



# Potential, challenges and future directions for deep learning in prognostics and health management applications

Olga Fink <sup>a,\*</sup>, Qin Wang <sup>a</sup>, Markus Svensén <sup>b</sup>, Pierre Dersin <sup>c</sup>, Wan-Jui Lee <sup>d</sup>, Melanie Ducoffe <sup>e</sup>

<sup>a</sup> Intelligent Maintenance Systems, ETH Zurich, Switzerland

<sup>b</sup> GE Aviation Digital Group, United Kingdom

<sup>c</sup> Alstom, France

<sup>d</sup> Dutch Railways, Netherlands

<sup>e</sup> Airbus, France

## ARTICLE INFO

### Keywords:

Deep learning  
Prognostics and health management  
GAN  
Domain adaptation  
Fleet PHM  
Deep reinforcement learning  
Physics-induced machine learning

## ABSTRACT

Deep learning applications have been thriving over the last decade in many different domains, including computer vision and natural language understanding. The drivers for the vibrant development of deep learning have been the availability of abundant data, breakthroughs of algorithms and the advancements in hardware. Despite the fact that complex industrial assets have been extensively monitored and large amounts of condition monitoring signals have been collected, the application of deep learning approaches for detecting, diagnosing and predicting faults of complex industrial assets has been limited. The current paper provides a thorough evaluation of the current developments, drivers, challenges, potential solutions and future research needs in the field of deep learning applied to Prognostics and Health Management (PHM) applications.

## 1. Today's challenges in PHM applications

The goal of Prognostics and Health Management (PHM) is to provide methods and tools to design optimal maintenance policies for a specific asset under its distinct operating and degradation conditions, achieving a high availability at minimal costs. PHM can be considered as a holistic approach to an effective and efficient system health management (Lee et al., 2014). PHM integrates the detection of an incipient fault (fault detection), its isolation, the identification of its origin and the specific fault type (fault diagnostics) and the prediction of the remaining useful life (prognostics). However, PHM does not end with the prediction of the remaining useful life (RUL). The system health management goes beyond the predictions of failure times and supports optimal maintenance and logistics decisions by considering the available resources, the operating context and the economic consequences of different faults. Health management is the process of taking timely and optimal maintenance actions based on outputs from diagnostics and prognostics, available resources and operational demand. PHM focuses on assessing and minimizing the operational impact of failures, and controlling maintenance costs (Lee et al., 2014).

Nowadays, the condition of complex systems is typically monitored by a large number of different types of sensors, capturing e.g. temperature, pressure, flow, vibration, images or even video streams of system conditions, resulting in very heterogeneous condition monitoring data at different time scales. Additionally, the signals are affected

by measurement and transmission noise. In many cases, the sensors are partly redundant, having several sensors measuring the same system parameter. Not all of the signals contain information on a specific fault type since different fault types are affecting different signals and the correspondence is generally not one-to-one.

In most cases, using raw condition monitoring data in machine learning applications will not be conducive to detect faults or predict impending failures. Hence, successful PHM applications typically require manual pre-processing to derive more useful representations of signals in the data, a process known as *feature engineering*. Feature engineering involves combinations of transforming raw data using e.g. statistical indicators or other signal processing approaches, such as time-frequency analysis, and steps to reduce the dimensionality of the data, where applicable, either with manual or automatic feature selection (filter, wrapper or embedded approaches) (Forman, 2003).

Feature selection depends on the past observations or knowledge of the possible degradation types and their effect on the signals. Selecting too few or too many features may result in missed alarms, particularly for those fault types that have not been previously observed. At the same time, the number of false alarms must be minimized since this would otherwise impact the credibility of the developed model negatively. Thus, feature selection in a supervised fault classification problem must contribute to minimizing the false alarm rate (false positives) and maximizing the detection rate (true positives).

\* Corresponding author.

E-mail address: [ofink@ethz.ch](mailto:ofink@ethz.ch) (O. Fink).

The concept of feature extraction is also closely linked to the concept of condition indicators (Sharma and Parey, 2016). A condition indicator is defined as a feature of condition monitoring system data whose behaviour changes in a predictable way as the system degrades or operates in different operational modes. Therefore, a condition indicator can be any feature that enables to distinguish normal from faulty conditions or for predicting the remaining useful life (RUL). While there can be several condition indicators for one system, a more attractive way of monitoring the system health condition is by integrating several condition indicators into one health indicator, a value that represents the health status of the component to the end user. Health indicators have to follow some desired characteristics, including monotonicity, robustness and adaptability (Hu et al., 2016b). Several ways have been proposed to design health indicators and partly even subsequently using them for predicting RUL (Guo et al., 2017). Recently, also an approach for learning the healthy system condition and using the distance measure to the learned healthy condition as an unbounded health indicator was proposed (Michau et al., 2020; Arias Chao et al., 2019).

