



# Machine learning in manufacturing: advantages, challenges, and applications

Thorsten Wuest<sup>a</sup> , Daniel Weimer<sup>b</sup>, Christopher Irgens<sup>c</sup> and Klaus-Dieter Thoben<sup>d</sup>

<sup>a</sup>Industrial and Management Systems Engineering, West Virginia University, Morgantown, WV 26506, USA;

<sup>b</sup>ICT Applications for Production, BIBA – Bremer Institut fuer Produktion und Logistik, 28359 Bremen, Germany; <sup>c</sup>Design, Manufacture & Engineering Management, University of Strathclyde, Glasgow G1 1XJ, UK;

<sup>d</sup>Department of Integrated Product Development, University of Bremen, 28359 Bremen, Germany

## ABSTRACT

The nature of manufacturing systems faces ever more complex, dynamic and at times even chaotic behaviors. In order to being able to satisfy the demand for high-quality products in an efficient manner, it is essential to utilize all means available. One area, which saw fast pace developments in terms of not only promising results but also usability, is machine learning. Promising an answer to many of the old and new challenges of manufacturing, machine learning is widely discussed by researchers and practitioners alike. However, the field is very broad and even confusing which presents a challenge and a barrier hindering wide application. Here, this paper contributes in presenting an overview of available machine learning techniques and structuring this rather complicated area. A special focus is laid on the potential benefit, and examples of successful applications in a manufacturing environment.

## ARTICLE HISTORY

Received 8 April 2016

Accepted 18 May 2016

## KEYWORDS

Manufacturing; machine learning; intelligent manufacturing systems; smart manufacturing

## 1. Introduction

The manufacturing industry today is experiencing a never seen increase in available data (Chand & Davis, 2010). These data comprise a variety of different formats, semantics, quality, e.g. sensor data from the production line, environmental data, machine tool parameters, etc. (Davis et al., 2015). Different names are used for this phenomenon, e.g. Industrie 4.0 (Germany), Smart Manufacturing (USA), and Smart Factory (South Korea). This increase and availability of large amounts of data is often referred to as Big Data (Lee, Lapira, Bagheri, & Kao, 2013). The availability of, e.g. quality-related data offers potential to improve process and product quality sustainably (Elangovan, Sakthivel, Saravanamurugan, Nair, & Sugumaran, 2015). However, it has been recognized that much information can also propose a challenge and may have a negative impact as it can, e.g. distract from the main issues/causalities or lead to delayed or wrong conclusions about appropriate actions (Lang, 2007). Overall, it can be safely concluded, the manufacturing industry has to accept that in order to benefit from the increased data availability, e.g. for quality improvement initiatives,

manufacturing cost estimation and/or process optimization, better understanding of the customer's requirements, etc., support is needed to handle the high dimensionality, complexity, and dynamics involved (Davis et al., 2015; Loyer, Henriques, Fontul, & Wiseall, 2016; Wuest, 2015).

New developments in certain domains like mathematics and computer science (e.g. statistical learning) and availability of easy-to-use, often freely available (software) tools offer great potential to transform the manufacturing domain and their grasp on the increased manufacturing data repositories sustainably. One of the most exciting developments is in the area of machine learning (incl. data mining (DM), artificial intelligence (AI), knowledge discovery (KD) from databases, etc.). However, the field of machine learning is very diverse and many different algorithms, theories, and methods are available. For many manufacturing practitioners, this represents a barrier regarding the adoption of these powerful tools and thus may hinder the utilization of the vast amounts of data increasingly being available.

In accordance to that, the paper aims to:

- argue from a manufacturing perspective why machine learning is an appropriate and promising tool for today's and future challenges;
- introduce the terminology used in the respective fields;
- present an overview of the different areas of machine learning and propose an overall structuring;
- provide the reader with a high-level understanding of the advantages and disadvantages of certain methods with respect to manufacturing application.

In the following section, the current challenges manufacturing faces are illustrated. This provides a basis for the later argumentation of machine learning being an appropriate tool to for manufacturers to face those challenges head on.

### **1.1. Challenges of the manufacturing domain**

Manufacturing is a very established industry, however the importance of it cannot be rated high enough. Several mature economies experienced a reduction of the manufacturing contribution toward their GDP over the last decades. However, in the last years, several initiatives to revamp the manufacturing sector were started. Examples are the US through 'Executive Actions to Strengthen Advanced Manufacturing in America' (White House, 2014) and the European Union with their 'Factories of the Future' (European Commission, 2016) initiative. The challenges manufacturing faces today are different from the challenges in the past.

There are several studies available proposing key challenges of manufacturing on a global level. The key challenges most of the researchers agree upon (Dingli, 2012; Gordon & Sohal, 2001; Shiang & Nagaraj, 2011; Thomas, Byard, & Evans, 2012) are the following:

- Adoption of advanced manufacturing technologies.
- Growing importance of manufacturing of high value-added products.
- Utilizing advanced knowledge, information management, and AI systems.
- Sustainable manufacturing (processes) and products.
- Agile and flexible enterprise capabilities and supply chains.
- Innovation in products, services, and processes.

- Close collaboration between industry and research to adopt new technologies.
- New manufacturing management paradigms.

These key challenges highlight the ongoing trend of the manufacturing domain to becoming more complex and dynamic. The apparent complexity is inherited not only in the manufacturing programs themselves but increasingly in the to-be-manufactured product as well as in the (business) processes of the companies and collaborative networks (Wiendahl & Scholtissek, 1994). Adding to the challenge is the fact that the dynamic business environment of today's manufacturing companies is affected by uncertainty (Monostori, 2003). Especially looking at domains most likely to being optimized, e.g. monitoring and control, scheduling and diagnostics, it becomes apparent that the increasing availability of data is adding another challenge: besides the large amounts of available data (e.g. sensor data), the high dimensionality and variety (e.g. due to different sensors or connected processes) of data as well as the NP complete nature of manufacturing optimization problems (Wuest, 2015) present a challenge.

To overcome some of today's major challenges of complex manufacturing systems, valid candidates are machine learning techniques. These data-driven approaches are able to find highly complex and non-linear patterns in data of different types and sources and transform raw data to features spaces, so-called models, which are then applied for prediction, detection, classification, regression, or forecasting.

In the following, first the main advantages and challenges of machine learning applications with regard to manufacturing, its challenges and requirements are illustrated. Then the current state of the art of machine learning, again with a focus on manufacturing applications is presented. Within that context, a structuring of different machine learning techniques and algorithms is developed and presented.

## ***1.2. Suitability of machine learning application with regard to today's manufacturing challenges***

Before looking into the suitability of machine learning (ML) based on the previously derived requirements toward a future solution approach, the used terms are briefly introduced. ML is known for its ability to handle many problems of NP-complete nature, which often appear in the domain of smart manufacturing (Monostori, Hornyák, Egresits, & Viharos, 1998).

The application of ML techniques increased over the last two decades due to various factors, e.g. the availability of large amounts of complex data with little transparency (Smola & Vishwanathan, 2008) and the increased usability and power of available ML tools (Larose, 2005). Nevertheless, the main definition of ML, allowing computers to solve problems without being specifically programmed to do so (Samuel, 1959) is still valid today. ML is connected to other terms, like DM, KD, AI, and others (Alpaydin, 2010). Today, ML is already widely applied in different areas of manufacturing, e.g. optimization, control, and troubleshooting (Alpaydin, 2010; Pham & Afify, 2005).

Many ML techniques (e.g. Support Vector Machine [SVM]) are designed to analyze large amounts of data and capable of handling high dimensionality (>1000) very well (Yang & Trewn, 2004). However, accompanying issues like possible over-fitting has to be considered (Widodo & Yang, 2007) during the application. If dimensionality proves to be an issue despite it being unlikely due to the power of the algorithms, there are methods available to

reduce the dimensions. These claim to reduce the impact of the reduction of the dimensionality on the expected results (Kotsiantis, 2007; Manning, Raghavan, & Schütze, 2009). The importance of using ML, in this case SVM is that dimensionality is not a practical problem and therefore the need for reducing dimensionality is reduced. This implies the possibility of being more liberal in including seemingly irrelevant information available in the manufacturing data that may turn out to be relevant under certain circumstances. This may have a direct effect on the existing knowledge gap described previously (Alpaydin, 2010; Pham & Afify, 2005).

Applying ML in manufacturing may result in deriving pattern from existing data-sets, which can provide a basis for the development of approximations about future behavior of the system (Alpaydin, 2010; Nilsson, 2005). This new information (knowledge) may support process owners in their decision-making or be used automatically to improve the system directly. In the end, the goal of certain ML techniques is to detect certain patterns or regularities that describe relations (Alpaydin, 2010).

Given the challenge of a fast changing, dynamic manufacturing environment, ML, being part of AI and inherit the ability to learn and adapt to changes ‘the system designer need not foresee and provide solutions for all possible situations’ (Alpaydin, 2010). Therefore, ML provides a strong argument why its application in manufacturing may be beneficial given the struggle of most first-principle models to cope with the adaptability. Learning from and adapting to changing environments automatically is a major strength of ML (Lu, 1990; Simon, 1983).

