data graph based on the Handle system that connects information across the product lifecycle [8]. Much of this work forms the foundation of the Model-Based Enterprise (MBE) concept, which describes the use of digital models to support decision making throughout the product lifecycle [1]. MBE leverages notions of semantics, context, and viewpoint interoperability from the digital thread (i.e., linked systems across the product lifecycle [76]) to enable better knowledge extraction and subsequently improved decision making. MBE has become a topic of great interest with organizations such as American society of mechanical engineers (ASME), which has formed a Steering Group and Standards Committee to explore and develop MBE for implementation in the industry [2].

2.3.2. Semantic indexing

The prerequisite to intepret data and support subsequent data analysis, such as supervised learning, is proper data labeling. While syntax concerns with whether the data are valid from a data structure point of view, semantics refers to the meaning of data and is therefore directly related to data analysis. Semantic methods can ensure performance efficiency for large data collection [53]. For example, a semantically-enhanced cloud service environment can be developed using ontological methods to facilitate the discovery of resources that meet the users' needs.

Semantic indexing forms the basis for enhanced search processes for big data, and supports data variety, veracity, and value. Furthermore, it is also realistic in handling the generation of semantic annotations from unstructured documents.

Semantic annotation (also known as semantic tagging or semantic enrichment) is the process of attaching additional information to various concepts so that the annotated data are machine interpretable [103]. Written in machine-interpretable data language, these notes allow computers to perform operations such as classifying, linking, inferencing, searching, and filtering [39]. A example of semantic data defined by the MTConnect standard is shown in Fig. 9 [120].

In semantic-based indexing, each annotation of every document is stored in a database, and a weight is assigned to reflect how relevant the ontological entity is to the document meaning. The main idea is that the more semantically "close" two concepts are in a document, the higher the vector values [34].

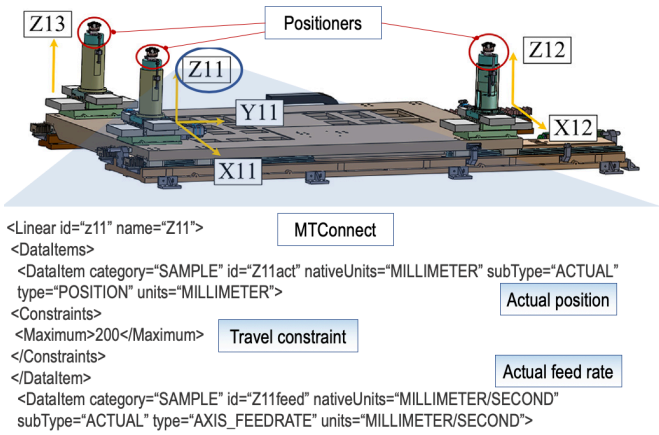

Fig. 9. MTConnect semantic data for axis Z11, adapted from [120]

A method for indexing semantic, non-transitory, computer-stored data was developed by [232] and comprises the following steps: (1) storing the data in a database, (2) representing the data in a structured framework having at least three elements derived from an ontology, (3) expressing each element as a hierarchical-index value based on an ontology such that semantic information is embedded therein, (4) combining the elements in a multi-dimensional index, and (5) converting the multi-dimensional index into a one-dimensional index. US Navy researchers extended this process (see Fig. 10) by first organizing the data in a database using the resource description framework (RDF) model [232]. The result is an one-dimensional index created from multiple RDF components embedding semantic information and providing facility for retrieving data from big data stores.

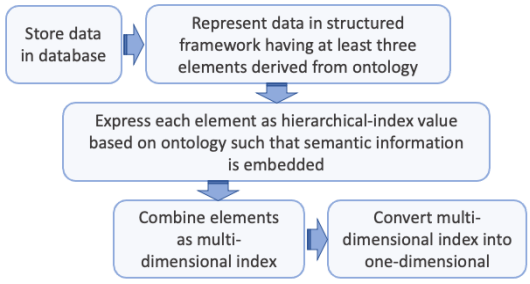

Fig. 10. Method for semantic indexing of big data using a multidimensional, hierarchical scheme, adapted from [232]

2.3.3. Data cleansing

Data cleasing refers to the process of detecting and correcting errors in a dataset to improve data quality [135]. It generally involves three steps: (1) define and determine error types, (2) search and identify error instances, and (3) correct the uncovered errors [135]. While many modern database systems support basic data cleansing, errors that involve relationships between one or more data attributes are often difficult to detect. For error detection, four general approaches are available [135]: (1) Statistical method, (2) clustering, (3) pattern-based method, and (4) association rules.

Despite their wide adoption, these approaches are generally designed for simple data tables and difficult to apply directly to manufacturing related big data. Specialized techniques have been developed recently to tackle data cleansing in a variety of data types. In [251], a cleansing technique for high-dimensional Radio Frequency Identification (RFID) data has been presented. The key steps involve a series of logic functions based on the domain knowledge of production schedule and event, which allows the algorithm to be efficiently executed in real time to process a large amount of RFID data and remove duplicate, merge redundant attributes and fill missing data point. In [236], local outlier factor (LOF) has been investigated for erroneous data detection in high-speed time series data. Using a sliding window the sensing signal is first divided into multiple segments which are considered as different objects, each of which having the attributes of time-domain statistical features, such as the mean and peak-to-peak value. Next, kernel-based LOF is computed to evaluate the degree of each segment being considered as erroneous. Finally, the erroneous segments are detected based on an application-specific threshold. The effectiveness of the method has been evaluated with case studies of wind turbine and gearbox. Advancement in the research of correcting erroneous image data has also been reported recently, with sparse representation and deep learning as representative techniques [166]. Successful applications include super-resolution, which is the process of upscaling to improve details within an image, and image recovery from missing pixels and significant noise [226]. Applications of these techniques in manufacturing have yet to be identified.

2.3.4. Data storage

Data storage management has become a fundamental concern as data volumes often exceed the capacity of storage hardware. Moreover, data variety, especially unstructured data such as images, makes the traditional relational databases unsuited [85]. The complexity of big data has motivated the development of new storage techniques capable of scaling with data quantity, optimizing availability, and improving retrieval speed [21].

The process of managing storage for rapidly growing amounts of data involves performing activities, such as data clustering, replication and indexing, in parallel to optimizing the storage process. The following are considered critical to address storage in big data management:

- *Clustering* is the process of summarizing large volumes of data into groups where similar feature entities are placed together. Clustering facilitates the accommodation of a large volume of data in relatively small and limited storage resources [200].
- *Replication* is the key factor in improving the availability and robustness of data in distributed systems. Replicated data are stored at multiple sites to enable consistent access even when some copies are not available due to site failures [7].
- *Indexing* improves the efficiency of data retrival. It addresses the challenge of obtaining optimized query execution results when a large volume of data are stored at distributed sites and facilitates the efficient use of data for decision making and value extraction [53].

Another important aspect of big data storage is data variety. Manufacturing data can be generally classified into structured, semi-structured, and unstructured data [23]. Traditionally, manufacturers have been focused on structured data storage in relational database since it can be difficult to manage unstructured data due to a lack of techniques. The advances in development of nonrelational (often "NoSQL") databases has provided an means to cope with big data heterogeneity in storage [31].

Selection from the four types of NoSQL databases: (1) key-value stores, (2) column-oriented databases, (3) document databases, and (4) graph databases depends on the properties of the data and type of application [85, 91]. For example, for fault detection in plastic injection molding machines [105], the data structure is heterogeneous, but the object of analysis (i.e., manufacturing cycle) is clearly defined. This suggests the use of a document-oriented NoSQL database, which allows each sample to have a completely different set of attributes.

3. Data Processing

After data is acquired, transmitted, and stored, data analysis is performed to generate knowledge about the process [67]. Data analysis methods can be classified into two categories depending on function: data processing and learning.

Data processing is traditionally built on statistical models that aim at inferring process status and optimizing quality-related parameters. As data become more complex and applications more heterogeneous, these methods have gradually become insufficient and are increasingly enhanced by other techniques [73]. Conversely, data learning aims to learn from data the patterns related to quality [137] and has become increasingly important. With the continual increase in data volume and variety, the evaluation of data quality, e.g., uncertainty and redundancy, has also become a major focus of data processing [25]. This chapter focuses on recent developments in data processing, with the state-of-the-art summarized in Table 2.

3.1. Statistical analysis

Statistical methods, such as SPM and DoE, form the foundation of process monitoring and optimization [144]. When data volume and variety is low, these techniques are effective in their respective applications, but as data volume and variety increase, new techniques are needed to enhance the their capability.

3.1.1. Statistical process monitoring

Traditional SPM assumes that for an "in-control" process, process variables (e.g., pressure) and product quality (e.g., dimension) follow a Gaussian distribution. Accordingly, control limits can be set using statistics, such as t- or Hotelling's t-squared, to alert on the potential occurrence of faults or anomalies [134].

Table 2. State-of-the-art for data processing

Big data V's	Statistical analysis	Uncertainty handling	Data visualization
Volume	Definitive screening design [97-99]	Uncertainty decomposition [66, 96, 116)	Information graphics: sunburst chart, heat map, parallel coordinates chart, Sankey diagram, directed network [12, 20, 42, 79]
Velocity	Distributed data processing: edge computing [188]; distributed storage [22]; parallel algorithm (e.g., parallel particle filter)		
Variety	Multiple profiles sensor-based statistical process monitoring [243]	Multi-variate uncertainty propagation modeling: Bayesian network [150, 179, 181]	Information graphics: sunburst chart, heat map, parallel coordinates chart, Sankey diagram, directed network [12, 20, 42, 79]
Veracity	Non-Gaussian statistical process monitoring: statistical pattern-based, kernel independent component analysis-based, multiple profiles sensor-based [72, 118, 243]	Data filtering: multi-mode Kalman filtering, local search particle filtering [174, 222]	Correlation analysis: Maximal Information Coefficient [173]; Topological data analysis: Mapper clustering [62, 63, 191]

As data volume and variety increase, factors such as nonlinearity among variables and dynamics (e.g., correlation among time steps) can violate the Gaussian assumption [73, 170], making the traditional statistics insufficient to characterize the process and reducing the effectiveness of SPM. Research efforts to overcome this limitation have been focused on features that capture the dominant process characteristics. Furthermore, these methods alleviate issues created by high data volume and variety by condensing data into a compact feature set [73].

Lee *et al.* investigated kernel independent component analysis (KICA) for SPM to extract the dominant independent components capturing data nonlinearity [118]. The developed method has been evaluated for process fault detection using the Tennessee Eastman (TE) chemical production benchmark dataset, which consists of five operating units, 52 process variables, and 21 faults. The KICA-based method has shown to outperform traditional SPM in detection accuracy. Guo *et al.* reported an iterative control limit tuning method for feature-based process monitoring, which accounts for both univariate limit and multivariate limit among sensing signals [64]. It has been successfully applied in quality control of lithium-ion battery welding process to eliminate the Type II error, as shown in Fig. 11.

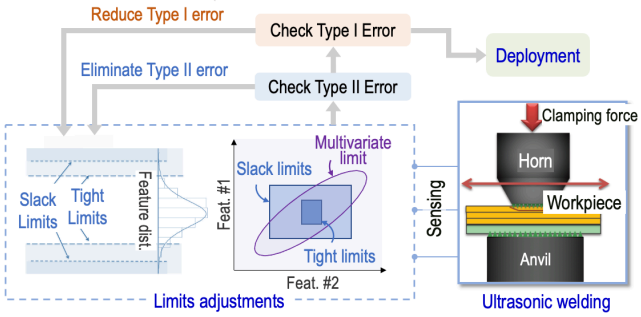

Fig. 11. Ultrasonic welding control limit tuning, adapted from [64]

He and Wang developed statistics pattern analysis (SPA) to account for both nonlinearity and dynamics by forming a comprehensive set of features to capture variable characteristics (e.g., mean), interactions (e.g., correlation), and dynamics (e.g., autocorrelation) [72, 74]. Principal component analysis (PCA) is used to quantify the dissimilarities among statistical patterns (SPs) to define detection index. When a new measurement is available, dissimilarity between its SP and training SPs is compared to a pre-defined threshold for fault detection.

One assumption in feature-based SPM research is that multi-variate signals have similar characteristics that can be described by shared features. However, significant sensor heterogeneity may occur, leading to data veracity issues, such as out-of-sync, drift, and inter-correlation. In [243], a SPM framework to account for sensor individuality is developed. Specifically, dynamic time warping is applied to aligning out-of-sync signals. Sensor drift is compensated for via numerical optimization, and clustering is carried out for sensor correlation. The developed method is evaluated in a manufacturing process monitored by 26 sensors, with out-of-control average run length (ARL) showing how fast an out-of-control process is detected as a performance indicator. Compared to other feature-based SPM, Significant reduction in ARL is shown.

3.1.2. Design of experiments

Process optimization requires the knowledge of influential parameters and their causal effect on quality. The effective use of DoE has proven crucial in screening candidate parameters and determining causal effect for optimization, leading to products with higher quality and reliability [144]. However, when the number of parameters becomes large, experimental time and data volume increases significantly and quickly become an issue. The most common solution has long been fractional-factorial designs [144]. Despite its popularity, these designs suffer from several limitations pertinent to big data, including (1) difficulty to estimate curvature in response surface, which is expected in complex processes, (2) undesirable confounding effects requiring additional experimental runs, and (3) poor scalability since the number of runs increases exponentially with the number of process parameters [97, 144]. Recently, definitive screening design (DSD) has been developed and shown the potential of extending DoE methodology to big data [97]. Desirable properties of DSD include: (1) quadratic effects are quantifiable, (2) main effects are completely independent of two-factor interactions and these interactions are never completely confounded with each other, and (3) number of required experimental runs only increases linearly with process parameters. Different from traditional experimental design, DSD is constructed via numerical optimization. Recent work has further extended DSD for nominal inputs and blocking, which arranges experimental runs into groups (termed blocks) that are similar to one another [98, 99].

DSD has been successfully applied in a variety of fields. Erler *et al.* investigated DSD to determine the effect of six process parameters on a formylation protein-crosslink reaction [45]. The authors reported that the number of required experimental runs could be reduced from over 70 with a traditional approach to 17, corresponding to a significant reduction in data volume [45]. In manufacturing, Patil investigated DSD to establish a predictive model to optimize weld tensile strength and hardness in gas tungsten arc welding [165]. DSD has shown to be effective in revealing all main effects and quadratic effects and interactions, leading to models with good fit. Several recent publications on DSD are focused on AM process optimization by determining causal influence of process parameters on part quality [132].

3.2. Uncertainty handling

With the expansion of data volume and variety, data quality has become a critical issue in big data analysis [201]. Data quality can be described by several measures, including uncertainty, which may be caused by several factors (see Fig. 12). This section first presents uncertainty quantification (UQ) methods, followed by filtering for data quality improvement.

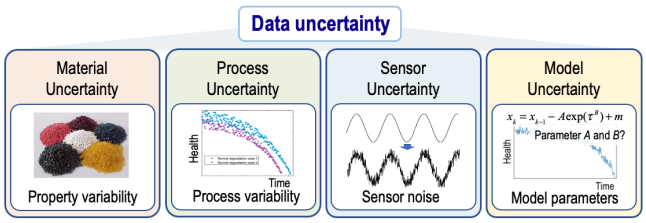

Fig. 12. Source of data uncertainty, adapted from [149]

3.2.1. Uncertainty quantification

The "standard" UQ metric is the standard deviation [96]. When a measurement is indirectly obtained from other variables, the combined uncertainty equals the root of the sum of the variances and covariances of these variables weighted by the partial derivatives (e.g., how the result varies as these variables change). This principle has been widely used since a direct UQ of a variable of interest is often difficult to obtain. In [116], a UQ method for micro gear measurement is developed using reference geometries, such as cylinders, to address the issue of a lack of micro-scale master gears. The sub-level uncertainties involve calibration of cylindrical standard, production process change, and system deviation. In [66], uncertainty in micro gear lifetime prediction is evaluated from two sources, tooth root stress from FEA and Weibull distribution parameters, with each having its uncertainty

further decomposed. For example, the uncertainty of tooth root stress is associated with FEA input and mesh discretization.

Increasingly, Bayesian network (BN) has shown to be effective in estimating system uncertainty propagation [179]. BN is a probabilistic graphical model consisting of nodes and arcs. The nodes represent system variables and arcs denote the conditional probability between the nodes. The parameters of the conditional probabilities are estimated from the collected data using the maximum likelihood (i.e., parameters that maximize the likelihood of generating the data) [181]. After parameter estimation, uncertainty can be propagated from variables to output with a distribution estimated via Markov Chain Monte Carlo (MCMC). BN can also reveal the contribution of each variable to the output uncertainty, which is useful to guide uncertainty reduction. A UQ case study for evaluating injection molding energy consumption was presented in [150], which took into account polymer properties and process parameters. The constructed BN is shown in Fig. 13. Sensitivity analysis has shown that polymer density is the most dominant contributor of uncertainty in total energy consumption.