Due to the multiplicity of the possible fault types that can occur, feature engineering may face limitations to design a set of representative features that is able to depict the differences between all possible fault types. Handcrafted features do not necessarily offer generalization ability and transferability from one system to another or even to other fault types. They also have a limited scalability due to the expert-driven manual approach. Additionally, the performance of feature engineering is highly dependent on the experience and expertise of the domain experts performing the task. The quality of the extracted features highly influences the performance of machine learning approaches that are using the extracted features. As the number of monitored parameters increases so does the difficulty of feature engineering for diagnostics engineers and consequently there is an interest in automating this processes (Yan and Yu, 2015) or circumventing the need for feature engineering in the first place. Deep learning (DL) has the potential to incorporate feature engineering, or at least parts thereof, into the end-to-end learning processes.

While fault detection and fault diagnostics have been recently adopting DL approaches, prognostics has remained a rather difficult terrain for DL. Prognostics is the study and prediction of the future evolution of the health of the system being monitored, through the estimation of the RUL. Several directions have been proposed for estimating the RUL which broadly fall into the following categories (Lee et al., 2014; Fink, 2020): (a) model-based approaches (partly also referred to as physics-based), i.e. relying mainly on multi-physics models for asset normal operation and physical degradation laws (possibly modelled as stochastic processes); (b) data-driven approaches that are based on condition monitoring data (where DL approaches belong to); (c) knowledge-based approaches, relying on expert judgements. There are also combinations of these three directions that are typically referred to as "hybrid approaches". While any combinations are, in fact, possible, the term "hybrid approaches" is typically used for approaches combining data-driven techniques with physics-based approaches (Arias Chao et al., 2019).

Two different scenarios can be considered for the RUL prediction:

- Degradation prediction (immediate degradation onset, dependent on the operating conditions)
- Detection of fault onset (e.g. crack initiation) and a subsequent prediction of the fault progression (dependent on the operating conditions)

Earlier works in data-driven RUL prediction have been mainly focusing on the statistical models, such as the thorough mathematical treatment of the RUL estimation in Banjevic (2009), with statistical properties of the expectation and standard deviation of the RUL and their asymptotic behaviour, under broad assumptions. A fairly comprehensive survey of data-driven approaches for predicting RUL can

be found in Si et al. (2011). A frequently used approach for data-driven RUL prediction consists of modelling degradation trajectories by stochastic processes that capture more or less accurately degradation physics. An example of that approach is given by Zhai and Ye in Zhai and Ye (2017). One of the research directions is also on assessing the performance of prognostics algorithms and the uncertainty in predictions by defining key performance indicators, such as the work of Saxena et al. (2014). A recent contribution has highlighted the importance of changes in the coefficient of variation of extracted features in detecting faults and in segmenting time series for health assessment and prognostics (Atamuradov et al., 2017). This approach can be also linked to learning or extracting the health indicators and using them to detect the fault onset.

Ultimately, what PHM may expect from deep learning is either solving complex PHM problems that were not solvable with traditional approaches or improving the performance of the traditional approaches, automating the development of the applied models and making the usage of the models more robust and more cost-effective in industrial contexts.

Given its promising potential for PHM applications, DL is particularly well positioned to offer solutions to following issues:

- Ability to automatically process massive amounts of condition monitoring data
- Ability to automatically extract useful features from high-dimensional, heterogeneous data sources
- Ability to learn functional and temporal relationships between and within the time series of condition monitoring signals
- Ability to transfer knowledge between different operating conditions and different units.

In addition, there are several general problems that need to be overcome in order to win broader acceptance of machine learning models in general and DL models in particular by the different stakeholders. These challenges include:

- Ability to identify and deal with outliers
- Dynamic learning ability of the applied approaches: in particular, learning evolving operating conditions and distinguishing them from evolving faults
- Ability to detect novelty, i.e. hitherto unknown degradation types
- Robustness of the applied approaches, including cases under highly varying operating conditions
- Generalization ability of the developed models (ability to transfer knowledge between different operating conditions and different units)
- Interpretability of the obtained results for the domain experts
- Automated and optimal decision support taking the system condition and the relevant constraints, such as resource availability, into consideration.

Currently, the application of DL in PHM is mainly driven by the developments in other DL application domains, such as computer vision and natural language processing (NLP). Recent progress in DL has also been increasingly reflected in DL applications in PHM, although the transfer to industrial applications has been limited. While some of the requirements of PHM are similar to those encountered in other fields, other requirements become more specific due to the specificities of complex industrial systems and the collected condition monitoring data.

This review paper provides a thorough evaluation of the current developments, drivers, challenges and potential solutions in the field of DL applied to PHM applications. Due to the wide variety of possible topics and potential directions, we focused on the most promising aspects and directions with the potential of solving the current challenges in industry as pointed out in the above section. The content and structure of the paper are shown in Fig. 1.