ML techniques are designed to derive knowledge out of existing data (Alpaydin, 2010; Kwak & Kim, 2012). Alpaydin (2010) emphasizes that ‘stored data becomes useful only when it is analyzed and turned into information that we can make use of, for example, to make predictions’ (Alpaydin, 2010). This is especially true for manufacturing, given the struggle of obtaining real-time data during a live manufacturing program run with the technical, financial, and knowledge restrictions. This may also have an impact on issue of positioning of process checkpoints (Wuest, Liu, Lu, & Thoben, 2014). Whereas, it makes sense to select carefully checkpoints under the perspective of what data are useful, it may be obsolete given the analytical power of ML techniques to derive information from formerly considered useless data. This may result in the ability to determine more states, to capture data, along the overall manufacturing program. Whether this is beneficial is an open question, which has to be researched. Given the ability of ML to handle high-dimensionality data, the technical side of analyzing the additional data provides no problem. However, in terms of capturing data it may still be a problem, specifically the ability to capture the data. Once the data are available, determining state drivers in very high-dimensionality situations is not considered problematic, nor is repeating it frequently.

In the following table, a summary of the theoretical ability of ML techniques to answer the main challenges of manufacturing applications (requirements) is presented (Table 1).

Overall, as Monostori, Márkus, Van Brussel, and Westkämper (1996) emphasize, ‘intelligence is strongly connected with learning, and learning ability must be an indispensable feature of Intelligent Manufacturing Systems.’ ML provides strong arguments when it comes to the limitations and challenges the theoretical *product state concept* faces. Given the above-stated analysis, ML techniques seem to provide a promising solution based on the derived requirements. Most of the identified requirements are successfully addressed by ML.

**Table 1.** Summary of suitability of ML techniques in manufacturing application.

Manufacturing requirement	Theoretical ability of ML to meet requirements
Ability to handle high-dimensional problems and data-sets with reasonable effort	Certain ML techniques (e.g. SVM) are capable of handling high dimensionality (>1000) very well. However, accompanying issues like possible over-fitting has to be considered (Widodo & Yang, 2007; Yang & Trewn, 2004)
Ability to reduce possibly complex nature of results and present transparent and concrete advice for practitioners (e.g. monitor XX and parameter YY at checkpoint ZZ)	ML may be able to derive pattern from existing data and derive approximations about future behavior (Alpaydin, 2010). This new information (knowledge) may support process owners in their decision-making or used to automatically improve a system
Ability to adapt to changing environment with reasonable effort and cost. Ideally a degree auf 'automated' adaptation to changing condition	As ML is part of AI, and thus be able to learn and adapt to changes, 'the system designer need not foresee and provide solutions for all possible situations' (Alpaydin, 2010). Learning from and adapting to changing environments automatically is a major strength of ML (Lu, 1990; Simon, 1983)
Ability to further the existing knowledge by learning from results	ML can contribute to create new information and possibly knowledge by, e.g. identifying patters in existing data (Alpaydin, 2010; Pham & Afify, 2005)
Ability to work with the available manufacturing data without special requirements toward capturing of very specific information at the start	ML techniques are designed to derive knowledge out of existing data (Alpaydin, 2010; Kwak & Kim, 2012). 'The stored data becomes useful only when it is analyzed and turned into information that we can make use of, for example, to make predictions' (Alpaydin, 2010)
Ability to identify relevant process intra- and inter-relations & ideally correlation and/or causality	The goal of certain ML techniques is to detect certain patterns or regularities that describe relations (Alpaydin, 2010)

However, a more detailed analysis of available ML techniques as well as their strengths and limitations concerning the requirements has to be provided. Most of all, the possible compatibility with the theoretical product state concept and its perspective on the manufacturing program has to be elaborated further before a final judgment can be given. Furthermore, there are many questions to be answered like how ML techniques may handle qualitative information.

In the next section, the advantages and challenges of machine learning application in manufacturing are introduced based on the previous presented requirements.

## 2. Advantages and challenges of machine learning application in manufacturing

ML has been successfully utilized in various process optimization, monitoring and control applications in manufacturing, and predictive maintenance in different industries (Alpaydin, 2010; Gardner & Bicker, 2000; Kwak & Kim, 2012; Pham & Afify, 2005; Susto, Schirru, Pampuri, McLoone, & Beghi, 2015). ML techniques were found to provide promising potential for improved quality control optimization in manufacturing systems (Apte, Weiss, & Grout, 1993), especially in 'complex manufacturing environments where detection of the causes of problems is difficult' (Harding, Shahbaz, & Kusiak, 2006). However, often ML applications are found to be limited focusing on specific processes instead of the whole manufacturing program or manufacturing system (Doltsinis, Ferreira, & Lohse, 2012).

There are many different ML methods, tools, and techniques available, each with distinct advantages and disadvantages. The domain of ML has grown to an independent research

domain. Therefore, within this section, the goal is to find a suitable ML technique for application in manufacturing.

### 2.1. Advantages of machine learning application in manufacturing

The general advantages of ML have been established in previous sections stating that ML techniques are able to handle NP complete problems which often occur when it comes to optimization problems of intelligent manufacturing systems (Monostori et al., 1998). In the following, the focus is on the ability of ML techniques to handle high-dimensional, multi-variate data, and the ability to extract implicit relationships within large data-sets in a complex and dynamic, often even chaotic environment (Köksal, Batmaz, & Testik, 2011; Yang & Trewn, 2004). 'Since most engineering and manufacturing problems are data-rich but knowledge-sparse' (Lu, 1990), ML provides a tool to increase the understanding of the domain. In this section, the advantages are presented in an attempt of generalization for ML in total. However, it has to be understood, that the peculiarity of the advantages may differ depending on the chosen ML technique.

Overall it is agreed upon that ML allows to reduce cycle time and scrap, and improve resource utilization in certain NP-hard manufacturing problems. Furthermore, ML provides powerful tools for continuous quality improvement in a large and complex process such as semiconductor manufacturing (Monostori et al., 1998; Pham & Afify, 2005).

An advantage of ML algorithms is the ability to handle *high dimensional problems and data*. Especially with regard to the increasing availability of complex data (Yu & Liu, 2003) with little transparency in manufacturing (Smola & Vishwanathan, 2008), this will most likely become even more important in the future. However, as is true for most advantages and disadvantages of ML algorithms, this cannot be generalized. Some algorithms (e.g. SVM; Distributed Hierarchical Decision Tree) can handle high dimensionality better than others (Bar-Or, Wolff, Schuster, & Keren, 2005; Do, Lenca, Lallich, & Pham, 2010). As was stated previously, in manufacturing mostly those ML algorithms are applicable that are capable of handling high-dimensional data. Therefore, the ability to cope with high dimensionality is considered an advantage of ML application in manufacturing.

Another advantage of ML techniques is the increased *usability of application* of algorithms due to (often source) programs like Rapidminer. This allows (relatively) easy application in many cases and furthermore comfortable adjustment of parameters to increase the classification performance.

As previously stated, a major advantage of ML algorithms is to *discover formerly unknown (implicit) knowledge* and to identify implicit relationships in data-sets. Depending on the characteristic of the ML algorithm (supervised/unsupervised or Reinforcement Learning [RL]), the requirements toward the available data may vary. However, the overall ability of ML algorithm to achieve results in a manufacturing environment was successfully proven (e.g. Alpaydin, 2010; Filipic & Junkar, 2000; Guo, Sun, Li, & Wang, 2008; Kim, Kang, Cho, Lee, & Doh, 2012; Nilsson, 2005).

Given the specific nature of manufacturing systems being dynamic, uncertain, and complex. Here, ML algorithms provide the opportunity to learn from the dynamic system and adapt to the changing environment automatically to a certain extent (Lu, 1990; Simon, 1983). The adaptation is, depending on the ML algorithm, reasonably fast and in almost all cases faster than traditional methods.



Applying ML in manufacturing may result in deriving pattern from existing data-sets, which can provide a basis for the development of approximations about future behavior of the system (Alpaydin, 2010; Nilsson, 2005). This new information (knowledge) may support process owners in their decision-making or used to automatically improve the system directly. In the end, the goal of certain ML techniques is to detect certain patterns or regularities that describe relations (Alpaydin, 2010).

Kotsiantis (2007) compared several algorithms according to their specific performance in manufacturing application by different attributes. Even so, this presents the opportunity to get a first impression, it is not suggested to base the decision for a suitable ML algorithm solely on comparisons as presented in such a table. Each problem is different and the performance of each algorithm also depends on the data available and data pre-processing as well as the parameter settings. The best fitting algorithm has to be found in testing various ones in a realistic environment. This is discussed further in the next section.

## **2.2. Challenges of machine learning application in manufacturing**

A very common challenge of ML application in manufacturing is the acquisition of relevant data. This is also a limitation as the availability, quality, and composition (e.g. are meta-data included? are data labeled?) of the manufacturing data at hand have a strong influence on the performance of ML algorithms. Some challenges the data-set can contain are, e.g. high-dimensional data can represent for some ML algorithms, that is, it can contain a high degree of irrelevant and redundant information which may impact the performance of learning algorithms (Yu & Liu, 2003). Today, most machine learning techniques handle only data with continuous and nominal values (Pham & Afify, 2005). How significant the influence is, depends on various factors including the algorithm itself and the parameter settings. It can be considered a general challenge for most research in manufacturing and not only ML application, to get hold of any data due to, e.g. security concerns or a basic lack of data capturing during the process. Even though in most cases ML allows the extracting of knowledge and generates better results than most traditional methods with less requirements toward available data, certain aspects concerning the available data that can prevent the successful application still have to be considered. Together with the next point, this highlights the increased need to understand the data in order to apply ML. Hoffmann (1990) highlights that compared to traditional methods where a lot of time is spent to extract information, in ML a lot of time is spent on preparing the data.