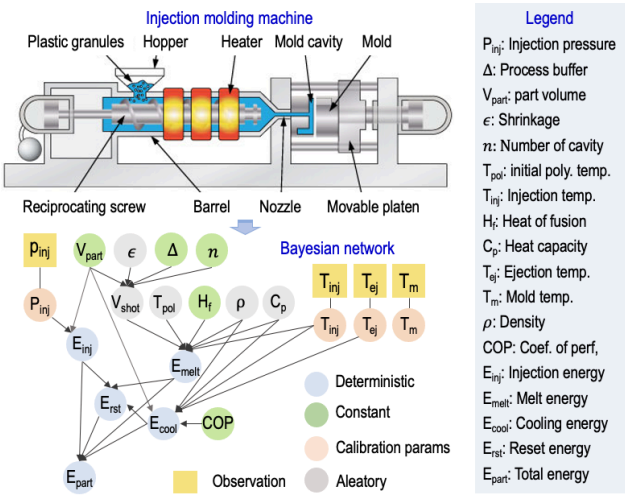

Fig. 13. BN for injection molding, adapted from [150]

3.2.2. Data filtering

One of the main purposes of UQ is to "update" raw sensor measurements to obtain a more accurate reading. Filtering is one key technique used. The concept of filtering is the alternating use of a state evolution model (relating current state to future state) and a measurement model (relating state to measurement) to more accurately approximate the true state. Kalman filter (KF) and particle filter (PF) are the most commonly investigated methods.

KF is based on the assumptions that (1) both process and measurement noise is Gaussian and (2) both state evolution and measurement model are linear [15]. It analytically combines estimators from these two models to improve estimation accuracy and reduce uncertainty [15]. Extended KF (EKF) and unscented KF (UKF) further extend KF to nonlinear systems [9, 16]. Recent developments of KF include a switching KF for bearing degradation phase identification [174] and an integrated KF and expectation-maximization algorithm to estimate both bearing degradation state and degradation model parameters [227]. Beyond state tracking, KF has also been widely used for simultaneous localization and mapping (SLAM), refering to the estimate of both robot and landmark locations, which is considered of great potential in human-robot collaboration [219].

Compared with KF, PF has the advantage of taking into account nonlinear and non-Gaussian property by design since it uses a set of weighted particles to sample the posterior distribution, regardless of its form [5]. This makes it attractive to tackle uncertainty in complex manufacturing systems. Furthermore, particle weights update are independent, meaning it can leverage parallel computing to tackle the data velocity issue. One major focus of PF research is the particle resampling strategy to maintain particle diversity for effective approximation of state posterior distribution. A weight-and-space based resampling method was developed for object location and velocity tracking in [124]. In [222], a local-search PF (LSPF) was presented that allows resampled particles to explore a wide range of values (see Fig. 14). It has shown to reduce tool wear prediction error from 11.7% to 3.5%, as compared to standard resampling [222]. PF has also been integrated with a total variation (TV) filter to detect abrupt faults in aircraft engines to support predictive maintenance [223].

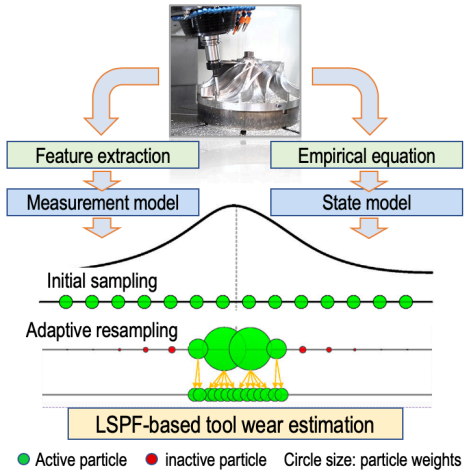

Fig. 14. LSPF-based tool wear estimation, adapted from [222]

3.3. Visualisation

Data visualization is the creation and study of visual representation of data [92]. Effective visual design helps to discover patterns and quickly gain insights. This section first presents representative graphics to improve the clarity in visualizing the structure of big data as compared to conventional charts. Then, two techniques to further understand data structure are presented.

3.3.1. Information graphics

Information graphics can be broadly categorized into five groups: time series, distribution, map, hierarchy, and network [79]. Conventional graphics, such as line chart, have become ineffective in visualizing the structure of big data that is often hierarachical and interconnected. Several less well-known but effective graphical methods have been reported recently. For example, alarm record analysis has been performed in a chemical production plant, which is divided into three levels: 1st level with 19 production units, 2nd level with 400 machines, and 3rd level with over 2,000 sensors/actuators distributed on the machines [42].

A sunburst chart, as shown in Fig. 15, is first utilized to provide an overview of the alarm distribution. This chart is well suited for displaying hierarchical data with each level of the hierarchy represented by a ring with the innermost as the top hierarchy. Each ring is further broken into its contributing segments (e.g., different units, machines, or sensors/actuators). The number in each segment represents the corresponding tag. For example, sensor/actuator #1688 on machine #467 in unit #14 has produced the largest percentage of alarms, as represented by the segment size, that is proportional to the alarm occurrence at that level.

To further reveal the alarm patterns, heat maps can be deployed, which a tool to reveal the relationship between two variables (as network) or the values of a function of two variables (as map) [20]. In this example, the frequency of alarm co-occurrence between two items is shown in Fig. 16. The color intensity represents the

alarm relative frequency. It is seen that most of the diagonal cells are in dark color, suggesting the alarm corresponding to these items frequently occured in pairs. On the other hand, dark-colored off-diagonal cells indicate potential causation between different items, e.g., between units #11 and #4.

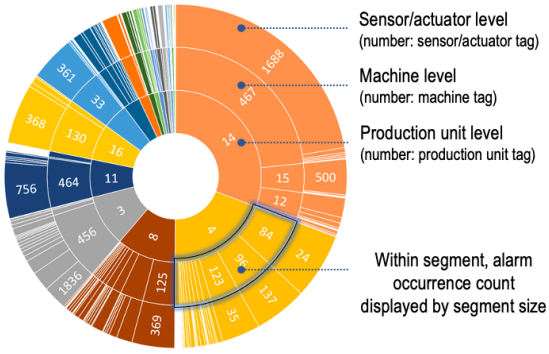

Fig. 15. Sunburst chart for overview of alarm record [42]

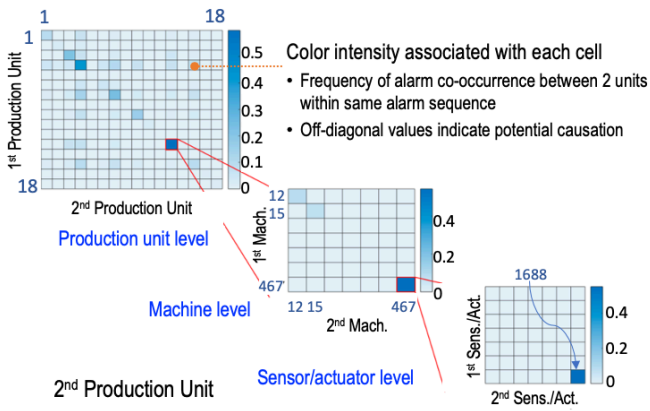

Fig. 16. Heat maps for alarm co-occurrence, adapted from [42]

Beyond pattern recognition in hierarchical data, another important application of visualization is high-dimensional data. For structured high-dimensional data, which can be stored in a table where each row represents a data point and each column represents a data attribute (e.g., feature of a product), techniques such as Spider web diagram and parallel coordinates plot [79] are commonly used for purpose of visualization. The former assigns data dimensions into different radial axes of the Spider web diagram, whereas the latter uses parallel vertical axes to represent data dimensions. The values from each dimension that correspond to the same data point are connected, and clustering of these connected lines allows the pattern of data at different dimension or groups of dimensions to be discerned. These two types of diagrams have been frequently used to compare features among different products for decision-making.

For unstructured high-dimensional data such as images, for which each of its dimension (e.g., pixel) does not represent a semantically meaningful data attribute, dimension reduction method is generally used by projecting the most essential information into low-dimensional space for visualization. One of the popular techniques developed recently is t-Distributed Stochastic Neighbor Embedding (t-SNE) [213]. This technique first constructs a probability distribution of pairs of high-dimensional data points such that similar data points have a high probability of being selected. Next, it defines a similar probability distribution over data points in a low-dimensional space and adjust the locations of these points by minimizing the Kullback–Leibler (KL) divergence between the two distributions, to mimic the structure of the high-dimensional data. t-SNE is widely used in big data analytics to verify data separation corresponding to various process or machine conditions in condition monitoring, which is described in Section 4.

Other noted graphics for big data visualization include the Sankey diagram [12, 112] and Directed Network [42]. Customized visual design is often required for specific tasks. Reported works include visualization of shop floor logistics [251, 252], production network [104] and production status [106, 107].

3.3.2. Correlation analysis

Data correlation is anticipated when the same process is measured using multiple sensors. How to distinguish the complementary information from redundancy has been an active research topic [216]. With big data, conventional correlation coefficients (e.g., Pearson) can become ineffective since the relationship among variables is often nonlinear [147]. Therefore, there is an increasing need for techniques to detect these relationships regardless of the forms [173].

Mutual information (MI) between two variables, a concept related to information theory and denoted as $I(X,Y)$, has provided a means of dealing with the increasing complexity in data correlation. $I(X,Y)$ can be interpreted as the change of information in X before/after having knowledge of Y with a small MI indicating a weak relationship. Reshef *et al.* developed a procedure to compute MI for large-scale dataset, termed maximal information coefficient (MIC) [173]. The idea is that if a relationship exists between two variables, then a grid can be drawn on the scatter plot to partition the data to encapsulate the relationship. Specifically, to compute MIC, all grids up to a maximal resolution (subject to a predefined threshold) are explored and the largest MI for each grid is selected and normalized to the corresponding grid dimension. Then, MIC is the maximum of these selected values [173]. Experimental evaluations over various functional relationships have shown that MIC provides a more effective way of detecting non-linear correlations, such as cubic, exponential, and sinusoidal, compared to Pearson coefficient. In manufacturing, MI has been widely considered as an important criterion in feature selection, for which a threshold is typically set up to determine whether additional feature provides substantially more information than features already selected [126, 216].

3.3.3. Topological data analysis

Data analysis often takes advantage of certain aspects of data structure, e.g., SVM finds a hyperplane that separates two clusters. Understanding data structure provides potential insights into data patterns. However, with big data, visualizing data structures to facilitate data analysis has become more difficult.

There are two ways of describing data structure: geometry and topology [59]. Geometry is focused on the metrics (e.g., distance) to determine the relationship. Topology concerns locality (e.g., whether points remain nearby) and therefore provides a means of analyzing data at a more refined level [50, 59]. Among various topological data analysis (TDA) methods, Mapper [191], a local clustering algorithm, has garnered considerable attention. Mapper features a filter function that guides clustering of high-dimensional data (e.g., large number of process variables) with three steps: (1) divide filter range (output range of filter function) into overlapping intervals, (2) cluster data per intervals, and (3) link clusters that have shared points [191]. This procedure is illustrated in Fig. 17 with a topology of a 1-D simplicial complex comprising vertices (0-simplex) and edges (1-simplex). This topology visualizes how large-scale data are organized and the resolution of the topology can be adjusted by changing the number of intervals [59].

TDA has found applications in many biomedical works [131], such as identification of diabetes subtypes [123] and pulmonary conditions [177]. The research in manufacturing is still limited. One of the reported works concerns variable selection for yield prediction in chemical plants [62, 63]. The concept is to first

recognize the structures that encode yield patterns. Then, by tuning the resolution, fundamental subgroups can be identified if the pattern persists over large changes, and statistical tests can be performed to identify the variables that best differentiate these subgroups. In this work, multidimensional scaling (MDS) is chosen as a filter function because it provides the smoothest variations in the yield. Each output dimension contains 14 intervals with 80% overlap, leading to 196 intervals in filter range.

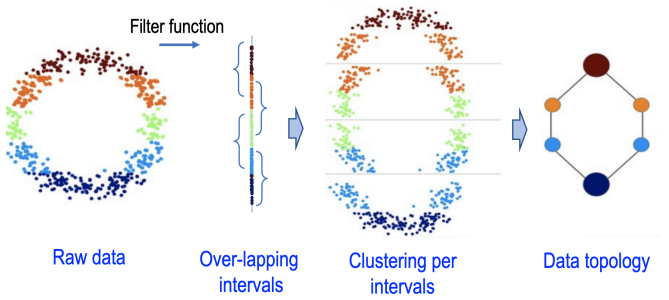

Fig. 17. Illustration of TDA steps [59]

The topological graph is shown in Fig. 18 (a). Each node is colour-coded by normalized mean yield for that node. High/low yield subgroups are isolated, and a Kolmogorov-Smirnov (KS) test is performed to select influential variables. One of the selected variables is visualized in Fig. 18 (b) for which the difference between subgroups B and D is significant. In total, 11 of 45 variables are selected as influential and are shown to have achieved comparable yield prediction accuracy as the case of using all 45 variables with 75% reduction in computational time.

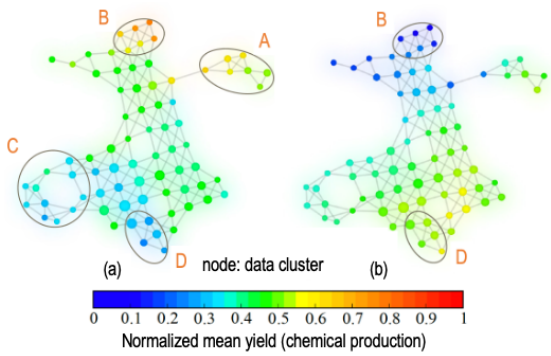

Fig. 18. Topological graph coloured by yield data (a) and an influential process variable (b) [62]

4. Data Learning

Data processing infers process quality via the underlying statistical model (e.g., distribution) that the collected data are expected to follow. While it has the advantage of building models through domain knowledge without being limited by data availability, it is challenged by the increasing heterogeneity and complexity in data [73]. Machine learning represents a shift from the statistical methods by allowing task-specific data pattern to be discovered via a set of representative training data, without relying on the assumption about the data [234]. ML can be broadly categoried into four categories as shown in Fig. 19. Recently, deep learning, which is a subset of ML has emerged as a powerful tool in data learning, It takes advantage of advanced computational infrastructure to optimize neural netowrks specifically designed for handling complexity embedded in big data (e.g., images) and of the increased data availability, providing rich training samples to better represent the individuality of each task [117]. DL has not only advanced the state-of-the-art in various common applications in manufacturing, such as condition monitoring, fault diagnosis and remaining useful life (RUL) prognosis [217, 250], but also opened up various new research opportunities, such as reinforcement leanring and transfer learning [163, 199].

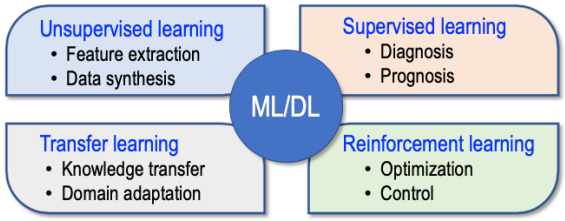

Fig. 19. Categories of ML/DL

4.1. From machine learning to deep learning

In manufacturing applications, both ML and DL have been studied to relate data patterns to improving product quality. The key difference lies in the manner of feature-based representation, either extracted manually in ML or automatically in DL [217]. ML techniques rely predominantly on the empirical knowledge that humans have acquired about the machines and processes, and become increasingly limited to data processing associated with modern manufacturing systems, due to the increasing complexity [250]. In comparison, DL has shown to hierarchically extract and decompose complex features into manageable levels for automatic and accurate data representation. This section starts with an overview of ML and DL, followed by unsupervised and supervised learning, the two most common types of learning.

4.1.1. Representative machine learning techniques

ML can be exemplied as a computer program that learns from experience (e.g., training data) with respect to certain tasks and improves its performance with experience [140]. The essence of ML is to transform a task into an optimization process applied to the corresponding (training) data. Therefore, substantial development of ML came as a result of advances in optimization methods such as quadratic programming (QP), classification and regression tree (CART), and backpropagation [71]. Using machine fault diagnosis as an exemple, decision tree formulates the task as a sequential decision-making process, for which decision thresholds are set for data attributes at each step (e.g., temperature > or ≤ 30°C). The objective is to find a series of threholds using CART that minimize the diagnostic error [94]. RF further extends the decision tree by utilizing ensemble of trees. It allows each tree to explore only a portion of the data attributes and then averages the results to improve model robustness [233]. In contrast, SVM formulates the task as finding a hyperplane that maximizes the margin of seperation between the data points representing different machine fault types using QP [216]. Artificial Neural Network, on the other hand, formulates a task a mapping from the data to conditional probabilities of the data belonging to different fault types. The objective is to adjust network weights using backpropagation, such that the probability of data belonging to the correct fault type is maximized [182].

4.1.2. Deep learning

While ML techniques work well on structured data with clearly defined data attributes, they become ineffective in processing unstructured data that are increasingly prevalent, such as images [117]. For years, tackling these types of data requires design of a suitable data representation based on domain expertise. Data representation refers to the transformation of raw data by data learning techniques (after they have been collected and processed as discussed in Sections 2 and 3, respectively) into a form that can be associated with relevant manufacturing tasks, e.g., a feature

vector. However, traditional ML tehcniques is limited in finding suitable data representation due to data complexity [117]. For example, many techniques have been developed to extract image features for surface defect recognition, such as edges [69], shapes [35], and peaks [11]. However, relying solely on these low-level features has shown to be insufficient since they are shared by various defects. Therefore, higher-level representations are required [117], which is at the core of DL.