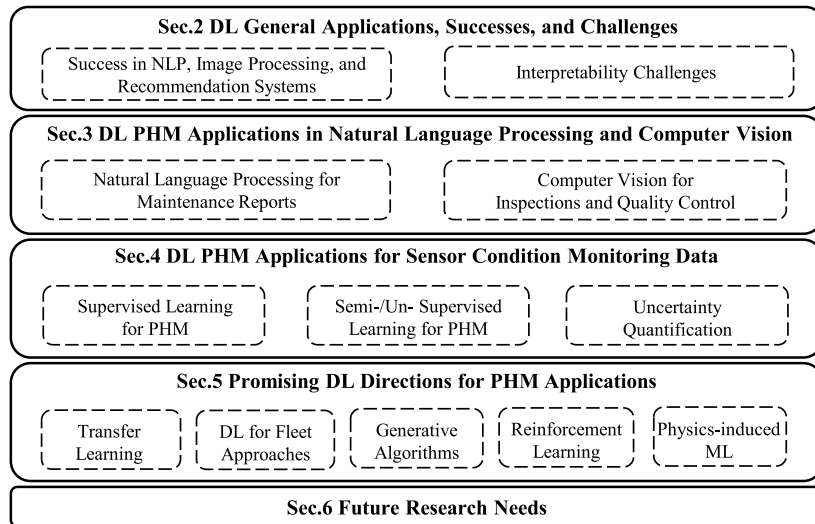


Fig. 1. Topics covered by this article.

## 2. Deep learning applications, successes and challenges

### 2.1. Introduction to deep learning

Deep learning is an overarching concept that encompasses new variants of a range of established learning models, known as neural networks (Bishop, 2007), now more commonly referred to as deep neural networks (DNNs) (Goodfellow et al., 2016; Zhang et al., 2019a). DNNs are called “deep” because they have multiple layers of computational units, whereas traditional, “shallow” neural networks usually only have a single layer. The extension from a single layer to multiple layers may seem rather obvious and various forms of DNNs were indeed tried already during the early-mid 1990s (Tesauro, 1992). However, these early attempts at training DNNs did not seem to offer any advantage compared to single-layer neural networks, rather the opposite, and consequently interest in this direction of research diminished. With time, though, new ways of training DNNs were gradually developed (Bengio et al., 2006; Lecun et al., 1998; Hinton et al., 2006), while the performance of computer hardware continued to improve. In 2012, there was a breakthrough, in terms of reaching beyond the neural networks and machine learning communities, when a deep convolutional neural network (CNN) set new top scores in image object recognition (Krizhevsky et al., 2012; He et al., 2015). This resulted in a surge of interest in DL, benefiting research that in many cases was already well underway, and resulting in rapid advancements of fields of applications for DL.

A common theme of DL is that it is used to tackle complex problems such as computer vision or NLP, for which near-human performance seemed a long way away a little more than a decade ago. These problems are tackled using intricate models with very large numbers of parameters, which are trained on very large datasets with only a limited amount of pre-processing required. Often, the architecture of these models reflects assumed structure in the data. For example, deep CNNs are typically used for image processing (Section 2.2.2), whereas recurrent neural networks (RNNs) are often used for sequential data, such as NLP data (Section 2.2.1). This structural match between model and data facilitates the identification of relevant features in the data during the learning process and DL models are frequently claimed to circumvent the need for feature engineering (Forman, 2003; Bengio et al., 2013). While several factors have contributed to the explosive growth of deep learning, three stand out as being particularly important:

- Availability of large labelled datasets required to train large DL models with large numbers of parameters (Deng et al., 2009).

Some of these datasets have been built over considerable time whereas others have been created through internet-scale crowd-sourcing efforts (Fiedler et al., 2018), while others yet are products of an increasingly digitized society. Many of these datasets are publicly (though not necessarily freely) available on the internet and several are used as benchmark datasets; some of them will be mentioned (with references) in the following sections.

- Emergence of fast, affordable hardware (Coates et al., 2013), well suited to carry out the large amount of computation required in training DL models. While this hardware was initially developed for the gaming console/PC mass-market, it is nowadays available specifically adapted for the purposes of DL (Haensch et al., 2018).
- The development of several sophisticated, freely available software libraries (Abadi et al., 2016; Jia et al., 2014; Seide and Agarwal, 2016; Al-Rfou’ et al., 2016; Paszke et al., 2017) has made building, training and evaluating DL models relatively straightforward. Although highly complex, several libraries include or are accompanied by wrapping libraries that make the creation of models with established architectures simple and intuitive, requiring only a small amount of coding.