After the available data are secured, the data often have to be pre-processed depending on the requirements of the algorithm of choice. *Pre-processing of data* has a critical impact on the results. However, there are many standardized tools available which support the most common pre-processing processes like normalizing and filtering the data. Also it has to be checked whether the training data are unbalanced. This can present a challenge for the training of certain algorithms. In manufacturing practice, it is a common problem that values of certain attributes are not available or missing in the data-set (Pham & Afify, 2005). These so-called missing values present a challenge for the application of ML algorithms. There are certain practical induction systems available which may fill the gap (Pham & Afify, 2005). However, each problem and later applied ML algorithm have specific requirements when it comes to replacing missing values. By replacing missing values, the original data-set is influenced. The goal is to reduce the bias and other negative influence as much as possible

in respect to the analysis goal. As this issue represents a very common challenge, there is a large amount of literature and practical solutions (e.g. in R) available (e.g. Graham, 2012; Kabacoff, 2011; Kwak & Kim, 2012; Li & Huang, 2009).

A major challenge of increasing importance is the question what ML technique and algorithm to choose (*selection of ML algorithm*). Even so, there were attempts to pursue the definition of 'general ML techniques,' the diverse problems and their requirements highlight the need for specialized algorithms with certain strength and weaknesses (Hoffmann, 1990). Especially due to the increased attention of practitioners and researchers for the field of ML in manufacturing, a large number of different ML algorithms or at least variations of ML algorithms is available. Adding to this already existing complexity, combinations of different algorithms, so-called 'hybrid approaches,' are becoming more and more common promising better results than 'individual' single algorithm application (e.g. Lee & Ha, 2009). Many studies are available highlighting a successful application of ML techniques for specific problems. At the same time the test data are not publically available in many cases. This makes a neutral and unbiased assessment of the results and therefore a final comparison challenging. As of today, the generally accepted approach to select a suitable ML algorithm for a certain problem is as follows:

- First, one looks at the available data and how it is described (labeled, unlabeled, available expert knowledge, etc.) to choose between a supervised, unsupervised, or RL approach.
- Secondly, the general applicability of available algorithms with regard to the research problem requirements (e.g. able to handle high dimensionality) has to be analyzed. A specific focus has to be laid on the structure, the data types, and overall amount of the available data, which can be used for training and evaluation.
- Thirdly, previous applications of the algorithms on similar problems are to be investigated in order to identify a suitable algorithm. The term 'similar' in this case means, research problems with comparable requirements e.g. in other disciplines or domains.

Another challenge is the *interpretation of the results*. It has to be taken into account that not only the format or illustration of the output is relevant for the interpretation but also the specifications of the chosen algorithm itself, the parameter settings, the 'planned outcome' and also the data including its pre-processing. Within the interpretation of the results, certain more distinct limitations (again depending on the chosen algorithm) can have a large impact. Among those are, e.g. immune to over-fitting (Widodo & Yang, 2007), bias, and variance (therefore bias–variance tradeoff) (Quadrianto & Buntine, 2011).

### 3. Structuring of machine learning techniques and algorithms

As previously stated, ML has developed into a wide and divers field of research over the past decades. This has led to a variety of different sub-domains, algorithms, theories, and application areas, etc. The relationship and structure between the different elements are not commonly agreed upon. Different researchers choose different approaches to structure the field. In Figure 1, the authors try to structure the ML domain of DM according to tasks on the one side and available algorithms on the other (Corne, Dhaenens, & Jourdan, 2012). This structure highlights the importance of differentiation of task (what is the goal) and algorithm (how can that goal be reached) within the ML field.



However, the presented overview in Figure 1 is falling short by not reflecting the commonly accepted differentiation of ML methods by the available feedback in supervised, unsupervised, and RL (Monostori, 1993; Kotsiantis, 2007; Monostori, 2003; Pham & Afify, 2005). Monostori (2003) described the three classes as follows:

- ‘reinforcement learning: less feedback is given, since not the proper action, but only an evaluation of the chosen action is given by the teacher;
- unsupervised learning: no evaluation [label] of the action is provided, since there is no teacher;
- supervised learning: the correct response [label] is provided by a teacher.’

This structure is widely accepted, however, there are still differences with regard to what falls under them or what these three classes fall under. For example, Pham and Afify (2005) map supervised, unsupervised, and RL as part of Neural Networks (NN) (see Figure 2). However, Pham and Afify (2005) also state that they only focus on supervised classification learning methods. This would correspond with Lu (1990) who states that inductive learning

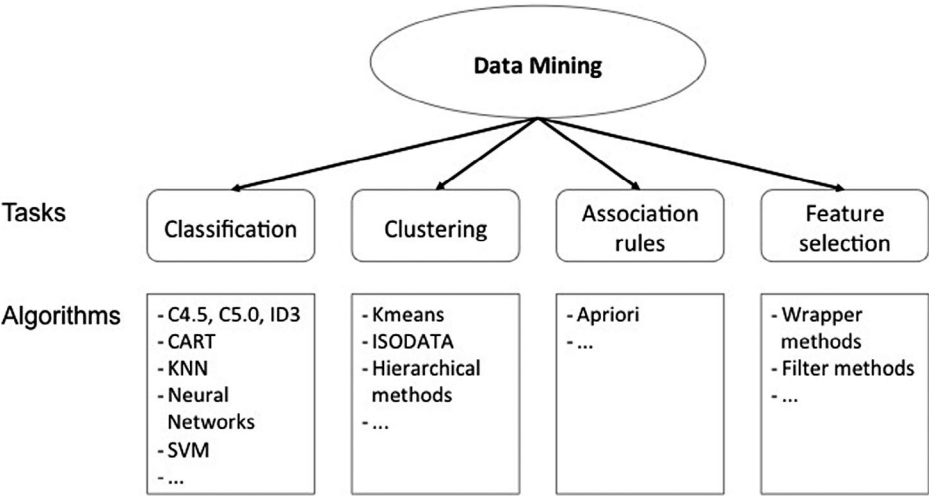


Figure 1. An overview of tasks and main algorithms in DM (Corne et al., 2012).

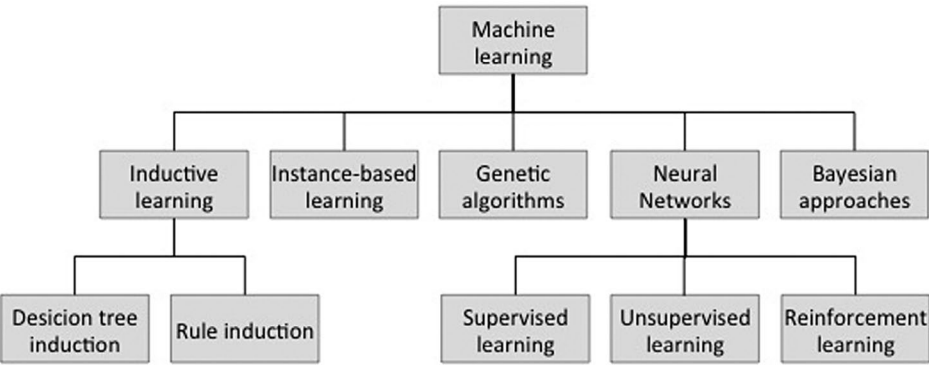


Figure 2. Classification of main ML techniques according to Pham and Afify (2005).

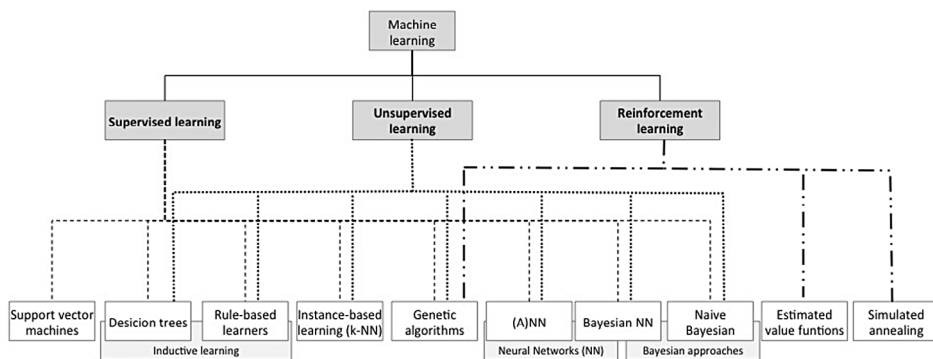
can be grouped in supervised and unsupervised learning. Other researchers differentiate between active and passive learning, stating that ‘active learning is generally used to refer to a learning problem or system where the learner has some role in determining on what data it will be trained’ (Cohn, 2011) whereas passive learning describes a situation where the learner has no control over the training set. Apparently, active learning is often used for problems where it is difficult (expensive and/or time-consuming) to obtain labeled training data. The advantage is to being able to achieve good performance needing less training data than other learners due to the sequentially identified useful examples by the active learner (Cohn, 2011). Active learning is mostly applied within supervised ML scenarios but was also found to be of valuable within certain RL problems (Cohn, 2011).

Some researchers like Kotsiantis (2007) focus only on supervised classification techniques and group NN as a learning algorithm as part of supervised learning. However, NN algorithms can also be applied in unsupervised learning and RL (Carpenter & Grossberg, 1988; Pham & Afify, 2005). This corresponds basically with Pham and Afify (2005), when the notion on top of the hierarchy is seen as ‘Supervised ML’ instead of the ‘Machine learning’ they originally stated.