DL refers to a series of neural networks consisting of multiple layers, which allow the decomposition of complex data into multiple levels and the assembly of multi-level features layer-by-layer into a high-level representation. DL can be considered an extension to ANN with various structrures specifically designed for unstructured data types, such as image, as shown in Table 3. For example, Convolutional neural networks (CNN) consists of a series of convolutional layers for image analysis, as shown in Fig. 20 (a) [109]. Neurons in each layer are connected to the local regions of the preceding layer through a set of weights called kernels. Local features are extracted through convolution at lower layers. The series of layers assembles local features into high-level features for image characterization [117]. Recurrent neural networks (RNN) and its variants – long short-term memory (LSTM) and gated recurrent units (GRU) – are suitable for analyzing sequence with each layer consisting of a series of cells. Each cell corresponds to a sequence step (e.g., a snapshot of machine state). Each layer is highlighted by the mechanism of maintaining past information, allowing the relationship among steps (e.g., degradation pattern) to be explicitly analyzed [84] as shown in Fig. 20 (b). The key aspect of DL is that the representation and feature are not designed manually by a human as compared to conventional ML. Instead, they are learned from data through backpropagation [117].

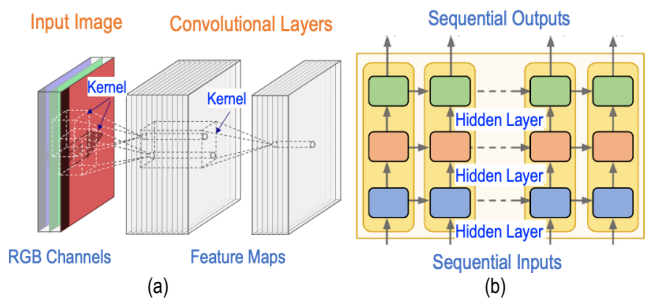

Fig. 20. (a) Convolutional layers, (b) Recurrent layers [58]

4.1.3. Unsupervised learning

Sensing data in manufacturing can be highly complex due to the manifestation of many interactions. Thus, it is essential that a learning method extract the most relevant information. This section hghlights the research of DL-based feature extraction followed by its application in data synthesis. Both tasks are "unsupervised" since there is no supervision as to what features and synthesised data will be used for.

4.1.3.1 Feature extraction

Intuitively, DL allows feature extraction by projecting data into a network with progressively reduced layer dimensions before reconstrucing the data themselves. As a result, the network is forced to find the most essential information and discard the rest. A widely-used realization of this concept is an Auto-encoder (AE) [250]. Several variants of AE exist. The most common variant is the sparse AE (SAE), which limits the number of non-zero elements in features to further improve discriminability [160]. The effectiveness of AE has been confirmed in [33] by comparing the features extracted from AE and SAE for bearing diagnosis.

In contrast, DBN extracts features by training RBMs on a layer-by-layer basis [224]. Shao *et al.* demonstrated the effectiveness of RBMs in obtaining features with progressively improved discriminability from vibration signal of induction motors with six different health conditions, as shown in Fig. 21 [186].

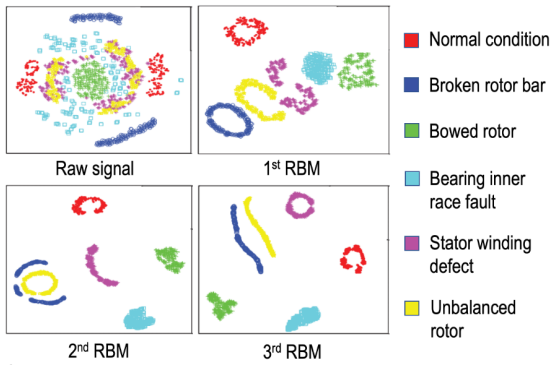

Fig. 21. t-SNE of features at different levels of RBMs [186]

4.1.3.2. Data synthesis

For certain tasks such as fault diagnosis, balanced data of various faults of interest is highly desired since unbalanced data can lead to serious learning bias [250]. However, balanced data collection is not always feasible since data related to faults is often limited [107]. The unbalance issue can be alleviated via data synthesis. However, tranditional methods (e.g., Synthetic Minority Over-sampling Technique or SMOTE) often rely on interpolation and cannot capture complex data characteristics [107]. A major breakthrough came with generative adversarial networks (GAN) [60], a DL method that is able to learn salient features and synthesize data with high fidelity. The idea of GAN is a competition between two components: a generator analyses real data to produce synthetic ones, and a discriminator distinguishes the synthetic data from the real ones. These two are trained alternately to improve the capability of each, and the final result is a equilibrium state, as shown in Fig. 22.

Implementation of GAN in manufacturing has been reported in the literature. Lee *et al.* compared the effectiveness of GAN and SMOTE in synthesizing data related to faults for motor fault

Table 3. Data learning for manufacturing process monitoring and prediction

Data type	Data form	Typical source	Main functional characteristics	Deep architecture	Typical task	Reference
Unstructured	2-D Image	Vision system	Shift/scale invariant to local features	Convolutional neural network (CNN)	Process & machine anomaly diagnosis/human activity recognition	[24, 65, 93, 230, 235, 240, 247]
		Time-frequency image				[40, 221, 250]
	Periodic 1-D sequence	Rotary device sensing				
	Non-periodic 1-D sequence	Machine health monitoring	Modeling relationship among sequence points	Recurrent neural network (RNN)	Trend prediction	[245, 248]
		Sequential process			Sequential process modeling	[146, 244]
Structured	Data table	Machine/process/ parameters	Handling many variables	Fully-connected network (ANN/DBN)	Non-sequential process modeling	[133, 224]

diagnosis [119]. It is shown that GAN consistently outperforms SMOTE in terms of fault classification accuracy across different unbalanced ratios. Similar work is reported in [187], in which the quality of synthetic faulted motor data has been confirmed through statistical analysis. Wang *et al*. investigated synthetic vibration signals for various gearbox faults. Spectral analysis showed that the synthetic data effectively captured important features, such as the characteristic frequencies [228].

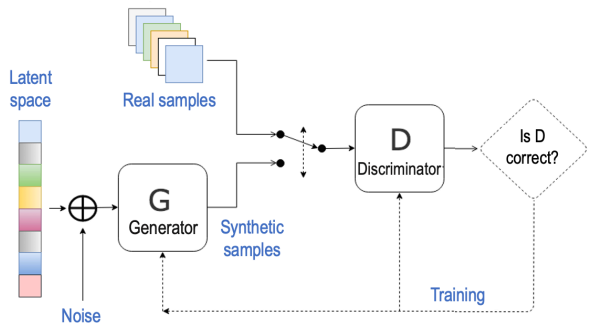

Fig. 22. Generative adversarial network, adapted from [60]

4.1.4. Supervised learning

Contrary to unsupervised learning where data transformation is carried out without supervision from specific manufacturing tasks, supervised learning establishes data transformation with the guidance (or supervision) from predefined data labels (e.g., process condition) associated with specific manufacturing tasks (e.g., process condition identification). With high flexibilities provided by DL architectures, supervised learning has seen considerable development in recent years. In particular, there has been an increasing effort in transforming data into suitable forms to better leverage the DL architectures and in customizing the DL architectures to better fit the manufacturing-related tasks. This section presents recent research efforts in CNN and RNN, which are two dominant methods in DL-based supervised learning.

4.1.4.1 Convolutional neural network

One of the main characteristics of big data in manufacturing is the increasing use of image data, which has significantly advanced the research field of in-situ process monitoring and made it one of the major research trends of big data analytics. Image data may contain information that are otherwise not captured by one-dimensional data (such as time series), thereby improving the observability of the object or process being monitored. From the published literature, CNN has shown to be the most popular and effective tool in extracting embedded information from image data. As an example, Weimer et al. reported on surface defect classification based on deep CNN (DCNN) that outperforms the traditional feature-based methods [230].

To account for the characteristics of various defect types embedded in different background textures, 1.3 million images were used for training. Dropout and l_2 regularization have been the two popular techniques to prevent "overfitting". Overfitting refers to the phenomenon that the perforamnce of the neural network, e.g., accuracy, at the training stage far exceeds that at testing stage. This is mainly due to the fact that model parameters (e.g., network weights) are over-sensitive to small variations in the training data (e.g., due to noise) and consequently, cannot effectively capture the underlying data pattern. The concept of dropout is to randomly drop neurons to "average out" noise effects. The technique of l_2 regularization prevents the weights from having excessive adjustment during the network training stage and therefore, making them less sensitive to data variations. Zhang *et al*. demonstrated an effective vision system based on DCNN for process deviation detection in selective laser melting (SLM) [247]. Process deviation affects powder melting and solidification, which is reflected as changing surface textures and recognized by DCNN. In [235], a vision system was developed to provide quality inspection at critical locations for automotive windshield glass priming. The work is highlighted by deploying DL in *edge* devices, which mitigated latency and facilitated real-time decision making [235]. Similar work can be found in [24, 65, 240]. DCNN has also shown to be effective in analyzing thermal images, which reflect the heat distribution associated with structural defect or improper maintenance (e.g., heat concentration due to improper lubrication) that is not captured by RGB cameras. Janssens *et al.* reported success by DCNN in detecting bearing anomalies such as outer raceway fault, particle contamination, and lubrication with high accuracy (91% ~ 95%), based on thermal image analysis [93]. Furthermore, the authors identified regions in the thermal images related to the specific anomalies by reversing the convolution operations to link to physical insight.

Beyond images captured by vision systems, DCNN has also been increasingly extended to the analysis of time series data. Wang et al. has built a condition monitoring system to quantify the severity level of gearbox faults, whereby vibration signals were first transformed into time-frequency images by means of wavelet transform, in order to leverage the DCNN architecture to extract fault-related patterns [221]. Similarly, in [40], the authors reported a method for both bearing fault type and severity level recognition using vibration signals, which were first converted to images using wavelet packet transform, before analyzed by DCNN for fault pattern recognition. Three fault types and four severity levels were identified using the developed method, and high accuracy has been achieved.

4.1.4.2 Recurrent neural network

Sequential data in manufacturing are commonly associated with machine degradation, which suits RNN and its variants by design. Several works on RNN-based machine health prediction have been reported. Zhang *et al*. developed a bi-directional LSTM for aircraft engine remaining useful life (RUL) estimation. To account for noisy measurements and engine heterogeneity, more than 30,000 sequences extracted from historical engine data were used for training. The bi-directional LSTM allows information to flow forward for prediction and backward for disturbance smoothing and has been shown to improve RUL prediction accuracy compared to uni-directional LSTM [245]. In [248], improved RUL estimation for lithium-ion battery has been achieved using ensemble LSTM, which allows estimates to be a probability distribution rather than a deterministic value.

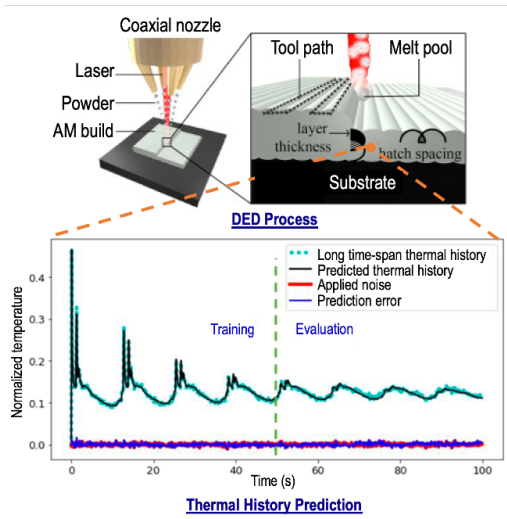

Fig. 23. DED thermal history prediction, adapted from [146]

More recently, as AM starts its transition into mainstream, RNN and its variants have been configured to model the sequential printing process by taking into account inter-layer effects. In [244], an LSTM-based model has been shown to improve part tensile strength prediction accuracy for fused deposition modelling (FDM). Specifically, each LSTM cell takes layer-wise sensing signals from IR sensor, thermocouple, and accelerometer as an input to represent each printing layer. Then, LSTM forward path models the inter-layer effects for joint prediction of part tensile strength. Layer-wise thermal history prediction in directed energy deposition (DED) has been reported in [146] with a GRU-based model. Over 250,000 training points were generated by FEA. Evaluation has shown that the developed model can accurately predict thermal history in DED (see Fig. 23).

4.2. Reinforcement learning

Dynamic optimization requires sequential process adjustment under specific conditions at each time step to optimize the final outcome. Manufacturing examples include shop floor scheduling, fixed-horizon process optimization, and robotic control. The learning technique for this task is reinforcement learning (RL), which aims at finding the optimal decision in dynamic settings through interactions with the environment [199, 210]. It is inspired by how a human worker learns to master complex tasks through a series of feedback loops between perception and action without a rigid rulebook. This section first presents the basics of RL with its recent development in scheduling and process optimization. Then, one of the most active research fields in RL, deep reinforcement learning (DRL), is presented under the context of robotic planning.

4.2.1. Dynamic programming

The framework of RL is shown in Fig. 24. The process can be summarized as an agent (e.g., order dispatcher, robot) interacts with an environment (e.g., shop floor, object cluster) and progressively learns to perform action (e.g., assign job, execute certain motion command) that maximizes long-term reward (e.g., minimizing job completion time, achieving object grasping), upon observing a state input (an instance of the environment).

RL distinguishes itself from unsupervised and supervised learning in two aspects: (1) data (e.g., state) are *self-generated* from previous interaction, and (2) the reward quantifies the *long-term* quality of state or action. Formally, it is quantified via state value function $v_\pi(s,a)$ or action value function $q_\pi(s,a)$, where s is the state, a is the action, and π represents the policy, or decision, for a given state [199]. RL is potentially data-intensive as it does not rely on the rules to manually limit the search space as in rule-based method. Therefore, the initial state space in RL is often large [199]. For example, if 10 variables are used to describe the state (e.g. ,10 machines), and each variable has 10 discrete values (e.g., utilization level), the initial number of states for RL to explore is 10^{10}. On the contrary, certain combinations of the state variables may be deemed by the rules as infeasible and therefore, excluded.

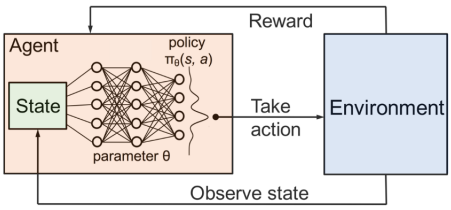

Fig. 24. Framework of reinforcement learning, adapted from [199]

If the environment dynamics, or model, is known, the optimal policy π_* can be found using dynamic programming (DP). The intuition is that any optimal value function can be obtained by collectively maximizing the reward from an immediate action and the value function from the successor state. π_* can then be determined from the optimal value function [199].

4.2.2. Model scheme generalization

When the environment model is unavailable as in most production applications, the realistic goal is to sample sequences of interactions to infer dynamics. One of the main model-free techniques is Q-learning, which iteratively updates action-value function based on sampled, immediate reward, and the maximum action-value function from the subsequent state. Model-free RL has been investigated in manufacturing for fixed-horizon process and shop floor scheduling.

Dornheim *et al.* developed Q-learning based optimal control for deep drawing where processing of a single workpiece (one episode) involves a constant number of discrete control steps (Fig. 25) [43]. The state is represented by stamp force, blank infeed, and blank-holder offset. The action determines blank holder force. The action-value function is approximated by ANN, and the reward is determined by internal stress, wall thickness, and material usage. RL is carried out in an FEA simulator, ensuring effective evaluation of the reward. Furthermore, process conditions are varied stochastically to simulate the real environment. It has been shown that RL-based methods outperformed baseline methods after 200 episodes of training. Stricker *et al.* reported an order dispatching system using Q-learning to maximize machine utilization with value function modeled by ANN [196]. The state is the number of orders, waiting time, and machine state. The action assigns order using ϵ-greedy criterion, allowing agent to "explore" random options with probability of ϵ to reduce the chance of settling in local minimum [199]. Evaluation has shown that 90% machine utilization rate is achieved, a 10% improvement over the rule-based method. Similar work on RL-based scheduling/planning can be found in [141, 142, 229].

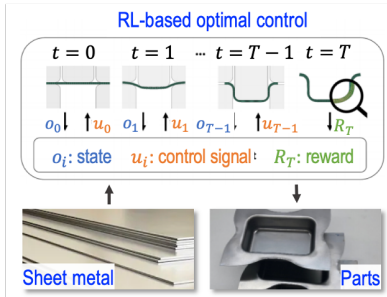

Fig. 25. RL-based control for deep drawing, adapted from [43]

4.2.3. Deep reinforcement learning

RL has been regarded as an effective approach to breaking the traditional rigid control scheme and improving the capability of robotics in dynamic situations. However, conventional model-free RL is limited in terms of the complexity of action-value functions that can be effectively approximated. The problem is further exacerbated in robotics where complex, multi-modal sensing data is essential to capturing both the environment and the interactions between the robot and human worker. Recent development in DRL, which is the combination of DL and RL, has shown great potential in achieving this objective due to the capability of DL in analyzing large volume and variety of sensor data and approximating complex action-value functions for robotic control.

Levine *et al.* presented a novel method for robotic object grasp learning [122]. The method consists of two components: a DCNN-based grasp predictor that analyses workspace images to determine the probability of a motion producing a successful grasp, and a servo mechanism that selects the motion to maximize the probability in the predictor, similar to action in Q-learning. The

reward is 1 for a successful grasp, which updates the network parameters in the direction to yield higher probability of success under similar input and 0 otherwise. Starting from random motion, after over 800,000 grasp attempts equivalent to millions of image inputs, the robot is able to intelligently handle various unseen object clusters [122]. In [242], a robot is shown to be able to learn more complex, human-like skills such as tossing. The authors leveraged physics to first compute analytical solutions of object release location and velocity. Then, a fully-convolutional network (FCN, a variant of CNN) is established to predict "residual" velocity to account for factors that can alter the trajectory. This learning process can be considered as two RLs with sequence of size 1, as picking up and throwing each requires only 1 control command in this research as compared to multiple motion adjustments designed in [122]. The reward is determined by residual velocity prediction error. After over 10,000 attempts, the robot exceeded a human in tossing accuracy [242].