An adapted and extended structuring of ML techniques and algorithms may be illustrated as follows:

Figure 3 does not include all available algorithms and algorithm variations. The purpose is to show the complex structure and the diverse nature of currently available and common ML techniques. Whereas the first selection of the main differentiation, supervised, unsupervised, and RL, suitable for the presented problem is in most cases possible, this is not necessarily the case when going further down the hierarchy. Additionally, it has to be kept in mind, that the different algorithms can be combined to maximize the classification power (Bishop, 2006). Pham and Afify (2005) state that ‘most of the existing machine-learning methods for generating multiple models can improve significantly on the accuracy of single models’ (Pham & Afify, 2005). That increases the complexity one has to face when in the process of selecting a suitable ML algorithm for a given problem, and thus the comprehensibility is hindered (Pham & Afify, 2005). Another interesting aspect is that many algorithms are applicable in both supervised and unsupervised learning (in adapted form).

The different algorithms and combinatory approaches often tend to be adapted to special problems. This makes it hard to compare them especially against their classification power for the given problem. A first indication can be comparing charts as can be found



**Figure 3.** Structuring of ML techniques and algorithms.

in Kotsiantis (2007). However, a more promising approach to select a suitable algorithm is to look for problems of similar nature and analyze what ML algorithm was used to solve it and what where the results. This is a good starting point. Once the algorithm is applied to the problem and first results are available, different methods can be applied and the results for the given problem can be compared. Modern computer tools support different kernels and make the switch (relatively) comfortable.

In the following, unsupervised machine learning, RL, and supervised machine learning are briefly described to being able to differentiate them from one another. Supervised machine learning later described in greater detail as it was found to have the best fit for challenges and problems faced in manufacturing applications and as manufacturing data is often labeled, meaning expert feedback is available (Lu, 1990).

### **3.1. Unsupervised machine learning**

Unsupervised machine learning is another large area of research. The defining attribute is that within unsupervised learning, there is no feedback from an external teacher/knowledgeable expert. The algorithm itself is supposed to identify clusters from existing data based on, e.g. conceptual cohesiveness of attributes (Lu, 1990). Kotsiantis (2007) introduced the rule that if instances are unlabeled (no known labels and corresponding correct outputs), it is most likely unsupervised learning. The goal is to discover unknown classes of items by clustering (Jain, Murty, & Flynn, 1999) whereas supervised learning is focused on classification (known labels). Basically, unsupervised ML describes any ML process that tries to learn 'structure in the absence of either an identified output [e.g. supervised ML] or feedback [e.g. RL]. Three typical examples of unsupervised learning are clustering, association rules, and self-organizing maps' (Sammut & Webb, 2011).

Especially in the Big Data context, unsupervised methods are becoming increasingly important. However, as in manufacturing application, the main assumption is that knowledgeable experts can provide feedback on the classification of states to identify the learning set in order to train the algorithm (Lu, 1990; Monostori, 2003). Thus, the focus will be laid on supervised methods. However, some aspects of unsupervised learning may be beneficial in manufacturing application after all. First, there is the possibility that in some cases there might be no expert feedback available or, in the future, desirable. Another aspect is to realize hybrid approaches, combining the 'best of both worlds' which gain importance due to the fast increase in unlabeled data especially in manufacturing (Kang, Kim, & Cho, 2016). And finally, unsupervised methods can be and are being used to, e.g. identify outliers in manufacturing data (Hansson, Yella, Dougherty, & Fleyeh, 2016).

### **3.2. Reinforcement learning**

RL is defined by the provision of the training information by the environment. The information on how well the system performed in the respective turn is provided by a numerical reinforcement signal (Kotsiantis, 2007). Another defining characteristic is that the learner has to uncover which actions generate the best results (numerical reinforcement signal) by trying instead of being told. This distinguishes RL from most of the other ML methods (Sutton & Barto, 2012). However, RL is seen by some researchers as 'a special form of supervised learning' (Pham & Afify, 2005). However, different from supervised learning

problems, RL problems can be described by the absence of labeled examples of ‘good’ and ‘bad’ behavior (Stone, 2011). RL, based on sequential environmental response, emulates the process of learning of humans (Wiering & Van Otterlo, 2012). This ‘reward signal,’ which can be perceived in RL differentiates it from unsupervised ML (Stone, 2011). Different from supervised learning, RL is most adequate in situation where there is no knowledgeable supervisor. In such uncharted territory, an agent is needed to being able to learn from interaction and its own experience – this is where RL can utilize its advantages (Sutton & Barto, 2012).

As RL is based on feedback of actions, one interesting and also challenging issue is that certain actions have not or not only an immediate impact, but certain effects might show at a later time and/or during a following additional trial. Overall, RL ‘is defined not by characterizing learning methods, but by characterizing a learning problem. Any method that is well suited to solving that problem, [might be considered] to be a reinforcement learning method’ (Sutton & Barto, 2012).

A very specific challenge for RL is the tradeoff between exploration and exploitation. In order to achieve the goal, the agent has to ‘exploit’ the actions it learned to prefer and to identify those it has to ‘explore’ by actively trying new ways (Sutton & Barto, 2012). In manufacturing, RL is not widely applied and just a few examples of successful application exist as of today (Doltsinis et al., 2012; Günther, Pilarski, Helfrich, Shen, & Diepold, 2015). In the majority of manufacturing applications today, expert feedback is available. Therefore, even though RL is applicable in manufacturing applications, the focus in the following is on supervised techniques.

### 3.3. Supervised machine learning

In manufacturing application, supervised ML techniques are mostly applied due to the data-rich but knowledge-sparse nature of the problems (Lu, 1990). In addition, supervised ML may benefit from the established data collection in manufacturing for statistical process control purposes (Harding et al., 2006) and the fact that these data are mostly labeled. Basically, supervised ML ‘is learning from examples provided by a knowledgeable external supervisor’ (Sutton & Barto, 2012). This is partly due to the availability of (a) expert feedback (e.g. quality) and (b) the labeled instances. Supervised ML is applied in different domains of manufacturing, monitoring, and control being a very prominent one among them (e.g. Alpaydin, 2010; Apte et al., 1993; Harding et al., 2006; Kwak & Kim, 2012; Pham & Afify, 2005).

The general process of supervised ML contains several steps handling the data and setting up the training and test data-set by the teacher, hence supervised (Kotsiantis, 2007). Based on a given problem, the required data are identified and (if needed) pre-processed. An important aspect is the definition of the training set, as it influences the later classification results to a large extent. Even so it often appears as if the algorithm selection is always following the definition of the training data-set, the definition of the training data also has to take the requirements of the algorithm selection into account. Some algorithms allow for a so-called ‘kernel selection’ to adapt the algorithm to the specific nature of the problem. This highlights the adaptability of ML application and the variety of problems that can be tackled.

Similar requirements stand to some extent also true for the identification and pre-processing of the data as different algorithms have certain strength and weaknesses concerning the

handling of different data-sets (e.g. format, dimensions, etc.). After an algorithm is selected, it is trained using the training data-set. In order to judge the ability to perform the targeted task, the trained algorithm is then evaluated using the evaluation data-set. Depending on the performance of the trained algorithm with the evaluation data-set, the parameters can be adjusted to optimize the performance in the case the performance is already good. In case the performance is not satisfying, the process has to be started over at an earlier stage, depending on the actual performance.

A rule of thumb is that 70% of the data-set is used as a training data-set, 20% as an evaluation data-set (in order to adjust the parameters – e.g. bias) and final 10% as a test data-set.

In the following section, supervised learning algorithms are illustrated in more detail as they are the most commonly used algorithms in manufacturing application today. A major reason being the availability of ‘labels’ based on quality inspections in many manufacturing application.

#### 4. Supervised machine learning algorithms in manufacturing application

As can be seen in the previously presented figures, there are several supervised ML algorithms available. Each of these algorithms has specific advantages and limitations concerning the application in manufacturing. A major challenge is to select a suitable algorithm for the requirements of the manufacturing research problem at hand. First, the general applicability of a ML algorithm with the requirements may be derived from more general comparisons (e.g. presented by Kotsiantis (2007)). However, due to the individual nature, most research problems represent the specific characteristics of ML algorithms as well as their adapted ‘siblings’, it is not advisable to base the decision for a ML algorithm solely on such a theoretical and general selection. In order to being able to identify a suitable ML algorithm for the problem at hand, the next step involves a careful analysis of previous applications of ML algorithms on research problems with similar requirements. The research problems do not have to be located within the same domain, the major issue in this selection is the matching of the identified requirements, in this case the ability to handle multi-variate, high-dimensional data-sets and the ability to continuously adapt to changing environments (updating the learning set). A brief presentation of the main advantages and limitations of the different ML algorithms is presented in order to pre-select a group of potentially suitable techniques.

A very promising and fitting supervised ML algorithm for manufacturing research problem is *Statistical Learning Theory* (SLT). Within the theory of supervised learning, meaning the training of a machine to enable it (without being explicitly programmed) to choose a (performing) function describing the relation between inputs and output (Evgeniou, Pontil, & Poggio, 2000). SLT focuses on the question of ‘how well the chosen function generalizes, or how well it estimates the output for previously unseen inputs’ (Evgeniou et al., 2000). Several more practical algorithms are based on the theoretical background of SLT, e.g. NNs, SVMs, and Bayesian modeling (Brunato & Battiti, 2005). A major advantage of SLT algorithms is the variety of possible application scenarios and possible application strategies (Evgeniou, Poggio, Pontil, & Verri, 2002). SLT allows to reduce the number of needed samples in certain cases (Koltchinskii, Abdallah, Ariola, & Dorato, 2001). SLT is also able to overcome issues like observer variability better than other methods (Margolis, Land, Gottlieb, & Qiao, 2011). In some other cases, SLT still needs a large number of samples to perform (Cherkassky & Ma, 2009; Koltchinskii et al., 2001). Another challenge



for the application of SLT is the likelihood of over-fitting in some realizations (Evgeniou et al., 2002). However, Steel (2011) found that the Vapnik–Chernovnenkis dimension is a good predictor for the chance of over-fitting using STL. Furthermore, the computational complexity is not eliminated using SLT but rather avoided by relaxing design questions (Koltchinskii et al., 2001).