4.3. Transfer learning

Intensive research on DL has produced a large number of solutions that are highly task-specific. Most of them were built from scratch, using large-scale datasets and intensive computation. Recently, researchers have started to investigate the possibility of knowledge transfer across different tasks to reduce the effort of building new solutions and maximize the value of data that have already been generated. The relevant technique is termed transfer learning, which is the process of optimizing the task performance in the target domain by using the knowledge transferred from the task performed in the source domain [163]. This section introduces this technique and related results.

4.3.1. Semi-supervised learning

Semi-supervised learning bridges unsupervised feature extraction and supervised tasks with the motivation that features extracted through unsupervised learning over a large dataset can be translated to various supervised tasks. This establishes the theoretical foundation for realizing transfer learning.

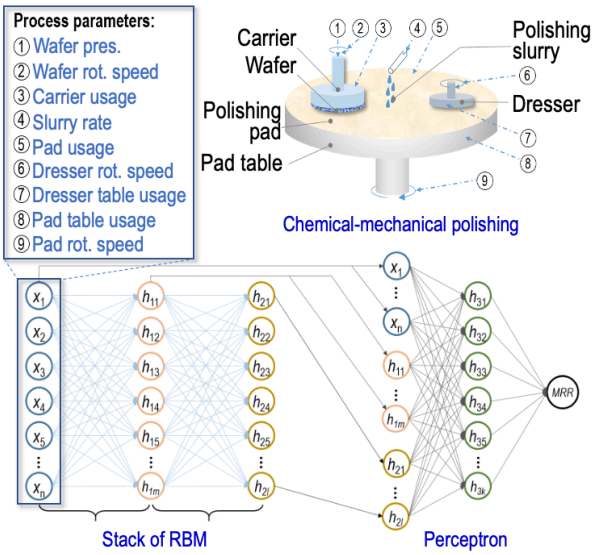

Fig. 26. DBN for MRR prediction, adapted from [224]

Sun *et al.* investigated semi-supervised learning involving a sparse AE for extracting current signal features and a multi-layer perceptron (MLP) for motor fault classification [198]. In the unsupervised stage, partial signal corruption was performed to improve the feature robustness. Good classification accuracy was reported. Wang *et al.* investigated semi-supervised learning for predicting material removal rate (MRR) in chemical-mechanical polishing of wafers [224]. Specifically, a stack of RBMs [83] were constructed for feature extraction using historical data from over 2,000 wafers, and a perceptron carried out supervised MRR prediction (Fig. 26). It was shown that the developed method improved MRR prediction accuracy over other ML techniques. The combination of RBM and perceptron was also investigated for bearing degradation phase recognition in [133].

4.3.2. Domain adaptation

Domain adaptation differs from semi-supervised learning in that it aims at generalizing well-performing models learned from a source domain to a target domain (in which the learning task shares similarity with the source domain) rather than transferring the generated features. In an unsupervised framework, Sun *et al.* proposed a method to transfer an SAE across machine tools [197]. Three strategies were developed to enhance the transferability: (1) weights learned in source domain SAE are first copied to target domain SAE, (2) Kullback–Leibler divergence between the activation from the two SAEs is used as a constraint for feature transfer, and gradient of the weights in target domain SAE are obtained by minimizing the divergence, and (3) using the gradients, layer-wise weights update in target domain SAE is performed and generate final features. The framework of the method is illustrated in Fig. 27.

The most common research direction of domain adaptation is in supervised framework. The procedure is similar, involving network structure transfer and target domain weights fine-tuning. Hasan and Kim demonstrated the effectiveness of such method for bearing fault diagnosis under different working conditions [70]. Specifically, four load conditions, from 1 to 4 HP, and six bearing health conditions are evaluated. First, a source DCNN is trained on vibration data from one of the conditions (i.e., source domain). Then, final DCNN layer is fine-tuned with data from the other three conditions (i.e., target domains). It has been shown that the transferred DCNN outperformed the model trained only on target domain data based on diagnostic accuracy and reduced training time by half. Similar results were reported in [246].

Motivated by the observation that features generated from early layers of deep network are very generic as compared to those generated later, researchers have recently started to investigate the off-the-shelf deep network as source model, such as AlexNet and VGG-16, to take advantage of their feature generation capability achieved by training over massive datasets (e.g., over 10 million images). Although these datasets are not related to manufacturing, state-of-the-art results in domain adaptation nevertheless have been achieved with only a small amount of target domain data required for fine-tuning for bearing fault diagnosis [29], surface defect classification [102], human motion recognition [128, 225], and part identification [110]. These results suggest potential deeper connections among features learned across completely different domains, which constitue a future research direction.

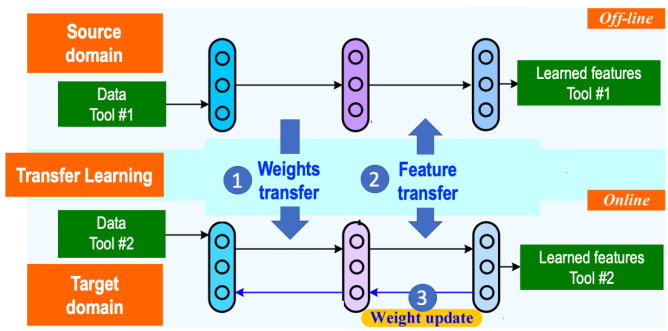

Fig. 27. Framework of transfer of SAE-based model, adapted from [197]

5. Data Security

The increasing collection and use of big data on the factory floor have introduced risks that have made data security a critical need [14, 46, 55, 231]. Historically, industry has been hesitant to exploit data from manufacturing systems due to the perceived threat of cyberattacks. Well-publicized events, such as Stuxnet, and more traditional attacks, such as phishing, have created a strong sense of unease, especially when coupled with growing government regulations on the distribution and use of data. Information technology (IT) professionals share even greater concern as the age, obsolescence, and variety of operating systems and control technologies that are central to manufacturing have created a large number of attack surfaces that may be exploited by bad actors [80]. Thus, data security is essential to enable the technologies and solutions that have been presented in Sections 2-4. The goal of this section is to provide manufacturers with the foundation needed to understand all security considerations and risks so that the right combination of technologies can be deployed.

5.1. Relevant perspectives

The primary focus of data security in manufacturing has been the protection of sensitive information (e.g., intellectual property, customer information) and the security of networked devices [55, 231]. While these areas remain critical, it is important to note other considerations that arise from the use of big data, such as the trust that must be established to ensure the quality and value of subsequent data analysis [77, 78, 108].

5.1.1. Data management perspective

The traditional perspective on data security can be described using the "CIA triad" as shown in Fig. 28 [4]:

- *Confidentiality (C)* reflects the ability to limit access to data to authorized users
- *Integrity (I)* reflects the ability to ensure the accuracy, authenticity, and completeness of data
- *Availability (A)* reflects the ability to make data available to authorized users on demand

These principles can also be considered by their negation: disclosure, alteration, and denial. By mapping these considerations onto the general process of transferring a part design to a manufacturing supplier, security can be achieved if unauthorized disclosure of a part model is prevented, no part model is altered, and no part model is missing and thus denied to the supplier.

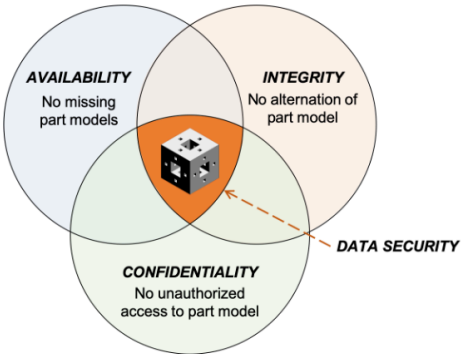

Fig. 28. CIA Triad mapped onto the process of transferring a part design to a manufacturing supplier, adapted from [4].

The considerations described by the CIA triad are critical in ensuring the general security of big data in smart factories. The impacts that may result from not enforcing these considerations go beyond the loss of critical information, including the loss of reputation, productivity, or even life [55; 231]. Traditional strategies to address these concerns in manufacturing include encryption (see Section 5.2.1) [46, 55, 231], network segmentation (e.g., firewalls), root authorization, secure boot, and virtual local area networks (VLAN) [55]. Many governmental policies and standards also exist to promote data security from the data management perspective, which will be the focus of Section 5.3.

5.1.2. Data use perspective

Beyond traditional data security considerations, the quality of any analysis depends entirely on the trustworthiness of data, which is based on three pillars that form the basis of the semantic web concept (see Fig. 29) [77, 78, 108]:

- *Authentication* reflects the ability to determine the integrity and provenance of data
- *Authorization* reflects the ability to determine permissions on the use of data
- *Traceability* reflects the ability to determine the history (e.g., generation, modification, and use) of data throughout lifecycle

The objective of authentication is to determine who did what to the data, e.g., the act of a designer acknowledging that the part model is his design. The objective of authorization is to determine what can be done to the data, e.g., the act of an authority approving the part model for production. The objective of traceability is to determine where or when some action was done to the data, e.g., the act of timestamping the part model based on when it was authorized for production. Collectively, these actions help establish trust in the part model such that a manufacturing supplier can determine if the part model has sufficient veracity to begin production. For example, if a relatively long period of time has passed since the part model was authorized for production, a supplier may decide that the risk of a bad actor altering the design is too great. Similarly, a supplier may not have trust in the part model if the approval mechanism (e.g., digital certificates) does not correctly correspond to the recognized authority.

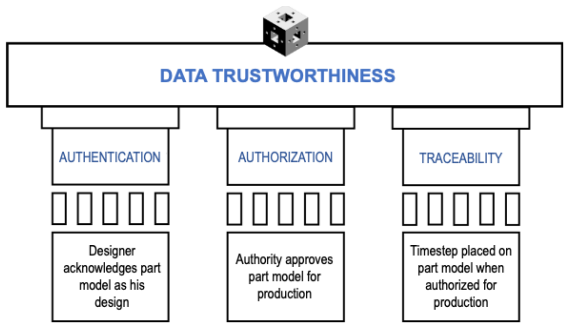

Fig. 29. Three pillars of trustworthiness mapped onto the process of transferring a part design to a manufacturing supplier

The area of data trustworthiness has been a growing topic of interest in the design and manufacturing community, especially in highly regulated sectors such as aerospace and biomedical industries where product traceability is an essential requirement [77, 78, 108]. The digital thread paradigm (i.e., linking design and manufacturing systems across the product lifecycle) has driven much of this research, which has focused on technologies such as data certification [77, 78] and blockchain [108]. AM has also motivated much of this research because it relies on digital data and deviates in workflow relative to subtractive processes. Since much of the state-of-the-art has been driven by highly regulated industries, several areas of standards development exist that target data trustworthiness, especially traceability, which is the focus of Section 5.3.

5.1.3. Risk management

It is essential that a risk management approach be followed when assessing data security requirements [14]. Risk management involves systematically identifying, assessing, and addressing risks. Typical data security practice in many organizations is to apply blanket policies to ensure no breach or attack, but such an approach tends to limit the ability to generate value from big data without substantially improving safety. For example, the majority of data breaches typically occur within a company even when that company uses cloud-based platforms to manage data, which limits the effectiveness of tools such as encryption since insiders may have the requisite keys to decipher data [46]. Thus, the confidentiality risk typically associated with a cloud-based platform may be overblown relative to the value of the data being managed, and instead of focusing on higher levels of confidentiality, it is more important to have a well-designed key management system to avoid recognized risks.

Numerous tools exist to support industry through the risk management process, such as root cause analysis, Pareto diagrams, and probabilistic risk assessments [14]. Several industry-based, consensus-driven frameworks have also been created to help a variety of organizations identify and respond to risks effectively given their priorities and goals. A well-known framework is the NIST Cybersecurity Framework (see Section 5.3.3) [155].

5.2. Practical examples

Various data security technologies and standards relevant to the use of big data are available. This section presents the basics of three technologies that have been deployed or are of strong interest to manufacturing. Representative textbooks such as that by Andress [4] and cybersecurity frameworks serve as references for manufacturers to explore solutions from the state-of-the-art.

5.2.1. Cryptographic hashing and encryption

Hashing is a fundamental element of modern cryptography that uses an algorithm to transform (or "cipher") data of an arbitrary length (or "plaintext") into a typically fixed and shortened message (or "ciphertext") without the possibility of retrieving the original data [4]. It is an essential component of various cybersecurity technologies, such as digital passwords, digital certificates, and blockchain. A common use of hashing is when different datasets need to be compared but storing and/or sharing the datasets as plaintext presents a security risk, e.g., passwords (see Fig. 30). In this case, hashing can be used to cipher the plaintext while still enabling a comparison between the datasets. Similarly, hashing can be useful when a dataset is large (e.g., images or music) and would be better shortened for comparison. Another good use of hashing is when the validity of dataset needs to be assessed, and assessment can be conducted after hashing the original data so that it will not be disclosed. For many manufacturing use cases, hashing is sufficient to ensure data integrity.

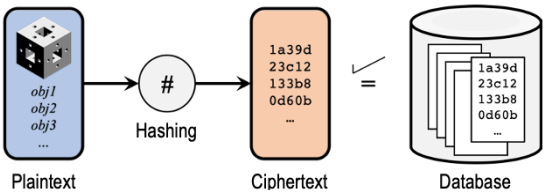

Fig. 30. Example of hashing used to compare two datasets without disclosing the data itself

Encryption is another subfield of cryptography that uses an algorithm to cipher plaintext into ciphertext with one or more "keys" [4]. The key difference between hashing and encryption is that encryption provides a means of reversing the cipher so that the original data may be retrieved. Encryption is a fundamental technology that enables the sharing of data with a reduced risk of disclosure. The most common implementation of encryption is asymmetric encryption (see Fig. 31), which relies on pairs of public and private keys.

A public key can be shared to encrypt data, but encrypted data cannot be read without a private key, which is generated per user and is the essential component to protect to ensure privacy. Many well-known applications rely on asymmetric encryption, including Secure Sockets Layer (SSL)/Transport Layer Security (TLS) and secure/multipurpose internet mail extensions (S/MIME) [77].

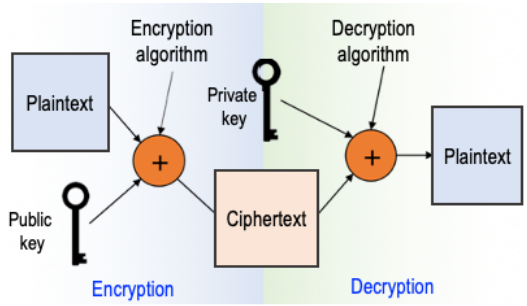

Fig. 31. Schematic representation of asymmetric encryption

Despite the power of encryption, it is not a panacea since any actor with an appropriate key can decrypt data. Therefore, key management is critical to successfully implement encryption [46]. Figure 32 provides a basic overview of the essential processes of a key management system.

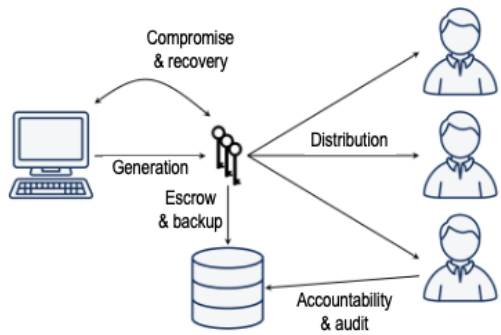

Fig. 32. Essential processes of key management, adapted from [46]

5.2.2. Digital certificates

A digital certificate (or public key certificate) is an electronic document that contains metadata about a public key and a "signature" from an authority that validates this metadata [4, 204]. The X.509 standard (ISO/IEC 9594-8:2014) provides the most common format for digital certificates, which are primarily used to establish trust in encryption schemes by identifying the owner of a public key thereby authenticating it [204].

A hierarchy of trust is essential for the implementation of digital certificates. It consists of hardware, software, people, policies, and procedures [108, 205]. Figure 33 provides an example of a hierarchy of trust proposed by Hedberg *et al.* [78] to apply digital certificates to support certification and traceability of product and manufacturing data. A root certificate authority empowers an audit service to sign certificates for various entities to provide different services. These entities can then be authorized to provide other services, such as validating the type of data being used (native or derivative) or the actual use of that data.

Extending the typical X.509 implementation with additional metadata provides authorization as well as traceability, which establishes firm trust in the data the certificate is attached to [77,

108]. Hedberg *et al*. described a Digital Manufacturing Certificate (DMC) Toolkit for manufacturers to incorporate digital certificates into four open-data formats for production: ISO 10303-242:2014 (STEP AP242), ISO 6983-1:2009 (G code), ANSI/DMSC Quality Information Framework, and the combined ISO 32000 (Portable Document Format or PDF) and ISO 14739 (Product Representation Compact or PRC) [77].

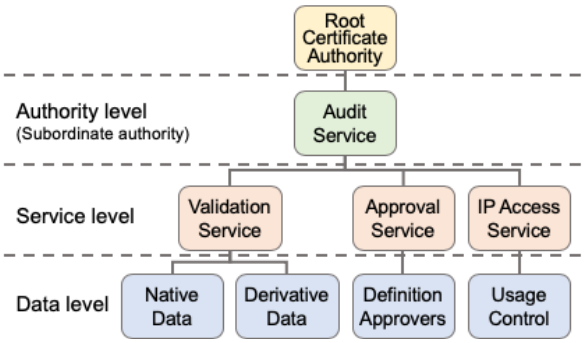

Fig. 33. Hierarchy of trust to enable digital certificates, adapted from [78]

5.2.3. Blockchain

Blockchain is a distributed ledger system (or database) that contains a growing list of records called "blocks" [237]. The distributed ledger is managed by anonymous peers adhering to a protocol that enables the verification of transactions without disclosing the participants of the transactions. The blockchain itself is designed to be resistant to manipulation, which establishes trust despite the anonymity of transactions. Figure 34 provides a basic example of how blockchain can be used to facilitate the secure transfer of product data (e.g., a part model) from a designer to a manufacturer, who can validate data integrity by using the blockchain to determine whether the data originated from the expected organization and whether the data transaction was completed without issue. To expand on this concept, a reference information model that establishes blockchain-based traceability for product and manufacturing data has been proposed by Hedberg *et al*. [78].

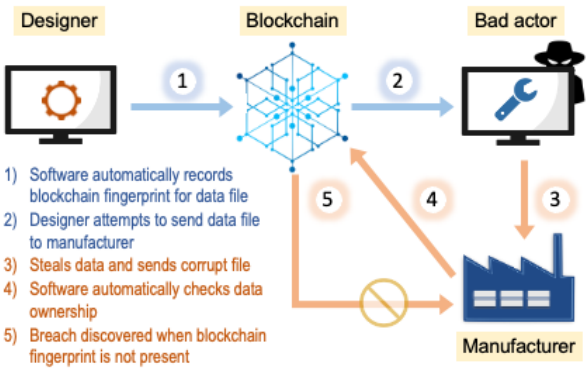

Fig. 34. Example of using blockchain to identify breach in transmission of data file to a manufactuer, adapted from [157]

5.3. Regulatory policies and standards

Regulatory policies and standards add further requirements on the collection and use of big data for smart factories and introduce additional challenges for manufacturers as they may vary by jurisdiction or industry sector. This section presents a basic understanding of the types of regulatory and standards considerations that may affect manufacturing with a focus on the US market. However, similar policies can be expected to exist in other jurisdictions.

5.3.1. Government regulations

A primary concern of government regulation on the collection and use of data is to limit the disclosure of sensitive information for reasons of national or economic security. In the US, such regulations are broadly referred to as "export controls" and are described by the Export Administration Regulations (EAR) [211]. "Export" refers to the transfer of technologies, including physical items, designs, software, or data and information, to a foreign national or entity either within or outside of the US. The EAR are not exhaustive since they do not apply to all services and technologies; e.g., defense services and technologies are regulated by the US Department of State via regulations such as the International Traffic in Arms Regulations (ITAR) [212].

Another aspect of government regulation on the collection and use of data in manufacturing is the traceability of product data in highly-regulated sectors such as aerospace, biomedical, and food and drug manufacturing [77, 78, 108]. Regulations exist in these sectors that require companies to keep records of the design and manufacture of relevant products, including information such as the specific tools and machines used for each manufacturing process performed on the product, the time when these process occurred, and the operator who managed the process.

Finally, government regulation is increasingly addressing privacy concerns when collecting and using data of any kind. For example, the European Union (EU) General Data Protection Regulation (GDPR) or EU Regulation 2016/679 was implemented in 2018 to protect the personal data of all citizens in the EU and European Economic Area (EEA) [48]. These types of regulations generally require the disclosure of any data collection, the lawful reasons for data collection, the length of time that collected data are to be retained, and the means by which collected data may be shared with third parties. Such regulations may affect the collection and use of data in manufacturing if the data include information on customers or shop-floor personnel. In Table 4, common regulatory considerations and their potential implication on manufacturers, with sample requiremnts, have been summarized.

5.3.2. Standards

Beyond regulatory policies, many standards also exist to harmonize different IT security technologies. These standards have been released by different parties, including government agencies and consensus-based standards development organizations (SDOs). Some examples include:

Table 4. Regulatory considerations and potential implication on manufacturers.

Regulatory consideration	Potential implication	Sample requirement
Export, re-export, or transfer of sensitive technologies	Limits collection, distribution, and use of data and information on products deemed sensitive for economic or security reasons	Export Administration Regulations (US); International Traffic in Arms Regulations (US)
Traceability of product data	Requires long-term data collection and management over the lifecycle of products in sectors such as aerospace, biomedical and food	Federal Aviation Administration (US); Food and Drug Administration (US)
Privacy of personally identifiable information	Requires disclosure, management and protection of data and information on customers or shop-floor personel	General Data Protection Regulation (EU)

- Federal Information Processing Standards (FIPS) [156]:

  FIPS publications are released by the National Institute of Standards and Technology (NIST) when required by statue or when a compelling need exists within the US Federal Government (although they may be used by private-sector organizations as desired). Two example FIPS publications relevant to encryption are FIPS 197 (Advanced Encryption Standard) [153] and FIPS 140-2 (Security Requirements for Cryptographic Modules) [152].

- NIST Special Publication 800-series (NIST SP 800) [154]:

  NIST SP 800 provides a series of guidelines, recommendations, technical specifications, and annual reports of NIST's cybersecurity activities. As with FIPS, NIST SP 800 publications are developed to address the security needs of the US Federal Government, but they may be used by non-government and private-sector organizations.

- ISO/IEC 27001:2013 (Information technology – Security techniques – Information security management systems – Requirements) [88]:

  Published jointly by the International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC), ISO/IEC 27001:2013 is part or the larger ISO/IEC 27000 family of standards focused on information security management systems. It provides the requirements for these types of management systems, which address the security of a variety of information, including financial information, IP, or employee and customer information.

Standards are generally voluntary unless mandated by statue. They may be contractually or otherwise required, which would necessitate compliance by a manufacturer who entered into such a business transaction. Otherwise, standards provide guidance that manufacturers may leverage to ensure that they follow appropriate security requirements and best practices.

5.3.3. Frameworks

Given the growing importance of data security, the variety of government regulation and standards on the topic, and the large number of security solutions on the market, it can be an enormous challenge for manufacturers to determine the correct approach and technologies needed for their specific production systems. To address this need and provide tools and guidance, several framework approaches have been developed to educate the broader community adopting data security practices.

One early example is the Common Weaknesses Enumeration (CWE) released by the MITRE Corporation [208]. The CWE provides a formal classification of weaknesses and security flaws collected and documented from software as well as distilled "lessons learned" and solutions to these flaws. Another important framework is the Cybersecurity Framework published by NIST [155]. The Cybersecurity Framework is voluntary and includes standards, guidelines, and best practices for managing cybersecurity-related risks. It provides a way for organizations to describe their current security posture and target state and then assess progress towards meeting its goals.

Figure 35 describes the NIST Cybersecurity Framework Core, which is composed of five essential functions and the different categories of topics relevant for each function. Included with the Cybersecurity Framework is NISTIR 8183A (Cybersecurity Framework Manufacturing Profile), which was developed specifically for manufacturers to be able to manage risk within discrete-based productions systems [195]. These tools enable manufacturers to develop plans and practices to ensure security when using big data for smart factories.

Fig. 35. Structure of the NIST Cybersecurity Framework Core, adapted from [151]

6. Case Studies

The ultimate goal of big data analytics is to have various technologies developed for different stages of data lifecycle successfully translated into realization of smart factory. To this end, three industry case studies are presented in this chapter.

6.1 Applying text-based data for decision making

Textual data (e.g., written logs) are one type of collected data that often go unused in manufacturing. This is especially true for maintenance where data, especially historical data, are often collected through maintenance work orders and service tickets [184]. While the data can be rich in historical knowledge, they can be difficult to analyze because the data are not computable (see Section 2.3.2). Textual data in manufacturing are often informal and unstructured and contains technical jargon, abbreviations, and misspellings that challenge the application of existing natural language processing (NLP) techniques due to data variety and veracity. These issues also often challenge humans who attempt to analyze the data to identify trends and best practices from previous activities. The result is that it usually takes more time to diagnose a problem in a manufacturing system than it does to resolve any problem that is found.

Enabling the analysis of textual data in manufacturing requires formulating consistent, reusable semantics around the data such that the data are structured with commonly understood meaning (i.e., reduce variety and improve veracity). Often, this process involves the manual application of "tags" (or annotations) to the data based on input from a domain expert, but this process can be time-consuming and costly [184]. Sexton *et al.* presented an alternative hybrid approach that augments NLP techniques with human guidance to decompose and tag the textual data in maintenance work orders [184, 185] (see Fig. 36).

This approach shifts the effort from manual tagging to the creation of a domain-specific dictionary that contains relevant terms and the knowledge that these terms represent. The dictionary can then be used with NLP techniques to significantly reduce the time and effort needed to make textual data computable: a case study with an industry partner of annotating 3438 raw text descriptions required over 18 hours manually versus one hour when applying the hybrid approach. Furthermore, the initial use of this hybrid approach can greatly simply future data collection efforts by identifying effective data tags that can be incorporated into text-based documents, such as maintenance work orders, so that these documents already contain computable data ready for analysis. An open-source toolkit – Nestor [159] – has been developed from this research and is currently being studied

by several manufacturers, including those in the automotive and aerospace sectors.

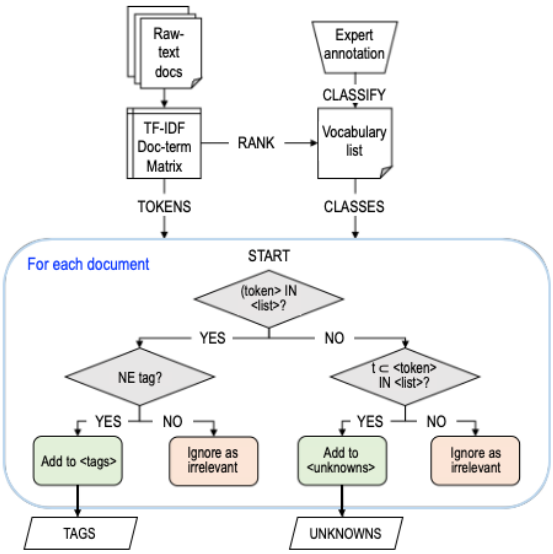

Fig. 36. Flow chart of hybrid annotating process [184]

6.2. Defect prevention-based job scheduling

Current shop floor scheduling methods mainly focus on the availability of machining resources when assessing the time and cost of task execution instead of potential defect during operation [141, 229]. However, defects in machines may result in significant task delay and production loss. This case study discusses defect prevention-based scheduling, which reassign high-risk tasks based on both the historical database and data collected from on-going production operations (see Fig. 37) [94].

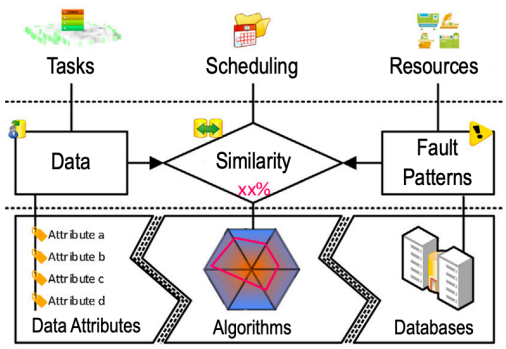

Fig. 37. Fault prediction enabled scheduling [94]

Data associated with a machining task may involve more than 40 attributes, containing information on machine tools, workpieces, machining processes, operation time, results, and operators, which all affect the production [94]. To cope with data complexity, a 3-level data management structure has been developed to learn defect-related patterns from local data, local network data, and cloud data (i.e., from machine-level to system-level). Local data analysis considers the patterns of a single machine tool and local network data analysis considers operation patterns of a class of machine tools. Cloud data analysis reveals the pattern at the shop floor level.

The established learned pattern is then used to compare with data from an incoming/on-going task to evaluate the similarity and risk probability, which provide the basis for task scheduling or rescheduling. A sample risk probability evaluation process is illustrated in Fig. 38, which shows that different scheduling of task #1 can lead to different risk probability, thereby providing guidance for final decision making. The algorithm is based on a decision tree in which each tree bifurcation represents a decision based on certain numbers of data attributes.

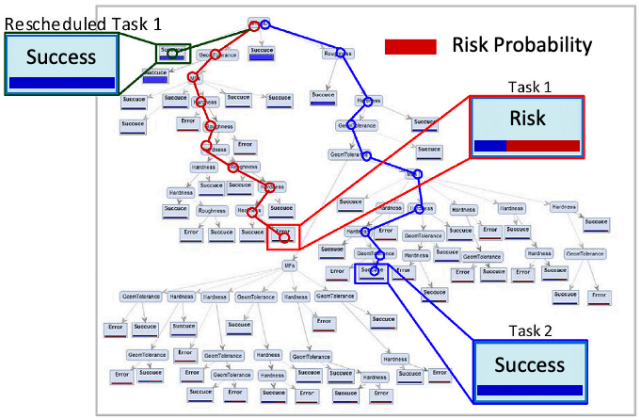

Fig. 38. Sample risk probability evalutation process for scheduling based on decision tree, adapted from [94]

6.3. Digital twin

The digital twin is an emerging concept that leverages data and information collected from a physical system to create a digital representation of that system that may be used to generate some desired control action (see Fig. 39) [194]. The growth of data collection in manufacturing has enabled the potential use of digital twins for variety of situations, including optimization of production system performance, prediction of maintenance-related faults and failures, and virtual verification and validation of production equipment. In this way, the digital twin concept can be one way for manufacturers to realize the following four key use cases for big data: (1) imaging, (2) prognosis, (3) maintenance, and (4) supply chains and assembly.

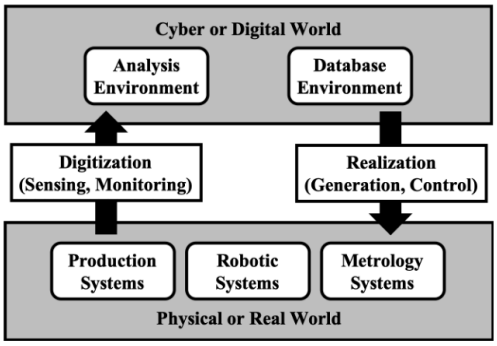

Fig. 39. Schematic a digital twin, adapted from [75]

One recent example of the implementation of digital twins was the Operate, Orchestrate, and Originate (O3) Project funded by MXD (formerly the Digital Manufacturing and Design Innovation Institute or DMDII) in the US [38]. This project was a collaboration between STEP Tools, International TechneGroup (ITI), Mitutoyo America, and VIMANA. The goal of the effort was to enable automated, real-time conformance validation of manufacturing processes by analyzing data from design, production, and inspection [194]. A prototype digital twin was developed by the project partners that combined several standards-based data pipelines, including ISO 10303-242 (Standard for the Exchange of Product Model Data or STEP AP242) from design, MTConnect from manufacturing, and Quality Information Framework (QIF) from inspection [68, 193]. Universally unique identifiers (UUIDs) were used to merge these data types and address the challenges presented by a variety of data types [193]. Prototype digital twins created through the O3 Project (see Fig. 40) can be explored online

and highlight two other significant challenges: (1) the need for potentially large data volumes to support sufficient analysis, and (2) the need to ensure veracity of the data and analysis to support effective decision making. If addressed successfully, the resulting digital twin can be used to adjust a non-conforming process.

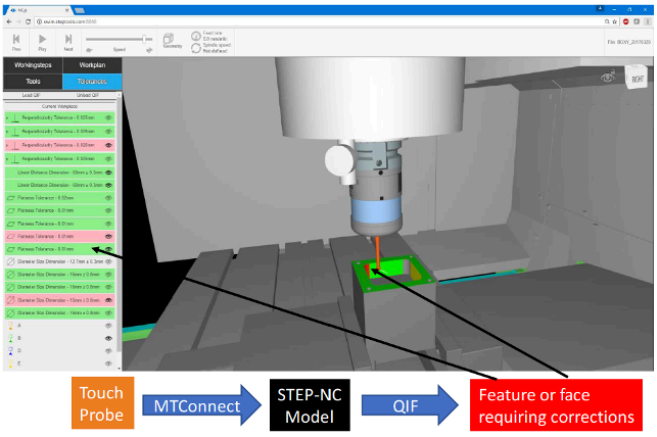

Fig. 40. Example of a visualization created from a digital twin that can be used for automate conformance validation of a machining process [69]

7. Future directions

Manufacturing continues to evolve toward optimization, as companies are increasingly able to capture data from various aspects of manufacturing processes and transform them into actionable insights [113, 218]. However, multiple gaps still exist that should be addressed to ensure that big data analytics are successfully leveraged for value-addedness in realizing smart factories of the future. Seven topics related to big data are summarized here as recommendations for future research.

7.1 Improving data quality

To maximize the value of big data, methods of data collection should be closely aligned and correlated with domain knowledge. Furthermore, there has been a growing realization that manufacturing lacks data of sufficient quality to identify and model causality. To address these needs, efforts should be directed to leveraging methods such as linked data, graph theory, and category theory, to connect different concepts inherent in data so that domain knowledge can inform and guide the data analysis process. It is also important to have sufficient semantics to ensure that the analysis of data provides value and can be reused for future analyses. Several techniques developed for data contextualization and semantic indexing have been discussed in Section 2, but more research efforts are needed for the democratization of these important techniques.

7.2 Scaling data collection

Data collection in manufacturing face two types of scalability challenges. Compared to areas where techniques exists for system scaling down, verification, validation, and conformance testing (e.g., wind tunnels), manufacturing lacks such tools. As a result, it is difficult to collect meaningful data to support research and many outcomes are only tangential to the real need of manufacturers [113]. Similarly, data collection under large scale manufacturing (e.g., 100s or more machines) has not truly been investigated, and such large-scale deployment can potentially involve additional issues such as heterogeneity of machines and systems under various industrial standards. Therefore, future research on scalable data collection is urgently needed to provide meaningful guidance to manufactuers. Furthermore, research efforts on data collection has been so far largely focused on machines and processes. Other aspects of manufacturing operations, such as assembly where data associated with fixturing, alignment, and tolerancing are generated, should also be considered to ensure scalability in data collection on the manufacturing shop floors.