*Bayesian Networks* (BNs) may be defined as a graphical model describing the probability relationship among several variables (Kotsiantis, 2007). BNs are among the most well-known applications of SLT (Brunato & Battiti, 2005). Naïve Bayesian Networks represent a rather simple form of BNs, being composed of directed acyclic graphs (one parent, multiple children) (Kotsiantis, 2007). Among the advantages of BN are the limited storage requirements, the possibility to use it as an incremental learner, its robustness to missing values, and the easiness to grasp output. However, the tolerance toward redundant and interdependent attributes is understood to be very limited (Kotsiantis, 2007).

*Instance-Based Learning* (IBL) (Kang & Cho, 2008; Okamoto & Yugami, 2003) or *Memory-Based Reasoning* (MBR) (Kang & Cho, 2008) are mostly based on k-nearest neighbor (k-NN) classifiers and applied in, e.g. regression and classification (Kang & Cho, 2008). Even though IBL/MBR techniques have proven to achieve high accuracy of classification in some cases (Akay, 2011), a stable and good performance (Gagliardi, 2011; Zheng, Li, & Wang, 2010) and were found to be applicable in many different domains (Dutt & Gonzalez, 2012), when looking at the previously identified requirements they seem not to be the best match. Reasons why IBL/MBR are excluded from further investigation are, among other things, their difficulty to set the attribute weight vector in little known domains (Hickey & Martin, 2001), the complicated calculations needed if large numbers of training instances/test patterns and attributes are involved (Kang & Cho, 2008; Okamoto & Yugami, 2003), less adaptable learning procedures (tends to over-fitting with noisy data) (Gagliardi, 2011), task-dependency (Dutt & Gonzalez, 2012; Gonzalez, Dutt, & Lebiere, 2013), and time-sensitive to complexity (Gonzalez et al., 2013).

NN or Artificial Neural Networks are inspired by the functionality of the brain. The brain is capable of performing impressive tasks (e.g. vision, speech recognition), tasks that may prove beneficial in engineering application when transferred to a machine/artificial system (Alpaydin, 2010). NN simulate the decentralized ‘computation’ of the central nervous system by parallel processing (in reality or simulated) and allow an artificial system to perform unsupervised, reinforcement, and supervised learning tasks (e.g. pattern recognition) (Corne et al., 2012; Pham & Afify, 2005). Decentralization makes use of a high ‘number of simple, highly interconnected processing elements or nodes and incorporates the ability to process information by a dynamic response of these nodes and their connections to external inputs’ (Cook, Zobel, & Wolfe, 2006). These NN play an important role in today’s ML research (Nilsson, 2005). Today’s application of NN can be seen as being on the representation and algorithm level (Alpaydin, 2010). NN are applied in various fields of manufacturing (e.g. semiconductor manufacturing) and diverse problems (e.g. process control) (Harding et al., 2006; Lee & Ha, 2009; Wang, Chen, & Lin, 2005) which highlights their main advantage: their wide applicability (Pham & Afify, 2005). Besides the wide applicability, NN are capable of handling high-dimensional and multi-variate data on a similar rate to the later introduced SVM (Kotsiantis, 2007). Manallack and Livingstone (1999) found NN to ‘offer high accuracy in most cases but can suffer from over-fitting the training data’ (Manallack & Livingstone, 1999). However, in order to achieve the high accuracy, a



large sample size is required by NN (similar to SVM) (Kotsiantis, 2007). Over-fitting, connected to the high-variance algorithms is commonly accepted as a drawback of NN (again partly similar to SVMs) (Kotsiantis, 2007). Other challenges of applying NN include the complexity of the models they produce, the intolerance concerning missing values and the (often) time-consuming training (Kotsiantis, 2007; Pham & Afify, 2005).

The previously described SLT builds the theoretical foundation of a rather new and very promising ML algorithm that attracts increasing attention in recent years due to its generally high performance, ability to achieve high accuracy, and ability to handle high-dimensional, multi-variate data-sets – SVM. SVMs were introduced by Cortes and Vapnik (1995) as a new machine learning technique for two-group classification problems. Burbidge, Trotter, Buxton, and Holden (2001) found SVM to be a ‘robust and highly accurate intelligent classification technique well suited for structure–activity relationship analysis.’ SVM can be understood as a practical methodology of the theoretical framework of STL (Cherkassky & Ma, 2009). SVMs have a proven track record for successfully dealing with non-linear problems (Li, Liang, & Xu, 2009). The idea behind it is that input vectors are non-linearly mapped to a very high-dimensional feature space (Cortes & Vapnik, 1995). SVM can be combined with different kernels and thus adapt to different circumstances/requirements (e.g. NNs; Gaussian) (Keerthi & Lin, 2003). SVM as a classification technique has its roots in SLT (Khemchandani & Chandra, 2009; Salahshoor, Kordestani, & Khoshro, 2010) and has shown promising empirical results in a number of practical manufacturing applications (Chinnam, 2002; Widodo & Yang, 2007) and works very well with high-dimensional data (Azadeh et al., 2013; Ben-hur & Weston, 2010; Salahshoor et al., 2010; Sun, Rahman, Wong, & Hong, 2004; Wu, 2010; Wuest, Irgens, & Thoben, 2014). Current literature suggests that the performance of SVM compared to other ML methods is still very competitive (Jurkovic, Cukor, Brezocnik, & Brajkovic, 2016). Another aspect of this approach is that it represents the decision boundary using a subset of the training examples, known as the support vectors.

*Ensemble Methods* are a class of machine learning algorithms that combine a weighted committee of learners to solve a classification or regression problem. The committee or ensemble contains a number of base learners like NNs, trees, or nearest neighbor (Dietterich, 2000; Opitz & Maclin, 1999). In many cases, the base learners are from the same algorithm family, which is called a homogeneous ensemble. In contrast to that, a heterogeneous example is constructed by combining base learners of different types. For many machine learning problems, it is demonstrated that the ensemble leads to a better model generalization compared to a single base classifier (Zhou, 2012).

To construct the base classifiers, two main paradigms have demonstrated their predictive power. On the one hand, sequential ensemble methods use the output from a base classifier as an input of the following base classifier and therefore boost the output in a sequential way. AdaBoost, introduced by Freund and Schapire (1995), is a well-known example, where simple decision stumps are combined toward a complex boosting cascade. On the other hand, parallel adjustment of base classifiers leads to independent models, which is also named *Bagging*. One famous example of bagging methods is Random Forest (Breiman, 2001), which is a combination of randomly sampled tree predictors. In a first step, Random forest randomly selects a subset of the features space, and then performs a conventional split selection procedure within the selected feature subset.

*Deep Machine Learning* is a new area of machine learning that allows the processing of data in multiple processing layers toward highly non-linear and complex feature

representations. The field is mainly driven by the computer vision and language processing domain (LeCun, Bengio, & Hinton, 2015) but offers great potential to also boost data-driven manufacturing applications. Deep Convolutional Neural Networks (ConvNets) have demonstrated outstanding prediction performance in various fields of computer vision and won several contests, e.g. (Krizhevsky, Sutskever, & Hinton, 2012). In contrast to standard NNs, where each neuron from layer  $n$  is connected to all neurons in layer  $(n - 1)$ , a ConvNet is constructed by multiple filter stages with a restricted view and therefore well suited for image, video, and volumetric data (LeCun et al., 1989). From layer to layer, a ConvNet transforms the output of the previous layer in a higher abstraction by applying non-linear activation.

In manufacturing scenarios, data streams or data with temporal behavior are of major importance. Especially deep recurrent neural nets have demonstrated the ability to model temporal patterns, e.g. in time series data. Here, an important concept is the Long-Short-Term Memory Model which is a more general architecture of deep NNs (Hochreiter & Schmidhuber, 1997).

## 5. Application areas of supervised machine learning in manufacturing

As was illustrated in the previous section, there is a wide variety of different ML algorithms available. Each of them has specific advantages and disadvantages. In order to give an overview of successful applications of ML in manufacturing systems, selected applications of an exemplary supervised machine learning algorithm, SVMs, are illustrated.

A major application area of SVM in manufacturing is monitoring (Chinnam, 2002). Especially *tool/machine condition monitoring*, fault diagnosis, and tool wear are domains where SVM is continuously and successfully applied (Azadeh et al., 2013; Salahshoor et al., 2010; Sun et al., 2004; Widodo & Yang, 2007). Also quality monitoring in manufacturing is a field where SVMs were successfully applied (Ribeiro, 2005).

An application area of SVM with an overlap to manufacturing application is *image recognition* (e.g. character and face recognition) (Salahshoor et al., 2010; Widodo & Yang, 2007; Wu, 2010). In manufacturing, this can be utilized to identify (classify) damaged products (e.g. surface roughness) (Çaydaş & Ekici, 2010). Other application areas are, e.g. handwriting classification (Scheidat, Leich, Alexander, & Vielhauer, 2009). *Time series forecasting* is also a domain where SVM optimization is often applied (Guo et al., 2008; Salahshoor et al., 2010; Tay & Cao, 2002).