7.3 Quantifying uncertainty

Although various research efforts have been reported in data uncertainty quantification as described in Section 3, it remains an open and critical topic for data analysis, especially given the variety of data sources involved. Continued research on uncertainty quantification that enables scalable and systematic knowledge translation for incorporation into subsequent data analysis, such as data learning, and ultimately, decision making, will be valuable and beneficial. Accordingly, more fundamental research is needed to develop standards and tools to address this need and advance the state of manufacturing.

7.4 Leveraging physics-guided data learning

Despite reported success of data learning methods such as DL, as described in Section 4, limitations continue to remain, including; (1) the "black-box" nature of the decision logic of DL, which are not transparently linked to mechanisms driving manufacturing phenomena, and (2) their critical dependence on the quality of the available training data that are prone to spurious relationships not consistent with physics and therefore not generalizable. Conversely, the large amount of physical knowledge accumulated in manufacturing has yet to be fully incorporated into data learning. Taking advantage of both methods will be beneficial to facilitating innovation that is more robust and reliable [171].

7.5 Controlling false discovery rate

Data learning algorithms, especially DL, facilitate the screening of potentially relevant process parameters (i.e., input variables) by inductively finding their correlations to the quality metrics (output variables), which can be further tested to verify their causal influence. However, a potential problem is the lack of methods in data learning to estimate and control false discovery rate (FDR) when generating correlations. This may result in significant waste of effort in the subsequent confirmatory study if the FDR is high [10, 89]. Only very limited work has been reported in this field so far. One example is knockoff filter [10]. The basic concept is to construct fake input variables designed to mimic the correlation structure found within the existing data in a way to allow accurate FDR control [10]. Initial tests of the method has shown promising results. These studies should motivate researchers in manufacturing to develop more efficient and reliable procedures for continued process improvement.

7.6 Generalizing analysis methods

Much of the efforts reported on big data analysis have been specific solutions rather than generalizable models and methods that can be broadly deployed across the industry. This means that models of machine performance will likely be re-trained for each different type of machines, even if the models are identical, due to the individual characteristics of machines and their operational history. Future research on generalizable big data analysis methods, such as transfer learning, will be beneficial for minimizing initial efforts in implementation across applications.

7.7 Ensuring trustworthiness of manufacturing data

While increasing attention has been placed on new technologies such as blockchain, there are a variety of potentially simpler solutions that should meet many manufacturing requirements, and research on these solutions should be of high relevance to avoid unnecessary cost and effort. The growth of semantics in data

analytics in manufacturing may also provide potential opportunities to leverage big data for data security. It is important for manufacturers to understand and explore the right combination of technologies to ensure the trustworthiness of the data under various manufacturing scenarios.

8. Conclusions

It has been envisioned by McKinsey that data will play a central role in smart manufacturing [138] with critical technologies involving (1) automated in-plant logistics, (2) data collection across supply chain, (3) data-driven predictive maintenance, (4) automation and human-machine collaboration, (5) digitalized quality system and process control, (6) digital performance management, and (7) smart planning and agile operations. To facilitate the realization of smart factories of the future, this paper summarized the state-of-the-art of big data analytics from the perspective of the lifecycle of manufacturing data, from collection, transmission, management, processing, to learning. Within each stage, specific challenges posed by the volume, velocity, variety, and veracity of data are highlighted, and respective solutions are synthesized. In addition, the issue of big data security is discussed from the perspective of both technology and policy. Three industry case studies are presented to demonstrate the value added by big data. Recognizing significant interest from both academia and industry on data and the many fields critical for realizing the full potential of big data that have yet to be explored, future research directions have been proposed and explained.

The landscape of big data is rapidly expanding, and new data analytic methods are increasingly reported in the literature. Leveraging these new technologies to advance the state of understanding of manufacturing processes and systems will allow manufacturers to benefit from the rich information embedded in the vast amount and type of data available in a sensor-rich environment. This area is quickly becoming a research topic of high relevance to manufacturing, and facilitates the digital transformation towards smart factories of the future.

Disclaimer

The identification of commercial systems does not imply recommendation or endorsement by NIST or that the products identified are necessarily the best available for the purpose.

Acknowledgment

Gao acknowledges support from the National Science Foundation under award CMMI-1830295. Authors are grateful to Mr. Jianjing Zhang, Department of Mechanical and Aerospace Engineering, Case Western Reserve University, and Dr. Peng Wang, Department of Electrical and Computer Engineering, University of Kentucky, for their valuable inputs and assistance.

References

[1] American Society of Mechanical Engineers, 2019, Model Organization Practices, ASME Y14.47.
[2] American Society of Mechanical Engineers Model-Based Enterprise Steering Group, 2018, Model-Based Enterprise Standards Committee: Recommendation Report.
[3] Anderl, R., Haag, S., Schützer, K., Zancul, E., 2018, Digital Twin Technology – An Approach for Industrie 4.0 Vertical and Horizontal Lifecycle Integration, it - Inf. Technol., 60/3:125–132.
[4] Andress, J., 2011, The Basics of Information Security: Understanding the Fundamentals of InfoSec in Theory and Practice, Elsevier.
[5] Arulampalam, M. S., Maskell, S., Gordon, N., Clapp, T., 2007, A Tutorial on Particle Filters for Online Nonlinear/Nongaussian Bayesian Tracking, Bayesian Bounds Param. Estim. Nonlinear Filtering/Tracking, 50/2:723–737.
[6] Asmare, E., McCann, J. A., 2014, Lightweight Sensing Uncertainty Metric-Incorporating Accuracy and Trust, IEEE Sens. J., 14/12:4264–4272.
[7] Azeem, R., Khan, M. I. A., 2012, Techniques about Data Replication for Mobile Ad-Hoc Network Databases, Int. J. Multidiscip. Sci. Eng., 3/5:53–57.
[8] Bajaj, M., Hedberg, T., 2018, System Lifecycle Handler - Spinning a Digital Thread for Manufacturing, INCOSE Int. Symp., 28/1:1636–1650.
[9] Bar-Shalom, Y., Li, X.-R., Kirubarajan, T., 2003, Estimation with Applications to Tracking and Navigation, John Wiley & Sons.
[10] Barber, R. F., Candés, E. J., 2015, Controlling the False Discovery Rate via Knockoffs, Ann. Stat., 43/5:2055–2085.
[11] Bay, H., Tuytelaars, T., Van Gool, L., 2006, SURF: Speeded up Robust Features, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 3951 LNCS, 404–417.
[12] Behrendt, T., Zein, A., Min, S., 2012, Development of an Energy Consumption Monitoring Procedure for Machine Tools, CIRP Ann. - Manuf. Technol., 61/1: 43–46.
[13] Bernstein, W. Z., Hedberg Jr., T. D., Helu, M., Feeney, A. B., 2018, Contextualising Manufacturing Data for Lifecycle Decision-Making, Int. J. Prod. Lifecycle Manag., 10/4:326.
[14] Bhinge, R., Sutherland, J. W., Robinson, S. L., Dornfeld, D., Hutchins, M. J., Micali, M. K., 2015, Framework for Identifying Cybersecurity Risks in Manufacturing, Procedia Manufacturing, 1:47–63.
[15] Bishop, G., Welch, G., 2001, An Introduction to the Kalman Filter - Course 8, An Introduction to the Kalman Filter.
[16] Bisht, S. S., Singh, M. P., 2014, An Unscented Kalman Filter for Tracking Sudden Stiffness Changes, Mech. Syst. Signal Process., 49/1-2:181–195.
[17] Block, C., Lins, D., Kuhlenkötter, B., 2018, Approach for a Simulation-Based and Event-Driven Production Planning and Control in Decentralized Manufacturing Execution Systems, Procedia CIRP, 72:1351–1356.
[18] Bloom, Z., A Quick History of Digital Communication Before the Internet.
[19] Breiman, L., 2001, Statistical Modeling: The Two Cultures, Stat. Sci., 16/3:199–231.
[20] Brundage, M. P., Bernstein, W. Z., Morris, K. C., Horst, J. A., 2017, Using Graph-Based Visualizations to Explore Performance Indicator Relationships for Manufacturing Production Systems, Procedia CIRP, 61:451–456.
[21] Buza, K., Nagy, G. I., Nanopoulos, A., 2014, Storage-Optimizing Clustering Algorithms for High-Dimensional Tick Data, Expert Syst. Appl., 41/9:4148–4157.
[22] Caggiano, A., 2018, Cloud-based manufacturing process monitoring for smart diagnosis services, Int. J. Comput. Integr. Manuf., 31/7:612-623.
[23] Caggiano, A., Rimpault, X., Teti, R., Balazinski, M., Chatelain, J. F., Nele, L., 2018, Machine Learning Approach Based on Fractal Analysis for Optimal Tool Life Exploitation in CFRP Composite Drilling for Aeronautical Assembly, CIRP Ann., 67/1: 483–486.
[24] Caggiano, A., Zhang, J., Alfieri, V., Caiazzo, F., Gao, R., Teti, R., 2019, Machine Learning-Based Image Processing for on-Line Defect Recognition in Additive Manufacturing, CIRP Ann., 68/1:451–454.
[25] Cai, L., Zhu, Y., 2015, The Challenges of Data Quality and Data Quality Assessment in the Big Data Era, Data Sci. J., 14: 2.
[26] Candela, L., Castelli, D., Pagano, P., 2012, Managing Big Data through Hybrid Data Infrastructures, ERCIM News, (September 2015), 37–38.
[27] Candes, E. J., Tao, T., 2006, Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies?, IEEE Trans. Inf. Theory, 52/12: 5406–5425.
[28] Cao, J., Brinksmeier, E., Fu, M., Gao, R., Liang, B., Merklein, M., Schmidt, M., Yanagimoto, J., 2019, Manufacturing of Advanced Smart Tooling for Metal Forming, CIRP Ann. 68/2: 605-628.
[29] Cao, P., Zhang, S., Tang, J., 2018, Preprocessing-Free Gear Fault Diagnosis Using Small Datasets with Deep Convolutional Neural Network-Based Transfer Learning, IEEE Access, 6:26241–26253.
[30] Chen, B., Wan, J., Celesti, A., Li, D., Abbas, H., Zhang, Q., 2018, Edge Computing in IoT-Based Manufacturing, IEEE Com. Mag., 56/9: 103–109.
[31] Chen, M., Mao, S., Liu, Y., 2014, Big Data: A Survey, Mobile Networks and Applications, 12/2:171–209.
[32] Chen, M., Miao, Y., Hao, Y., Hwang, K., 2017, Narrow Band Internet of Things, IEEE Access, 5: 20557–20577.
[33] Chen, Z., Li, W., 2017, Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network, IEEE Trans. Instr. Meas. 66/7:1693-1702.
[34] Chien, C. F., Chuang, S. C., 2014, A Framework for Root Cause Detection of Sub-Batch Processing System for Semiconductor Manufacturing Big Data Analytics, IEEE Trans. Semicond. Manuf., 27/4: 475–488.
[35] Dalal, N., Triggs, B., 2005, Histograms of Oriented Gradients for Human Detection, Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 886–893.
[36] De Mauro, A., Greco, M., Grimaldi, M., 2016, A Formal Definition of Big Data Based on Its Essential Features, Libr. Rev., 65/3: 122–135.
[37] Deming, W.E., 1942. On a classification of the problems of statistical inference, J. Am. Stat. Assoc., 37/218:173-185.
[38] Digital Manufacturing and Design Innovation Institute, 2017, Operate, Orchestrate, and Originate.
[39] Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R. V, Jhingran, A., Kanungo, T., Rajagopalan, S., Tomkins, A., Tomlin, J. A., Zien, J. Y., 2003, SemTag and Seeker: Bootstrapping the Semantic Web via Automated Semantic Annotation, Proc. 12th World Wide Web Conf., 178–186.
[40] Ding, X., He, Q., 2017, "Energy-Fluctuated Multiscale Feature Learning with Deep ConvNet for Intelligent Spindle Bearing Fault Diagnosis, IEEE Trans. Instrum. Meas., 66/8: 1926–1935.
[41] Doll, W., Vonderembse, M., 1991, The Evolution of Manufacturing Systems: Towards the Post-Industrial Enterprise, Int. J. of Mgmt Sci, 19/5:401–411.
[42] Dorgo, G., Varga, K., Abonyi, J., 2018, Hierarchical Frequent Sequence Mining Algorithm for the Analysis of Alarm Cascades in Chemical Processes, IEEE Access, 6: 50197–50216.
[43] Dornheim, J., Link, N., Gumbsch, P., 2019, Model-Free Adaptive Optimal Control of Episodic Fixed-Horizon Manufacturing Processes Using Reinforcement Learning, arXiv:1809.06646v3 [cs.SY].
[44] Economist, 2017, The Most Valuable Resource Is No Longer Oil, but Data.

[45] Erler, A., de Mas, N., Ramsey, P., Henderson, G., 2013, Efficient Biological Process Characterization by Definitive-Screening Designs: The Formaldehyde Treatment of a Therapeutic Protein as a Case Study, Biotechnol. Lett., 35/3: 323–329.
[46] Esposito, C., Castiglione, A., Martini, B., Choo, K. K. R., 2016, Cloud Manufacturing: Security, Privacy, and Forensic Concerns, IEEE Cloud Comput., 3/4:16–22.
[47] Everton, S. K., Hirsch, M., Stavroulakis, P. I., Leach, R. K., Clare, A. T., 2016, Review of In-Situ Process Monitoring and in-Situ Metrology for Metal Additive Manufacturing, Mater. Des., 95: 431–445.
[48] European Commission, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation)
[49] Fan, Z., Ng, M. K., Gao, R., Cao, J., Smith, E. F., 2012, Real-Time Monitoring of Pressure Distribution in Microrolling through Embedded Capacitive Sensing, CIRP Ann. - Manuf. Technol., 61/1: 367–370.
[50] Ferri, M., 2018, Why Topology for Machine Learning and Knowledge Extraction?, Mach. Learn. Knowl. Extr., 1/1: 115–120.
[51] Fujishima, M., Ohno, K., Nishikawa, S., Nishimura, K., Sakamoto, M., Kawai, K., 2016, Study of Sensing Technologies for Machine Tools, CIRP J. Manuf. Sci. Technol., 14: 71–75.
[52] Furumoto, T., Ueda, T., Alkahari, M. R., Hosokawa, A., 2013, Investigation of Laser Consolidation Process for Metal Powder by Two-Color Pyrometer and High-Speed Video Camera, CIRP Ann. - Manuf. Technol., 62/1: 223–226.
[53] Gani, A., Siddiqa, A., Shamshirband, S., Hanum, F., 2016, A Survey on Indexing Techniques for Big Data: Taxonomy and Performance Evaluation, Knowl. Inf. Syst., 46/2: 241–284.
[54] Gao, R., Fan, Z., Asadizanjani, N., Kazmer, D., 2013, Acoustic-Based Wireless Signal Transmission for Precision Metrology: Accuracy and Reliability, CIRP Ann. - Manuf. Technol., 62/1: 519–522.
[55] Gao, R., Wang, L., Teti, R., Dornfeld, D., Kumara, S., Mori, M., Helu, M., 2015, Cloud-Enabled Prognosis for Manufacturing, CIRP Ann. - Manuf. Technol., 64/2:749–772.
[56] General Electric, 2012, The Rise of Industrial Big Data.
[57] German Machine Tool Builders' Association and Mechanical Engineering Industry Association, 2020, umati – on its way to becoming the global language of production, Press Release.
[58] Geron, A., 2017, Hands-On Machine Learning, O'Reilly.
[59] Gilmore, A., 2015, Building Better Predictive Models Using Topology.
[60] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014, Generative Adversarial Networks, Adv. Neural Inf. Process. Syst. 2672-2680.
[61] Groche, P., Brenneis, M., 2014, Manufacturing and Use of Novel Sensoric Fasteners for Monitoring Forming Processes, Meas. J. Int. Meas. Confed., 53: 136–144.
[62] Guo, W., Banerjee, A. G., 2016, Toward Automated Prediction of Manufacturing Productivity Based on Feature Selection Using Topological Data Analysis, 2016 IEEE International Symposium on Assembly and Manufacturing, ISAM 2016.
[63] Guo, W., Banerjee, A. G., 2017, Identification of Key Features Using Topological Data Analysis for Accurate Prediction of Manufacturing System Outputs, J. Manuf. Syst., 43: 225–234.
[64] Guo, W., Shao, C., Kim, T. H., Hu, S. J., Jin, J., Spicer, J. P., Wang, H., 2016, Online Process Monitoring with Near-Zero Misdetection for Ultrasonic Welding of Lithium-Ion Batteries: An Integration of Univariate and Multivariate Methods, J. Manuf. Syst., 38: 141–150.
[65] Guo, Z., Ye, S., Wang, Y., Lin, C., 2017, Resistance Welding Spot Defect Detection with Convolutional Neural Networks, Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 169–174.
[66] Haefner, B., Lanza, G., 2017, Function-Oriented Measurements and Uncertainty Evaluation of Micro-Gears for Lifetime Prognosis, CIRP Ann. - Manuf. Technol., 66/1: 475–478.
[67] Harding, J. A., Shahbaz, M., Srinivas, Kusiak, A., 2006, Data Mining in Manufacturing: A Review, J. Manuf. Sci. Eng., 128/4: 969.
[68] Hardwick M. 2018, Requirements for a Digital Twin Manufacturing Framework, In: Proceedings of the 2018 Model-Based Enterprise (MBE) Summit, NIST Advanced Manufacturing Series 100-22.
[69] Harris, C., Stephens, M., 1988, A Combined Corner and Edge Detector, Procedings of the Alvey Vision Conference 1988, 23.1-23.6.
[70] Hasan, M. J., Kim, J.-M., 2018, Bearing Fault Diagnosis under Variable Rotational Speeds Using Stockwell Transform-Based Vibration Imaging and Transfer Learning, Appl. Sci., 8/12:2357.
[71] Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
[72] He, Q. P., Wang, J., 2011, Statistics Pattern Analysis: A New Process Monitoring Framework and Its Application to Semiconductor Batch Processes, AIChE J., 57/1: 107–121.
[73] He, Q. P., Wang, J., 2018, Statistical Process Monitoring as a Big Data Analytics Tool for Smart Manufacturing, J. Process Control, 67: 35–43.
[74] He, Q. P., Wang, J., Shah, D., 2019, Feature Space Monitoring for Smart Manufacturing via Statistics Pattern Analysis, Comput. Chem. Eng., 126: 321–331.
[75] Hedberg T., Helu M., Sprock T. 2018, A Standards and Technology Roadmap for Scalable Distributed Manufacturing Systems. In: Proceedings of the ASME 2018 Manufacturing Science and Engineering Conference, pp. V003T02A019.
[76] Hedberg, T., Feeney, A. B., Helu, M., Camelio, J. A., 2017, Toward a Lifecycle Information Framework and Technology in Manufacturing, J. Comput. Inf. Sci. Eng., 17/2:021010.
[77] Hedberg, T. D., Krima, S., Camelio, J. A., 2016, Embedding X.509 Digital Certificates in Three-Dimensional Models for Authentication, Authorization, and Traceability of Product Data, J. Comput. Inf. Sci. Eng., 17/1: 011008.
[78] Hedberg, T. D., Krima, S., Camelio, J. A., 2019, Method for Enabling a Root of Trust in Support of Product-Data Certification and Traceability, J. Comput. Inf. Sci. Eng., In Press.
[79] Heer, J., Bostock, M., Ogievetsky, V., 2010, A Tour through the Visualization Zoo, Commun. ACM, 53/6: 59.
[80] Helu, M., Hedberg Jr, T., 2017, Connecting, Deploying, and Using the Smart Manufacturing Systems Test Bed, NIST Adv. Manuf. Ser.
[81] Helu, M., Joseph, A., Hedberg, T., 2018, A Standards-Based Approach for Linking as-Planned to as-Fabricated Product Data, CIRP Ann., 67/1: 487–490.
[82] Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., 2016, The Age of Analytics: Competing in a Data-Driven World.
[83] Hinton, G. E., Osindero, S., Teh, Y. W., 2006, A Fast Learning Algorithm for Deep Belief Nets, Neural Comput., 18/7: 1527–1554.
[84] Hochreiter, S., Schmidhuber, J., 1997, Long Short-Term Memory, Neural Comput., 9/8:1735–1780.
[85] Hu, H., Wen, Y., Chua, T. S., Li, X., 2014, Toward Scalable Systems for Big Data Analytics: A Technology Tutorial,"IEEE Access, 2:652–687.
[86] Hu, S. J., 2013, Evolving Paradigms of Manufacturing: From Mass Production to Mass Customization and Personalization, Procedia CIRP, 7: 3–8.
[87] International Organization for Standardization, 2014, Industrial Automation Systems and Integration – Product Data Representation and Exchange – Part 242: Application Protocol: Managed Model-Based 3D Engineering, ISO 10303-242:2014.
[88] International Organization for Standardization and International Electrotechnical Commission, 2019, Information Technology – Security Techniques – Information Security Management Systems – Requirements, ISO/IEC 27001:2013.
[89] Ioannidis, J. P. A., 2014, Discussion: Why 'An Estimate of the Science-Wise False Discovery Rate and Application to the Top Medical Literature' Is False, Biostatistics, 15/1:28–36.
[90] Irani, K. B., Cheng, J., Fayyad, U. M., Qian, Z., 1993, Applying Machine Learning to Semiconductor Manufacturing, IEEE Expert, 1/1:41–47.
[91] Ismail, A., Truong, H. L., Kastner, W., 2019, Manufacturing Process Data Analysis Pipelines: A Requirements Analysis and Survey, J. Big Data, 6:1.
[92] Jalali, A., 2016, Reflections on the Use of Chord Diagrams in Social Network Visualization in Process Mining, Proceedings - International Conference on Research Challenges in Information Science.
[93] Janssens, O., Van De Walle, R., Loccufier, M., Van Hoecke, S., 2018, Deep Learning for Infrared Thermal Image Based Machine Health Monitoring, IEEE/ASME Trans. Mechatronics, 23/1: 151–159.
[94] Ji, W., Wang, L., 2017, Big Data Analytics Based Fault Prediction for Shop Floor Scheduling, J. Manuf. Syst., 43:187–194.
[95] Jin, X., Wah, B. W., Cheng, X., Wang, Y., 2015, Significance and Challenges of Big Data Research, Big Data Res., 2/2: 59–64.
[96] Joint Committee for Guides in Metrology, 2008, Guide to the expression of uncertainty in measurement.
[97] Jones, B., Nachtsheim, C. J., 2011, A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects, J. Qual. Technol., 43/1: 1–15.
[98] Jones, B., Nachtsheim, C. J., 2013, Definitive Screening Designs with Added Two-Level Categorical Factors, J. Qual. Technol., 45/2: 121–129.
[99] Jones, B., Nachtsheim, C. J., 2016, Blocking Schemes for Definitive Screening Designs, Technometrics, 58/1: 74–83.
[100] Kambatla, K., Kollias, G., Kumar, V., Grama, A., 2014, Trends in Big Data Analytics, J. Parallel Distrib. Comput., 74/7: 2561–2573.
[101] Kazmer, D. O., Gordon, G. W., Mendible, G. A., Johnston, S. P., Tang, X., Fan, Z., Gao, R., 2015, A Multivariate Sensor for Intelligent Polymer Processing, IEEE/ASME Trans. Mechatronics, 20/3: 1015–1023.
[102] Kim, S., Kim, W., Noh, Y. K., Park, F. C., 2017, Transfer Learning for Automated Optical Inspection, Proceedings of the International Joint Conference on Neural Networks, 2517–2524.
[103] Kiryakov, A., Popov, B., Terziev, I., Manov, D., Ognyanoff, D., 2004, Semantic Annotation, Indexing, and Retrieval, Web Semant., 2/1: 49–79.
[104] Kozjek, D., Vrabič, R., Eržen, G., and Butala, P., 2018, "Identifying the Business and Social Networks in the Domain of Production by Merging the Data from Heterogeneous Internet Sources," Int. J. Prod. Econ., 200, pp. 181–191.
[105] Kozjek, D., Vrabič, R., Kralj, D., Butala, P., 2017, A Data-Driven Holistic Approach to Fault Prognostics in a Cyclic Manufacturing Process, Procedia CIRP, 63: 664–669.
[106] Kozjek, D., Vrabič, R., Rihtaršič, B., Butala, P., 2018, Big Data Analytics for Operations Management in Engineer-to-Order Manufacturing, Procedia CIRP, 72: 209–214.
[107] Kozjek, D., Vrabič, R., Kralj, D., Butala, P., 2017, Interpretative Identification of the Faulty Conditions in a Cyclic Manufacturing Process, J. Manuf. Syst., 43: 214–224.
[108] Krima, S., Hedberg Jr, T., Barnard Feeney, A., 2018, Securing the Digital Thread for Smart Manufacturing: A Reference Model for Blockchain-Based Product Data Traceability, NIST Adv. Manuf. Ser.
[109] Krizhevsky, A., Sutskever, I., Hinton, G., 2012, Imagenet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems, 1097–1105.
[110] Krüger, J., Lehr, J., Schlüter, M., Bischoff, N., 2019, Deep Learning for Part Identification Based on Inherent Features, CIRP Ann., 68/1: 9–12.
[111] Kurada, S., Bradley, C., 1997, A Review of Machine Vision Sensors for Tool Condition Monitoring, Comput. Ind., 34/1: 55–72.

[112] Kurle, D., Herrmann, C., Thiede, S., 2017, Unlocking Water Efficiency Improvements in Manufacturing — From Approach to Tool Support, CIRP J. Manuf. Sci. Technol., 19: 7–18.
[113] Kusiak, A., 2017, Smart Manufacturing Must Embrace Big Data, Nature, 544:23–25.
[114] Kwak, Y. H., Anbari, F. T., 2006, Benefits, Obstacles, and Future of Six Sigma Approach, Technovation, 26/5-6: 708–715.
[115] Lanza, G., Moser, R., 2014, Multi-Objective Optimization of Global Manufacturing Networks Taking into Account Multi-Dimensional Uncertainty, CIRP Ann. - Manuf. Technol., 63/1: 397–400.
[116] Lanza, G., Viering, B., 2011, A Novel Standard for the Experimental Estimation of the Uncertainty of Measurement for Micro Gear Measurements, CIRP Ann. - Manuf. Technol., 60/1: 543–546.
[117] Lecun, Y., Bengio, Y., Hinton, G., 2015, Deep Learning, Nature, 521: 436–444.
[118] Lee, J.-M., Qin, S. J., Lee, I.-B., 2008, Fault Detection of Non-Linear Processes Using Kernel Independent Component Analysis, Can. J. Chem. Eng., 85/4:526–536.
[119] Lee, Y. O., Jo, J., Hwang, J., 2018, Application of Deep Neural Network and Generative Adversarial Network to Industrial Maintenance: A Case Study of Induction Motor Fault Detection, Proceedings - 2017 IEEE International Conference on Big Data, 3248–3253.
[120] Lei, P., Zheng, L., Wang, L., Wang, Y., Li, C., Li, X., 2017, MTConnect Compliant Monitoring for Finishing Assembly Interfaces of Large-Scale Components: A Vertical Tail Section Application, J. Manuf. Syst., 45: 121–134.
[121] Leskovec, J., Rajaraman, A., Ullman, J. D., 2014, Mining of Massive Datasets: Second Edition.
[122] Levine, S., Pastor, P., Krizhevsky, A., Ibarz, J., Quillen, D., 2018, Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection, Int. J. Rob. Res., 37/4-5: 421–436.
[123] Li, L., Cheng, W. Y., Glicksberg, B. S., Gottesman, O., Tamler, R., Chen, R., Bottinger, E. P., Dudley, J. T., 2015, Identification of Type 2 Diabetes Subgroups through Topological Analysis of Patient Similarity, Sci. Transl. Med., 7/311: 311ra174.
[124] Li, T., Sattar, T. P., Sun, S., 2012, Deterministic Resampling: Unbiased Sampling to Avoid Sample Impoverishment in Particle Filters, Signal Processing, 92/7: 1637–1645.
[125] Li, X., Wan, J., Dai, H. N., Imran, M., Xia, M., Celesti, A., 2019, A Hybrid Computing Solution and Resource Scheduling Strategy for Edge Computing in Smart Manufacturing, IEEE Trans. Ind. Informatics, 15/7: 4225–4234.
[126] Li, Y., Yang, Y., Li, G., Xu, M., Huang, W., 2017, A Fault Diagnosis Scheme for Planetary Gearboxes Using Modified Multi-Scale Symbolic Dynamic Entropy and MRMR Feature Selection, Mech. Syst. Signal Process., 91: 295–312.
[127] Liu, C., Wu, X., Mao, J., Liu, X., 2017, Acoustic Emission Signal Processing for Rolling Bearing Running State Assessment Using Compressive Sensing, Mech. Syst. Signal Process., 91:395–406.
[128] Liu, H., Fang, T., Zhou, T., Wang, L., 2018, Towards Robust Human-Robot Collaborative Manufacturing: Multimodal Fusion, IEEE Access, 6: 74762–74771.
[129] Liu, R., Yang, B., Zio, E., Chen, X., 2018, Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review, Mech. Syst. Signal Process., 108: 33–47.
[130] Lu, Y., Wang, Y., 2018, Monitoring Temperature in Additive Manufacturing with Physics-Based Compressive Sensing, J. Manuf. Syst., 48: 60–70.
[131] Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., Carlsson, G., 2013, Extracting Insights from the Shape of Complex Data Using Topology, Sci. Rep., 3:1236
[132] Luzanin, O., Guduric, V., Ristic, I., Muhic, S., 2017, Investigating Impact of Build Parameters on Maximum Flexural Force in FDM Specimens - A Definitive Screening Design Approach, Rapid Prototyp. J., 23/6:1088–1098.
[133] Ma, M., Sun, C., Chen, X., 2017, Discriminative Deep Belief Networks with Ant Colony Optimization for Health Status Assessment of Machine, IEEE Trans. Instrum. Meas., 66/12: 3115–3125.
[134] MacGregor, J. F., Kourti, T., 1995, Statistical Process Control of Multivariate Processes, Control Eng. Pract., 3/3: 403–414.
[135] Maletic, J., Marcus, A., 2000, Data Cleansing: Beyond Integrity Analysis, 1–10.
[136] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A., 2011, Big Data: The Next Frontier for Innovation, Competition, and Productivity.
[137] Meng, X. L., 2018, Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 Us Presidential Election, Ann. Appl. Stat., 12/2: 685–726.
[138] McKinsey, 2017, Ops 4.0: Manufacturing's future, made by people.
[139] Merriam-Webster Incorporated, 2018, Merriam-Webster Dictionary.
[140] Mitchell, T., 1997, Machine Learning. McGraw Hill.
[141] Monostori, L., Csáji, B. C., 2006, Stochastic Dynamic Production Control by Neurodynamic Programming, CIRP Ann. - Manuf. Technol., 55/1:473–478.
[142] Monostori, L., Csáji, B. C., Kádár, B., 2004, Adaptation and Learning in Distributed Production Control, CIRP Ann. - Manuf. Technol., 53/1: 349–352.
[143] Monostori, L., Kádár, B., Bauernhansl, T., Kondoh, S., Kumara, S., Reinhart, G., Sauer, O., Schuh, G., Sihn, W., Ueda, K., 2016, Cyber-Physical Systems in Manufacturing, CIRP Ann., 65/2:621–641.
[144] Montgomery, D. C., 2012, Design and Analysis of Experiments. John Wiley & Sons.
[145] Mourtizis, D. A., Doukas, M., 2014, The Evolution of Manufacturing Systems: From Craftsmanship to the Era of Customisation, Handbook of Research on Design and Management of Lean Production Systems, 1–29.
[146] Mozaffar, M., Paul, A., Al-Bahrani, R., Wolff, S., Choudhary, A., Agrawal, A., Ehmann, K., Cao, J., 2018, Data-Driven Prediction of the High-Dimensional Thermal History in Directed Energy Deposition Processes via Recurrent Neural Networks, Manuf. Lett., 18: 35–39.
[147] Moutinho, L., Hutcheson, G., Moutinho, L., 2014, Correlation Analysis,"The SAGE Dictionary of Quantitative Management Research, 57–60.
[148] MTConnect Institute, 2018, MTConnect v1.4.0, ANSI/MTC1.4-2018.
[149] Nannapaneni, S., Mahadevan, S., 2015, Uncertainty Quantification in Performance Evaluation of Manufacturing Processes, Proceedings - 2014 IEEE International Conference on Big Data, 996–1005.
[150] Nannapaneni, S., Mahadevan, S., Rachuri, S., 2016, Performance Evaluation of a Manufacturing Process under Uncertainty Using Bayesian Networks, J. Clean. Prod., 113:947–959.
[151] National Institute of Standards and Technology, Cybersecurity framework, (Image credit: Hanacek, N.)
[152] National Institute of Standards and Technology, 2001, Security Requirements for Cryptographic Modules, NIST FIPS – 140-2.
[153] National Institute of Standards and Technology, 2011, Advanced Encryption Standard (AES), 2001, NIST FIPS – 197.
[154] National Institute of Standards and Technology, 2018, NIST Special Publication 800-Series General Information.
[155] National Institute of Standards and Technology, 2018, Framework for Improving Critical Infrastructure Cybersecurity.
[156] National Institute of Standards and Technology, 2018, FIPS General Information.
[157] National Institute of Standards and Technology, 2019, Blockchain Provides Security, Traceability for Smart Manufacturing (Image credit: Hanacek, N.)
[158] National Science Foundation, 2003, Sensors and sensing networks.
[159] Nestor: a toolkit for quantifying tacit maintenance knowledge, for investigatory analysis in smart manufacturing, National Institute of Standards and Technology, US Department of Commerce.
[160] Ng, A. Y., 2011, Sparse Autoencoder.
[161] Online etymology dictionary, https://www.etymonline.com/
[162] OPC Foundation, 2017, OPC Unified Architecture Specifications v1.04, OPC 10000.
[163] Pan, S. J., and Yang, Q., 2010, A Survey on Transfer Learning, IEEE Trans. Knowl. Data Eng., 22/10:1345–1359.
[164] Parab, N. D., Zhao, C., Cunningham, R., Escano, L. I., Fezzaa, K., Everhart, W., Rollett, A. D., Chen, L., Sun, T., 2018, Ultrafast X-Ray Imaging of Laser–Metal Additive Manufacturing Processes, J. Synchr Radiat., 25/5: 1467–1477.
[165] Patil, M. V., 2017, Multi Response Simulation and Optimization of Gas Tungsten Arc Welding, Appl. Math. Model., 42: 540–553.
[166] Peleg, T., and Elad, M., 2014, A Statistical Prediction Model Based on Sparse Representations for Single Image Super-Resolution, IEEE Trans. Image Process., 23/6:2569–2582.
[167] Philip Chen, C. L., Zhang, C. Y., 2014, Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data, Inf. Sci. (Ny)., 275: 314–347.
[168] Piatetsky-Shapiro, G., 1999, The Data Mining Industry Coming of Age, IEEE Intell. Syst.,14/6: 32–34.
[169] Qian, G., Lu, S., Pan, D., Tang, H., Liu, Y., Wang, Q., 2019, Edge Computing: A Promising Framework for Real-Time Fault Diagnosis and Dynamic Control of Rotating Machines Using Multi-Sensor Data, IEEE Sens. J., 19/11: 4211–4220.
[170] Qiu, P., 2017, Statistical Process Control Charts as a Tool for Analyzing Big Data, 123–138.
[171] Raissi, M., Perdikaris, P., Karniadakis, G. E., 2019, Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations, J. Comput. Phys., 378:686–707.
[172] Regli, W., Rossignac, J., Shapiro, V., Srinivasan, V., 2016, The New Frontiers in Computational Modeling of Material Structures, CAD Comput. Aided Des., 77: 73–85.
[173] Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M., Sabeti, P. C., 2011, Detecting Novel Associations in Large Data Sets, Science, 334/6062: 1518–1524.
[174] Reuben, L. C. K., Mba, D., 2014, Diagnostics and Prognostics Using Switching Kalman Filters, Struct. Heal. Monit., 13/3: 296–306.
[175] Richter, B., Blanke, N., Werner, C., Parab, N. D., Sun, T., Vollertsen, F., Pfefferkorn, F. E., 2019, High-Speed X-Ray Investigation of Melt Dynamics during Continuous-Wave Laser Remelting of Selective Laser Melted Co-Cr Alloy, CIRP Ann., 68/1: 229–232.
[176] Rosen, R., Von Wichert, G., Lo, G., Bettenhausen, K. D., 2015, About the Importance of Autonomy and Digital Twins for the Future of Manufacturing, IFAC-PapersOnLine, 567–572.
[177] Rucco, M., Sousa-Rodrigues, D., Merelli, E., Johnson, J. H., Falsetti, L., Nitti, C., Salvi, A., 2015, Neural Hypernetwork Approach for Pulmonary Embolism Diagnosis, BMC Res. Notes, 8(1).
[178] Ruemler, S. P., Zimmerman, K. E., Hartman, N. W., Hedberg, T., Barnard Feeny, A., 2016, Promoting Model-Based Definition to Establish a Complete Product Definition, J. Manuf. Sci. Eng., 139/5: 051008.
[179] Sankararaman, S., 2015, Significance, Interpretation, and Quantification of Uncertainty in Prognostics and Remaining Useful Life Prediction, Mech. Syst. Signal Process., 53/1: 228–247.
[180] Schleich, B., Anwer, N., Mathieu, L., Wartzack, S., 2017, Shaping the Digital Twin for Design and Production Engineering, CIRP Ann. - Manuf. Technol., 66/1:141–144.
[181] Scutari, M., 2010, Learning Bayesian Networks with the Bnlearn R Package, J. Stat. Softw., 35/3.
[182] Segreto, T., Simeone, A., Teti, R., 2014, Principal Component Analysis for Feature Extraction and NN Pattern Recognition in Sensor Monitoring of Chip Form during Turning, CIRP J of Manuf. Sci. and Tech., 7:. 202–209.
[183] Segreto, T., Teti, R., 2019, Machine Learning for In-Process End-Point Detection in Robot-Assisted Polishing Using Multiple Sensor Monitoring, Int. J. Adv. Manuf. Technol., 103:4173–4187.
[184] Sexton, T., Brundage, M. P., Hoffman, M., Morris, K., 2017, Hybrid datafication of maintenance logs from AI-assisted human tags, Proceedings of the 2017 IEEE International Conference on Big Data, 1769-1777.