Besides manufacturing and image recognition, SVMs are often used within the medicine domain. Among the many areas of application within this domain, the use of SVM in *cancer research* is standing out (Furey et al., 2000; Guyon, Weston, Barnhill, & Vapnik, 2002; Rejani & Selvi, 2009). Other medical application areas are, e.g. drug design (Burbidge et al., 2001) and detection of microcalcifications (El-naqa, Yang, Wernick, Galatsanos, & Nishikawa, 2002).

Further application areas include but are not limited to credit rating (Huang, Chen, Hsu, Chen, & Wu, 2004), food quality control (Borin, Ferrão, Mello, Maretto, & Poppi, 2006), classification of polymers (Li et al., 2009), and rule extraction (Martens, Baesens, Van Gestel, & Vanthienen, 2007). These examples from various industries and optimization problems highlight the wide applicability and adaptability of the SVM algorithm.

As it was shown exemplarily for the SVM algorithm, there are several successful applications of ML in manufacturing available and many are already in daily use in industrial applications worldwide.

## 6. Conclusion and outlook

In this paper, first the challenges of modern manufacturing systems, e.g. increasing complexity, dynamic, high dimensionality, and chaotic structures are highlighted. Following, machine learning limitations and advantages from a manufacturing perspective were discussed before a structuring of the diverse field of machine learning is proposed and an overview of the basic terminology of this inter-disciplinary field is presented. The structure is distinguishing unsupervised machine learning, RL, and supervised machine learning as a possible way to group the available algorithms and applications. It was argued that supervised learning is a good fit for most manufacturing applications due to the fact that the majority of manufacturing applications can provide labeled data. Based on this distinction, the most commonly used supervised machine learning algorithms are presented. Thereafter, an exemplary illustration of successful application in manufacturing of the supervised machine learning algorithm SVMs is presented. This overview highlights the adaptability and variety of usage opportunities in the field.

With fast paced developments in the area of algorithms and increasing availability of data (e.g. due to low cost sensors and the shift toward smart manufacturing) and computing power, the applications for machine learning especially in manufacturing will increase further at a rapid pace. As of today, supervised algorithms have the upper hand in most application in the manufacturing domain. However, with the fast increase in available data, thanks to more and better sensor technologies and increased awareness, unsupervised methods (including RL) may increase in importance in the future. Already today, hybrid approaches are being used that offer 'the best of both worlds.' This corresponds with the attention the Big Data developments received in recent years. Concluding, it can be said with confidence, ML is already a powerful tool for many applications within (intelligent) manufacturing systems and smart manufacturing and its importance will increase further in the future. Its interdisciplinary nature presents a big opportunity but also a significant risk at the same time as collaboration between different disciplines, like Computer Science, Industrial Engineering, Mathematics, and Electrical Engineering is necessary to drive progress.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## ORCID

Thorsten Wuest  <http://orcid.org/0000-0001-7457-7927>

Klaus-Dieter Thoben  <http://orcid.org/0000-0002-5911-805X>

## References

- Akay, D. (2011). Grey relational analysis based on instance based learning approach for classification of risks of occupational low back disorders. *Safety Science*, 49, 1277–1282. doi:<http://dx.doi.org/10.1016/j.ssci.2011.04.018>
- Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed.). Cambridge, MA: MIT Press.
- Apte, C., Weiss, S., & Grout, G. (1993). Predicting defects in disk drive manufacturing: A case study in high dimensional classification. In *IEEE Annual Computer Science Conference on Artificial Intelligence in Application* (pp. 212–218). Los Alamitos, CA.
- Azadeh, A., Saberi, M., Kazem, A., Ebrahimpour, V., Nourmohammadzadeh, A., & Saberi, Z. (2013). A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. *Applied Soft Computing*, 13, 1478–1485. doi:<http://dx.doi.org/10.1016/j.asoc.2012.06.020>
- Bar-or, A., Schuster, A., Wolff, R., & Keren, D. (2005). Decision tree induction in high dimensional, hierarchically distributed databases. In *Proceedings SI-AM International Data Mining Conference* (pp. 466–470). Newport Beach, CA.
- Ben-hur, A., & Weston, J. (2010). A user's guide to support vector machines. In *Data mining techniques for the life sciences methods in molecular biology* (Vol. 609, pp. 223–239). Totowa, NJ: Humana Press. [http://dx.doi.org/10.1007/978-1-60327-241-4\\_13](http://dx.doi.org/10.1007/978-1-60327-241-4_13)
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York, NY: Springer.
- Borin, A., Ferrão, M. F., Mello, C., Maretto, D. A., & Poppi, R. J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Analytica Chimica Acta*, 579, 25–32. doi:<http://dx.doi.org/10.1016/j.aca.2006.07.008>
- Breiman, L. (2001). Random forest. *Machine Learning*, 45, 5–32. doi:<http://dx.doi.org/10.1023/A:1010933404324>
- Brunato, M., & Battiti, R. (2005). Statistical learning theory for location fingerprinting in wireless LANs. *Computer Networks*, 47, 825–845. doi:<http://dx.doi.org/10.1016/j.comnet.2004.09.004>
- Burbidge, R., Trotter, M., Buxton, B., & Holden, S. (2001). Drug design by machine learning: Support vector machines for pharmaceutical data analysis. *Computers & Chemistry*, 26, 5–14.
- Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer*, 21, 77–88. doi:<http://dx.doi.org/10.1109/2.33>
- Çaydaş, U., & Ekici, S. (2010). Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel. *Journal of Intelligent Manufacturing*, 23, 639–650. doi:<http://dx.doi.org/10.1007/s10845-010-0415-2>
- Chand, S., & Davis, J. F. (2010, July). What is smart manufacturing? *Time Magazine*.
- Cherkassky, V., & Ma, Y. (2009). Another look at statistical learning theory and regularization. *Neural Networks*, 22, 958–969. doi:<http://dx.doi.org/10.1016/j.neunet.2009.04.005>
- Chinnam, R. B. (2002). Support vector machines for recognizing shifts in correlated and other manufacturing processes. *International Journal of Production Research*, 40, 4449–4466. doi:<http://dx.doi.org/10.1080/00207540210152920>
- Cohn, D. (2011). *Active learning* (p. 10). Sammut, C. & Webb, G. I. (Eds.) (2011). *Encyclopedia of machine learning* (C. Sammut & G. I. Webb, Eds.) (p. 1058). New York, NY: Springer. doi:<http://dx.doi.org/10.1007/978-0-387-30164-8>
- Cook, D. F., Zobel, C. W., & Wolfe, M. L. (2006). Environmental statistical process control using an augmented neural network classification approach. *European Journal of Operational Research*, 174, 1631–1642. doi:<http://dx.doi.org/10.1016/j.ejor.2005.04.035>
- Corne, D., Dhaenens, C., & Jourdan, L. (2012). Synergies between operations research and data mining: The emerging use of multi-objective approaches. *European Journal of Operational Research*, 221, 469–479. doi:<http://dx.doi.org/10.1016/j.ejor.2012.03.039>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Davis, J., Edgar, T., Graybill, R., Korambath, P., Schott, B., Swink, D., & Wetzel, J. (2015). Smart manufacturing. *Annual Review of Chemical and Biomolecular Engineering*, 6, 141–160. doi:<http://dx.doi.org/10.1146/annurev-chembioeng-061114-123255>

- Dietterich, T. G. (2000). Ensemble methods in machine learning. *Proceedings of Multiple Classifier Systems: First International Workshop (MCS)* (pp. 1–15). Berlin: Springer. doi:[http://dx.doi.org/10.1007/3-540-45014-9\\_1](http://dx.doi.org/10.1007/3-540-45014-9_1)
- Dingli, D. J. (2012). *The manufacturing industry – Coping with challenges* (Working Paper No. 2012/05). Maastricht: Maastricht School of Management.
- Do, T.-N., Lenca, P., Lallich, S., & Pham, N.-K. (2010). Classifying very-high-dimensional data with random forests of oblique decision trees. In F. Guillet, G. Ritschard, D. Zighed, & H. Briand (Eds.), *Advances in knowledge discovery and management* (pp. 39–55). Berlin: Springer.
- Doltsinis, S., Ferreira, P., & Lohse, N. (2012). Reinforcement learning for production ramp-up: A Q-batch learning approach. In *11th International Conference on Machine Learning and Applications* (pp. 610–615). Boca Raton, FL: IEEE. doi:<http://dx.doi.org/10.1109/ICMLA.2012.113>
- Dutt, V., & Gonzalez, C. (2012). Making instance-based learning theory usable and understandable: The instance-based learning tool. *Computers in Human Behavior*, 28, 1227–1240. doi:<http://dx.doi.org/10.1016/j.chb.2012.02.006>
- Elangovan, M., Sakthivel, N. R., Saravanamurugan, S., Nair, B. B., & Sugumaran, V. (2015). Machine learning approach to the prediction of surface roughness using statistical features of vibration signal acquired in turning. *Procedia Computer Science*, 50, 282–288. doi:<http://dx.doi.org/10.1016/j.procs.2015.04.047>
- El-naqa, I., Yang, Y., Wernick, M. N., Galatsanos, N. P., & Nishikawa, R. M. (2002). A support vector machine approach for detection of microcalcifications. *IEEE Transactions on Medical Imaging*, 21, 1552–1563. doi:<http://dx.doi.org/10.1109/TMI.2002.806569>
- European Commission. (2016). *Factories for the future*. Retrieved from [http://ec.europa.eu/research/industrial\\_technologies/factories-of-the-future\\_en.html](http://ec.europa.eu/research/industrial_technologies/factories-of-the-future_en.html)
- Evgeniou, T., Poggio, T., Pontil, M., & Verri, A. (2002). Regularization and statistical learning theory for data analysis. *Computational Statistics & Data Analysis*, 38, 421–432. doi:[http://dx.doi.org/10.1016/S0167-9473\(01\)00069-X](http://dx.doi.org/10.1016/S0167-9473(01)00069-X)
- Evgeniou, T., Pontil, M., & Poggio, T. (2000). Statistical learning theory: A primer. *International Journal of Computer Vision*, 38, 9–13. doi:<http://dx.doi.org/10.1023/A:1008110632619>
- Filipic, B., & Junkar, M. (2000). Using inductive machine learning to support decision making in machining processes. *Computers in Industry*, 43, 31–41. doi:[http://dx.doi.org/10.1016/S0166-3615\(00\)00056-7](http://dx.doi.org/10.1016/S0166-3615(00)00056-7)
- Freund, Y., & Schapire, R. E. (1995). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55, 119–139.
- Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 16, 906–914. doi:<http://dx.doi.org/10.1093/bioinformatics/16.10.906>
- Gagliardi, F. (2011). Instance-based classifiers applied to medical databases: Diagnosis and knowledge extraction. *Artificial Intelligence in Medicine*, 52, 123–139. doi:<http://dx.doi.org/10.1016/j.artmed.2011.04.002>
- Gardner, R., & Bicker, J. (2000). Using machine learning to solve tough manufacturing problems. *International Journal of Industrial Engineering-Theory Applications and Practice*, 7, 359–364.
- Gonzalez, C., Dutt, V., & Lebiere, C. (2013). Validating instance-based learning mechanisms outside of ACT-R. *Journal of Computational Science*, 4, 262–268. doi:<http://dx.doi.org/10.1016/j.jocs.2011.12.001>
- Gordon, J., & Sohal, A. S. (2001). Assessing manufacturing plant competitiveness. *International Journal of Operations & Production Management*, 21, 233–253.
- Graham, J. W. (2012). *Missing data* (p. 322). New York, NY: Springer.
- Günther, J., Pilarski, P. M., Helfrich, G., Shen, H., & Diepold, K. (2015). First steps towards an intelligent laser welding architecture using deep neural networks and reinforcement learning. *Procedia Technology*, 15, 474–483.
- Guo, X., Sun, L., Li, G., & Wang, S. (2008). A hybrid wavelet analysis and support vector machines in forecasting development of manufacturing. *Expert Systems with Applications*, 35, 415–422. doi:<http://dx.doi.org/10.1016/j.eswa.2007.07.052>



- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). A gene selection method for cancer classification using Support Vector Machines. *Machine Learning*, 46, 389–422. doi:<http://dx.doi.org/10.1155/2012/586246>
- Hansson, K., Yella, S., Dougherty, M., & Fleyeh, H. (2016). Machine learning algorithms in heavy process manufacturing. *American Journal of Intelligent Systems*, 6(1), 1–13. doi:<http://dx.doi.org/10.5923/j.ajis.20160601.01>
- Harding, J. A., Shahbaz, M., & Kusiak, A. (2006). Data mining in manufacturing: A review. *Journal of Manufacturing Science and Engineering*, 128, 969–976. doi:<http://dx.doi.org/10.1115/1.2194554>
- Hickey, R. J., & Martin, R. G. (2001). An instance-based approach to pattern association learning with application to the English past tense verb domain. *Knowledge-Based Systems*, 14, 131–136. doi:[http://dx.doi.org/10.1016/S0950-7051\(01\)00089-2](http://dx.doi.org/10.1016/S0950-7051(01)00089-2)
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9, 1735–1780. doi:<http://dx.doi.org/10.1162/neco.1997.9.8.1735>
- Hoffmann, A. G. (1990, August). General limitations on machine learning. In *Proceedings of the 9th European Conference on Artificial Intelligence* (pp. 345–347). Stockholm, Sweden.
- Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, 37, 543–558. doi:[http://dx.doi.org/10.1016/S0167-9236\(03\)00086-1](http://dx.doi.org/10.1016/S0167-9236(03)00086-1)
- Jain, A. K., Murty, M. N., & Flynn, P. (1999). Data clustering: A review. *ACM Computing Surveys*, 31, 264–323. doi:<http://dx.doi.org/10.1145/331499.331504>
- Jurkovic, Z., Cukor, G., Brezocnik, M., & Brajkovic, T. (2016). A comparison of machine learning methods for cutting parameters prediction in high speed turning process. *Journal of Intelligent Manufacturing*. doi:<http://dx.doi.org/10.1007/s10845-016-1206-1>
- Kabacoff, R. I. (2011). Advanced methods for missing data. In R. I. Kabacoff (Ed.), *R in action: Data analysis and graphics with R* (pp. 352–371). Shelter Island, NY: Manning Publications.
- Kang, P., & Cho, S. (2008). Locally linear reconstruction for instance-based learning. *Pattern Recognition*, 41, 3507–3518. doi:<http://dx.doi.org/10.1016/j.patcog.2008.04.009>
- Kang, P., Kim, D., & Cho, S. (2016). Semi-supervised support vector regression based on self-training with label uncertainty: An application to virtual metrology in semiconductor manufacturing. *Expert Systems with Applications*, 51, 85–106. doi:<http://dx.doi.org/10.1016/j.eswa.2015.12.027>
- Keerthi, S. S., & Lin, C.-J. (2003). Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation*, 15, 1667–1689. doi:<http://dx.doi.org/10.1162/089976603321891855>
- Khemchandani, R., & Chandra, S. (2009). Knowledge based proximal support vector machines. *European Journal of Operational Research*, 195, 914–923. doi:<http://dx.doi.org/10.1016/j.ejor.2007.11.023>
- Kim, D., Kang, P., Cho, S., Lee, H., & Doh, S. (2012). Machine learning-based novelty detection for faulty wafer detection in semiconductor manufacturing. *Expert Systems with Applications*, 39, 4075–4083. doi:<http://dx.doi.org/10.1016/j.eswa.2011.09.088>
- Köksal, G., Batmaz, İ., & Testik, M. C. (2011). A review of data mining applications for quality improvement in manufacturing industry. *Expert Systems with Applications*, 38, 13448–13467. doi:<http://dx.doi.org/10.1016/j.eswa.2011.04.063>
- Koltchinskii, V., Abdallah, C. T., Ariola, M., & Dorato, P. (2001). Statistical learning control of uncertain systems: Theory and algorithms. *Applied Mathematics and Computation*, 120, 31–43. doi:[http://dx.doi.org/10.1016/S0096-3003\(99\)00283-0](http://dx.doi.org/10.1016/S0096-3003(99)00283-0)
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31, 249–268.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- Kwak, D.-S., & Kim, K.-J. (2012). A data mining approach considering missing values for the optimization of semiconductor-manufacturing processes. *Expert Systems with Applications*, 39, 2590–2596. doi:<http://dx.doi.org/10.1016/j.eswa.2011.08.114>
- Lang, S. (2007). *Durchgängige Mitarbeiterinformation zur Steigerung von Effizienz und Prozesssicherheit in der Produktion* (Dissertation). Universität Erlangen-Nürnberg, Bamberg: Meisenbach Verlag.