[185] Sexton, T., Hodkiewicz, M., Brundage, M. P., Smoker, T., 2018, Benchmarking for keyword extraction methodologies in maintenance work orders, Proceedings of the Annual Conference of the Prognostics and Health Management Society 2018, 10/1.
[186] Shao, S., Sun, W., Wang, P., Gao, R., Yan, R., 2016, Learning Features from Vibration Signals for Induction Motor Fault Diagnosis, International Symposium on Flexible Automation, ISFA 2016, 71–76.
[187] Shao, S., Wang, P., Yan, R., 2019, Generative Adversarial Networks for Data Augmentation in Machine Fault Diagnosis, Comput. Ind., 106: 85–93.
[188] Shi, W., Cao, J., Zhang, Q., Li, Y., Xu, L., 2016, Edge Computing: Vision and Challenges, IEEE Internet Things J., 3/5:. 637–646.
[189] Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., Nasaruddin, F., 2016, A Survey of Big Data Management: Taxonomy and State-of-the-Art, J. Netw. Comput. Appl., 71: 151–166.
[190] Siemens, 2017, Digital Innovation Platform.
[191] Singh, G., Mémoli, F., Carlsson, G., 2007, Topological Methods for the Analysis of High Dimensional Data Sets and 3D Object Recognition, Eurographics Symp. Point-Based Graph., 91–100.
[192] Smolenicki, D., Boos, J., Kuster, F., Roelofs, H., Wyen, C. F., 2014, In-Process Measurement of Friction Coefficient in Orthogonal Cutting, CIRP Ann. - Manuf. Technol., 63/1: 97–100.
[193] Sobel W., Hardwick M., Maggiano L., Horst J. 2017, Enabling Automated CNC Measurement on the Digital Thread. In: Proceedings of the 2017 Model-Based Enterprise (MBE) Summit, NIST Advanced Manufacturing Series 100-11, Hedberg T., Carlisle M. (eds).
[194] Stark R. Damerau T. 2019, Digital Twin. In: The International Academy for Production Engineering, Chatti S., Tolio T. (eds) CIRP Encyclopedia of Production Engineering. Springer, Berlin, Heidelberg.
[195] Stouffer, K., Zimmerman, T., Tang, C. Y., Cichonski, J., Shah, N., Downard, W., 2019, Cybersecurity Framework Manufacturing Profile Low Security Level Example Implementations Guide: Volume 3 – Discrete-Based Manufacturing System Use Case.
[196] Stricker, N., Kuhnle, A., Sturm, R., Friess, S., 2018, Reinforcement Learning for Adaptive Order Dispatching in the Semiconductor Industry, CIRP Ann., 67/1:511–514.
[197] Sun, C., Ma, M., Zhao, Z., Tian, S., Yan, R., Chen, X., 2019, Deep Transfer Learning Based on Sparse Autoencoder for Remaining Useful Life Prediction of Tool in Manufacturing, IEEE Trans. Ind. Informatics. 15/4: 2416 – 2425.
[198] Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., Chen, X., 2016, A Sparse Auto-Encoder-Based Deep Neural Network Approach for Induction Motor Faults Classification, Meas. J. Int. Meas. Confed., 89:171–178.
[199] Sutton, R., Barto, A., 2018, Reinforcment Learning, MIT Press.
[200] Tan, P., Steinbach, M., Kumar, V., 2005, Introduction to Data Mining, Addison-Wesley Longman Publishing Co., Inc.
[201] Tao, F., Qi, Q., Liu, A., Kusiak, A., 2018, Data-Driven Smart Manufacturing, J. Manuf. Syst., 48: 157–169.
[202] Tao, F., Zhang, M., Liu, Y., Nee, A. Y. C., 2018, Digital Twin Driven Prognostics and Health Management for Complex Equipment, CIRP Ann., 67/1: 169–172.
[203] Tausendfreund, A., Borchers, F., Kohls, E., Kuschel, S., Stöbener, D., Heinzel, C., Fischer, A., 2018, Investigations on Material Loads during Grinding by Speckle Photography, J. Manuf. Mater. Process., 2/4: 71.
[204] Telecommunications Standardization Sector of ITU (2014) Information Technology – Open Systems Interconnection – The Directory – Part 8: Public-Key and Attribute Certificate Frameworks. International Organization for Standardization, ISO/IEC 9594-8:2014.
[205] Telecommunications Standardization Sector of ITU, Security in Telecommunications and Information Technology.
[206] Tesch, M., IoT and Predictive Analytics: Fog and Edge Computing for Industries vs. Cloud.
[207] Teti, R., Jemielniak, K., O'Donnell, G., Dornfeld, D., 2010, Advanced Monitoring of Machining Operations, CIRP Ann. - Manuf. Technol., 59/2: 717–739.
[208] The MITRE Corporation, 2019, CWE: Common Weakness Enumeration.
[209] Tong, F., Sun, Y., He, S., 2019, "n Positioning Performance for the Narrow-Band Internet of Things: How Participating ENBs Impact?, IEEE Trans. Ind. Informatics, 15/1: 423–433.
[210] Ueda, K., Hatono, I., Fujii, N., Vaario, J., 2000, Reinforcement Learning Approaches to Biological Manufacturing System, CIRP Ann. - Manuf. Technol., 49/1: 343–346.
[211] U.S. Department of Commerce, Export Administration Regulations, 15 CFR § 730 et Seq.
[212] U.S. Department of State, International Traffic in Arms Regulations, 22 CFR 120-130.
[213] Van Der Maaten, L., and Hinton, G., 2008, Visualizing Data Using T-SNE, J. Mach. Learn. Res., 9: 2579–2625.
[214] Verba, N., Chao, K. M., Lewandowski, J., Shah, N., James, A., Tian, F., 2019, Modeling Industry 4.0 Based Fog Computing Environments for Application Analysis and Deployment, Futur. Gener. Comput. Syst., 91:48–60.
[215] Vijayaraghavan, A., Dornfeld, D., 2010, Automated Energy Monitoring of Machine Tools, CIRP Ann. - Manuf. Technol., 59/1: 21–24.
[216] Wang, J., Liu, S., Gao, R., Yan, R., 2012, Current Envelope Analysis for Defect Identification and Diagnosis in Induction Motors, J. Manuf. Syst., 31/4:380–387.
[217] Wang, J., Ma, Y., Zhang, L., Gao, R., Wu, D., 2018, Deep Learning for Smart Manufacturing: Methods and Applications, J. Manuf. Syst., 48: 144–156.
[218] Wang, L., 2019, From Intelligence Science to Intelligent Manufacturing, Engineering.
[219] Wang, L., Gao, R., Váncza, J., Krüger, J., Wang, X. V., Makris, S., Chryssolouris, G., 2019, Symbiotic Human-Robot Collaborative Assembly, CIRP Ann. 68/2: 701-726.
[220] Wang, L., Ji, W., 2018, Cloud Enabled CPS and Big Data in Manufacturing, Lecture Notes in Mechanical Engineering, 265–292.
[221] Wang, P., Ananya, Yan, R., Gao, R., 2017, Virtualization and Deep Recognition for System Fault Classification, J. Manuf. Syst., 44: 310–316.
[222] Wang, P., Gao, R., 2015, Adaptive Resampling-Based Particle Filtering for Tool Life Prediction, J. Manuf. Syst., 37: 528–534.
[223] Wang, P., Gao, R., 2016, Markov Nonlinear System Estimation for Engine Performance Tracking, J. Eng. Gas Turbines Power, 138/9: 091201.
[224] Wang, P., Gao, R., Yan, R., 2017, A Deep Learning-Based Approach to Material Removal Rate Prediction in Polishing, CIRP Ann. - Manuf. Technol., 66/1: 429–432.
[225] Wang, P., Liu, H., Wang, L., Gao, R., 2018, Deep Learning-Based Human Motion Recognition for Predictive Context-Aware Human-Robot Collaboration, CIRP Ann., 67/1:17–20.
[226] Wang, Q., Zhang, X., Wu, Y., Tang, L., and Zha, Z., 2017, Nonconvex Weighted ℓp Minimization Based Group Sparse Representation Framework for Image Denoising, IEEE Signal Process. Lett., 24/11:1686–1690.
[227] Wang, Y., Peng, Y., Zi, Y., Jin, X., Tsui, K. L., 2016, A Two-Stage Data-Driven-Based Prognostic Approach for Bearing Degradation Problem, IEEE Trans. Ind. Informatics, 12/3: 924–932.
[228] Wang, Z., Wang, J., Wang, Y., 2018, An Intelligent Diagnosis Scheme Based on Generative Adversarial Learning Deep Neural Networks and Its Application to Planetary Gearbox Fault Pattern Recognition, Neurocomputing, 310: 213–222.
[229] Waschneck, B., Reichstaller, A., Belzner, L., Altenmüller, T., Bauernhansl, T., Knapp, A., Kyek, A., 2018, Optimization of Global Production Scheduling with Deep Reinforcement Learning, Procedia CIRP, 72:1264–1269.
[230] Weimer, D., Scholz-Reiter, B., Shpitalni, M., 2016, Design of Deep Convolutional Neural Network Architectures for Automated Feature Extraction in Industrial Inspection, CIRP Ann. - Manuf. Technol., 65/1: 417–420.
[231] Wells, L. J., Camelio, J. A., Williams, C. B., and White, J., 2013, Cyber-Physical Security Challenges in Manufacturing Systems, Manuf. Lett., 2/1: 74–77.
[232] Wroblewski, R., Wroblewski.
[233] Wu, D., Jennings, C., Terpenny, J., Gao, R., and Kumara, S., 2017, A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests, J. Manuf. Sci. Eng. Trans. ASME, 139(7).
[234] Wuest, T., Weimer, D., Irgens, C., Thoben, K. D., 2016, Machine Learning in Manufacturing: Advantages, Challenges, and Applications, Prod. Manuf. Res., 4/1: 23–45.
[235] WZL RWTH Aachen, 2017, Neuartiger Edge Computing Demonstrator Zeigt Potenzial Für Resiliente Und Selbstoptimierende Produktion.
[236] Xu, X., Lei, Y., Li, Z., 2019, An Incorrect Data Detection Method for Big Data Cleaning of Machinery Condition Monitoring, IEEE Trans. Ind. Electron., 1.
[237] Yaga, D., Mell, P., Roby, N., Scarfone, K., 2018, Blockchain Technology Overview.
[238] Yavari, A., Jayaraman, P.P., Georgakopoulos, D., 2016, Contextualised service delivery in the Internet of Things, IEEE 3rd World Forum on Internet of Things, 454–459.
[239] Yavari, A., Jayaraman, P.P., Georgakopoulos, D., Nepal, D., 2017, ConTaaS: An Approach to Internet-Scale Contextualisation for Developing Efficient Internet of Things Applications. ScholarSpace. Proc. 50th Hawaii Int. Conf. on System Sciences 2017 (HICSS-50) 5932-5940.
[240] Ye, R., Pan, C. S., Chang, M., Yu, Q., 2018, Intelligent Defect Classification System Based on Deep Learning, Adv. Mech. Eng., 10/3.
[241] Yuan, H., Lu, C., 2017, Rolling Bearing Fault Diagnosis under Fluctuant Conditions Based on Compressed Sensing, Struct. Control Heal. Monit., 24/5.
[242] Zeng, A., Song, S., Lee, J., Rodriguez, A., Funkhouser, T., 2019, TossingBot: Learning to Throw Arbitrary Objects with Residual Physics, arXiv:1903.11239.
[243] Zhang, C., Yan, H., Lee, S., Shi, J., 2018, Multiple Profiles Sensor-Based Monitoring and Anomaly Detection, J. Qual. Technol., 50/4: 344–362.
[244] Zhang, J., Wang, P., Gao, R., 2019, Deep Learning-Based Tensile Strength Prediction in Fused Deposition Modeling, Comput. Ind., 107: 11–21.
[245] Zhang, J., Wang, P., Yan, R., Gao, R., 2018, Long Short-Term Memory for Machine Remaining Life Prediction, J. Manuf. Syst., 48:78–86.
[246] Zhang, R., Tao, H., Wu, L., Guan, Y., 2017, Transfer Learning with Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions, IEEE Access, 5:14347–14357.
[247] Zhang, Y., Hong, G. S., Ye, D., Zhu, K., Fuh, J. Y. H., 2018, Extraction and Evaluation of Melt Pool, Plume and Spatter Information for Powder-Bed Fusion AM Process Monitoring, Mater. Des., 156: 458–469.
[248] Zhang, Y., Xiong, R., He, H., Pecht, M. G., 2018, Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries, IEEE Trans. Veh. Technol., 67/7: 5695–5705.
[249] Zhao, C., Fezzaa, K., Cunningham, R. W., Wen, H., De Carlo, F., Chen, L., Rollett, A. D., Sun, T., 2017, Real-Time Monitoring of Laser Powder Bed Fusion Process Using High-Speed X-Ray Imaging and Diffraction, Sci. Rep., 7:3602.
[250] Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R., 2019, Deep Learning and Its Applications to Machine Health Monitoring, Mech. Syst. Signal Process., 115:213–237.
[251] Zhong, R. Y., Huang, G. Q., Lan, S., Dai, Q. Y., Chen, X., Zhang, T., 2015, A Big Data Approach for Logistics Trajectory Discovery from RFID-Enabled Production Data, Int. J. Prod. Econ., 165: 260–272.
[252] Zhong, R. Y., Lan, S., Xu, C., Dai, Q., Huang, G. Q., 2016, Visualization of RFID-Enabled Shopfloor Logistics Big Data in Cloud Manufacturing, Int. J. Adv. Manuf. Technol., 84/1-4: 5–16
[253] Zietsch, J., Büth, L., Juraschek, M., Weinert, N., Thiede, S., Herrmann, C., 2019, Identifying the Potential of Edge Computing in Factories through Mixed Reality, Procedia CIRP, 81:1095–1100