- Larose, D. (2005). *Discovering knowledge in data – An introduction to data mining*. Hoboken, NJ: Wiley.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521, 436–444. doi:<http://dx.doi.org/10.1038/nature14539>
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1, 541–555.
- Lee, J., & Ha, S. (2009). Recognizing yield patterns through hybrid applications of machine learning techniques. *Information Sciences*, 179, 844–850. doi:<http://dx.doi.org/10.1016/j.ins.2008.11.008>
- Lee, J., Lapira, E., Bagheri, B., & Kao, H. (2013). Recent advances and trends in predictive manufacturing systems in big data environment. *Manufacturing Letters*, 1, 38–41. doi:<http://dx.doi.org/10.1016/j.mfglet.2013.09.005>
- Li, H., Liang, Y., & Xu, Q. (2009). Support vector machines and its applications in chemistry. *Chemometrics and Intelligent Laboratory Systems*, 95, 188–198. doi:<http://dx.doi.org/10.1016/j.chemolab.2008.10.007>
- Li, T.-S., & Huang, C.-L. (2009). Defect spatial pattern recognition using a hybrid SOM–SVM approach in semiconductor manufacturing. *Expert Systems with Applications*, 36, 374–385. doi:<http://dx.doi.org/10.1016/j.eswa.2007.09.023>
- Loyer, J.-L., Henriques, E., Fontul, M., & Wiseall, S. (2016). Comparison of machine learning methods applied to the estimation of manufacturing cost of jet engine components. *International Journal of Production Economics*, 178, 109–119. doi:<http://dx.doi.org/10.1016/j.ijpe.2016.05.006>
- Lu, S. C.-Y. (1990). Machine learning approaches to knowledge synthesis and integration tasks for advanced engineering automation. *Computers in Industry*, 15, 105–120. doi:[http://dx.doi.org/10.1016/0166-3615\(90\)90088-7](http://dx.doi.org/10.1016/0166-3615(90)90088-7)
- Manallack, D. T., & Livingstone, D. J. (1999). Neural networks in drug discovery: Have they lived up to their promise? *European Journal of Medicinal Chemistry*, 34, 95–208.
- Manning, C. D., Raghavan, P., & Schütze, H. (2009). *An introduction to information retrieval*. Cambridge: Cambridge University Press.
- Margolis, D., Land, W. H., Gottlieb, R., & Qiao, X. (2011). A complex adaptive system using statistical learning theory as an inline preprocess for clinical survival analysis. *Procedia Computer Science*, 6, 279–284. doi:<http://dx.doi.org/10.1016/j.procs.2011.08.052>
- Martens, D., Baesens, B., Van Gestel, T., & Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European Journal of Operational Research*, 183, 1466–1476. doi:<http://dx.doi.org/10.1016/j.ejor.2006.04.051>
- Monostori, L. (1993). A step towards intelligent manufacturing: Modelling and monitoring of manufacturing processes through artificial neural networks. *CIRP Annals*, 42, 485–488. doi:[http://dx.doi.org/10.1016/S0007-8506\(07\)62491-3](http://dx.doi.org/10.1016/S0007-8506(07)62491-3)
- Monostori, L. (2003). AI and machine learning techniques for managing complexity, changes and uncertainties in manufacturing. *Engineering Applications of Artificial Intelligence*, 16, 277–291. doi:[http://dx.doi.org/10.1016/S0952-1976\(03\)00078-2](http://dx.doi.org/10.1016/S0952-1976(03)00078-2)
- Monostori, L., Hornyák, J., Egresits, C., & Viharos, Z. J. (1998). Soft computing and hybrid AI approaches to intelligent manufacturing. *Tasks and Methods in Applied Artificial Intelligence Lecture Notes in Computer Science*, 1416, 765–774. doi:[http://dx.doi.org/10.1007/3-540-64574-8\\_463](http://dx.doi.org/10.1007/3-540-64574-8_463)
- Monostori, L., Márkus, A., Van Brussel, H., & Westkämper, E. (1996). Machine learning approaches to manufacturing. *CIRP Annals*, 45, 675–712.
- Nilsson, N. J. (2005). *Introduction to machine learning*. Stanford, CA.
- Okamoto, S., & Yugami, N. (2003). Effects of domain characteristics on instance-based learning algorithms. *Theoretical Computer Science*, 298, 207–233. doi:[http://dx.doi.org/10.1016/S0304-3975\(02\)00424-3](http://dx.doi.org/10.1016/S0304-3975(02)00424-3)
- Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11, 169–198.
- Pham, D. T., & Afify, A. A. (2005). Machine-learning techniques and their applications in manufacturing. Proceedings of the Institution of Mechanical Engineers. Part B: *Journal of Engineering Manufacture*, 219, 395–412. doi:<http://dx.doi.org/10.1243/095440505X32274>

- Quadrianto, N., & Buntine, W. L. (2011). *Regression* (pp. 838–842). Sammut, C. & Webb, G. I. (2011). *Encyclopedia of machine learning* (C. Sammut & G. I. Webb, Eds.) (p. 1058). New York, NY: Springer. doi:<http://dx.doi.org/10.1007/978-0-387-30164-8>
- Rejani, Y. I. A., & Selvi, S. T. (2009). Early detection of breast cancer using SVM classifier technique. *International Journal on Computer Science and Engineering*, 1, 127–130.
- Ribeiro, B. (2005). Support vector machines for quality monitoring in a plastic injection molding process. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 35, 401–410. doi:<http://dx.doi.org/10.1109/TSMCC.2004.843228>
- Salahshoor, K., Kordestani, M., & Khoshro, M. S. (2010). Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers. *Energy*, 35, 5472–5482. doi:<http://dx.doi.org/10.1016/j.energy.2010.06.001>
- Sammut, C., & Webb, G. I. (2011). *Encyclopedia of machine learning* (C. Sammut & G. I. Webb, Eds.) (p. 1058). New York, NY: Springer. doi:<http://dx.doi.org/10.1007/978-0-387-30164-8>
- Samuel, A. (1959). Some studies in machine learning using the game of checkers. *IBM Journal*, 3, 210–229. doi:<http://dx.doi.org/10.1147/rd.33.0210>
- Scheidat, T., Leich, M., Alexander, M., & Vielhauer, C. (2009). Support vector machines for dynamic biometric handwriting classification. In *Proceedings of AIAI Workshops* (pp. 118–125). Thessaloniki, Greece.
- Shiang, L. E., & Nagaraj, S. (2011). Impediments to innovation: Evidence from Malaysian manufacturing firms. *Asia Pacific Business Review*, 17, 209–223. doi:<http://dx.doi.org/10.1080/13602381.2011.533502>
- Simon, H. A. (1983). Why should machines learn? In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine learning* (pp. 25–37). Charlotte, NC: Tioga Press.
- Smola, A., & Vishwanathan, S. V. N. (2008). *Introduction to machine learning*. Cambridge: Cambridge University Press.
- Steel, D. (2011). Testability and statistical learning theory. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Handbook of the philosophy of science* (Vol. 7, pp. 849–861). Amsterdam: Elsevier. doi:<http://dx.doi.org/10.1016/B978-0-444-51862-0.50028-9>
- Stone, P. (2011). *Reinforcement learning* (pp. 849–851). Sammut, C., & Webb, G. I. (Eds.) (2011). *Encyclopedia of machine learning* (C. Sammut & G. I. Webb, Eds.) (p. 1058). New York, NY: Springer. doi:<http://dx.doi.org/10.1007/978-0-387-30164-8>
- Sun, J., Rahman, M., Wong, Y., & Hong, G. (2004). Multiclassification of tool wear with support vector machine by manufacturing loss consideration. *International Journal of Machine Tools and Manufacture*, 44, 1179–1187. doi:<http://dx.doi.org/10.1016/j.ijmachtools.2004.04.003>
- Susto, G. A., Schirru, A., Pampuri, S., McLoone, S., & Beghi, A. (2015). Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics*, 11, 812–820. doi:<http://dx.doi.org/10.1109/TII.2014.2349359>
- Sutton, R. S., & Barto, A. G. (2012). *Reinforcement learning: An introduction* (2nd ed.). Cambridge, MA: MIT Press.
- Tay, F. E. H., & Cao, L. J. (2002). Modified support vector machines in financial time series forecasting. *Neurocomputing*, 48, 847–861. doi:[http://dx.doi.org/10.1016/S0925-2312\(01\)00676-2](http://dx.doi.org/10.1016/S0925-2312(01)00676-2)
- Thomas, A. J., Byard, P., & Evans, R. (2012). Identifying the UK's manufacturing challenges as a benchmark for future growth. *Journal of Manufacturing Technology Management*, 23, 142–156. doi:<http://dx.doi.org/10.1108/17410381211202160>
- Wang, K.-J., Chen, J. C., & Lin, Y.-S. (2005). A hybrid knowledge discovery model using decision tree and neural network for selecting dispatching rules of a semiconductor final testing factory. *Production Planning & Control*, 16, 665–680. doi:<http://dx.doi.org/10.1080/09537280500213757>
- White House. (2014, October 27). FACT SHEET: President Obama announces new actions to further strengthen U.S. manufacturing. Retrieved from <https://www.whitehouse.gov/the-press-office/2014/10/27/fact-sheet-president-obama-announces-new-actions-further-strengthen-us-m>
- Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. *Mechanical Systems and Signal Processing*, 21, 2560–2574. doi:<http://dx.doi.org/10.1016/j.ymssp.2006.12.007>

- Wiendahl, H.-P., & Scholtissek, P. (1994). Management and control of complexity in manufacturing. *CIRP Annals*, 43, 533–540. doi:[http://dx.doi.org/10.1016/S0007-8506\(07\)60499-5](http://dx.doi.org/10.1016/S0007-8506(07)60499-5)
- Wiering, M., & Van Otterlo, M. (2012). *Reinforcement learning: State-of-the-art*. New York, NY: Springer.
- Wu, Q. (2010). Product demand forecasts using wavelet kernel support vector machine and particle swarm optimization in manufacture system. *Journal of Computational and Applied Mathematics*, 233, 2481–2491. doi:<http://dx.doi.org/10.1016/j.cam.2009.10.030>
- Wuest, T. (2015). *Identifying product and process state drivers in manufacturing systems using supervised machine learning* (Springer theses). New York, NY: Springer Verlag.
- Wuest, T., Irgens, C., & Thoben, K.-D. (2014). An approach to monitoring quality in manufacturing using supervised machine learning on product state data. *Journal of Intelligent Manufacturing*, 25, 1167–1180. doi:<http://dx.doi.org/10.1007/s10845-013-0761-y>
- Wuest, T., Liu, A., Lu, S. C.-Y., & Thoben, K.-D. (2014). Application of the stage gate model in production supporting quality management. *Procedia CIRP*, 17, 32–37. doi:<http://dx.doi.org/10.1016/j.procir.2014.01.071>
- Yang, K., & Trewn, J. (2004). *Multivariate statistical methods in quality management*. New York, NY: McGraw-Hill.
- Yu, L., & Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)* (pp. 8). Washington, DC.
- Zheng, Y., Li, S., & Wang, X. (2010). An approach to model building for accelerated cooling process using instance-based learning. *Expert Systems with Applications*, 37, 5364–5371. doi:<http://dx.doi.org/10.1016/j.eswa.2010.01.020>
- Zhou, Z.-H. (2012). *Ensemble methods – Foundations and algorithms*, Machine Learning & Pattern Recognition Series. Florida, FL: Chapman & Hall/CRC. ISBN: 978-1-4398-3003-1.