The need for large computational efforts and, especially, large labelled datasets may be too difficult or costly to satisfy for many learning problems. Fortunately, it has been shown that DL models already trained on one task can often be retrained to perform a similar task. Such retraining, commonly known as *transfer learning* (Yosinski et al., 2014), typically requires much less data and computational resources, making DL applicable to a potentially much wider range of practical tasks.

### 2.2. Deep learning successes

Before looking at the current and future applications of DL within the field of PHM, we first review some of the most significant successes that DL has achieved so far in other fields. These serve as inspiration as well as examples for how to move the field forward.

#### 2.2.1. Speech recognition, machine translation and NLP

Many people probably do not realize that they already own one or more devices that implement DL technology. The speech recognition systems on modern smartphones, tablets and computers were possibly the first widely available applications of DL. These first systems used deep belief networks (DBNs) (Hinton et al., 2012) in conjunction with hidden Markov models (HMM) (Rabiner, 1989) to deliver a significant

performance improvement on the until then prevailing technology, which had used Gaussian mixture models (Bishop, 2007) in place of the DBNs. More recent DL based systems have moved on to use variants of RNNs (Amodei et al., 2015; He et al., 2018) in so-called end-to-end systems that implement the complete speech recognition process within a single model. Most recent speech recognition systems have been trained on large, proprietary datasets, but there is a number of publicly available datasets such as TIMIT (Garofolo et al., 1993) and Switchboard (Godfrey and Holliman, 1997), published by the Linguistic Data Consortium (LDC),<sup>1</sup> which have been frequently used to benchmark speech recognition models.

DL has also had a significant impact on processing of natural language in written form. Various types of RNNs have successfully been applied to the task of machine translation (Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014; Wu et al., 2016; Hieber et al., 2017), achieving performance that is comparable to or exceeding that of carefully handcrafted models. RNNs have also been shown as being capable of capturing aspects such as style of writing and syntactic rules from raw text data ranging from Shakespeare plays to the C source code of the Linux kernel (Karpathy, 2015). While these results are remarkable, a popular and possibly better alternative to raw text data is offered by new representations for language components, such as the word2vec embedding of words (Mikolov et al., 2013), which encodes words as vectors, such that words that share aspects of their meaning will be mapped to similar vectors. Embeddings at the character and sentence level have also been proposed. Using this kind of representations, DL models have been successfully applied to classic NLP tasks, such as part-of-speech tagging and named entity recognition, as well as sentiment analysis, and question-answering and dialogue systems (Young et al., 2017).<sup>2</sup>

### 2.2.2. Image processing

In the last decade, deep neural networks, in particular deep CNNs (Lecun et al., 1998), have revolutionized the field of image processing. These models were proposed as early as 1989 (LeCun et al., 1989) for the purpose of image classification and were trained to recognize images of numbers from handwritten ZIP-codes collected by the US post office, the so-called MNIST dataset (LeCun and Cortes, 2005). However, it was only in 2012 that these models truly caught the attention of wider research communities, after a deep CNN, subsequently named AlexNet, established a new top score in object recognition (Krizhevsky et al., 2012). The model was trained on the then recently released ImageNet (Deng et al., 2009) dataset, which contains more than a million images that have been labelled using a crowd-sourcing effort. There are also several smaller datasets, such as CIFAR (Krizhevsky, 2009) and Caltech-101 (Li et al., 2004), that have been used extensively in DL research. The AlexNet paper was followed by several papers (Simonyan and Zisserman, 2014; Szegedy et al., 2014; He et al., 2015), exploring increasingly deeper and intricate network topologies, with in some cases more than hundred layers and millions of parameters. Focus has also moved on from simple object detection to object instance segmentation, where multiple objects, including individual instances of the same category, are segmented out at the pixel level (He et al., 2017). Researchers have also tackled joint NLP-image-processing problems by combining RNNs and CNNs into systems for automatic generation of captions to images (Karpathy and Fei-Fei, 2015).

### 2.2.3. Recommendation systems

Many of the remarkable successes of DL have been in applications where humans have traditionally excelled above computers, such as those described above. However, it has also been used to successfully

tackle problems that have arisen out of modern Internet technologies, such as large-scale recommendation systems (Adomavicius and Tuzhilin, 2005; Ricci et al., 2011). As the name suggests, these systems seek to provide recommendations on content to present to users, who are often acting in the role of consumers. This enables automatic personalization of the recommended content, product or services to a large number of users. An example are the additional items offered on sites such as Amazon.com once a user has bought, or merely expressed an interest in an initial item; another are the film recommendations users receive on Netflix (Bennett and Lanning, 2007). Common to all such systems is that they learn user preferences from large, often vast collections of data. Hence, they are well-suited to the application of DL technologies (Zhang et al., 2019c). For example, DNNs have been used to improve the recommendation systems for the YouTube online video content platform (Covington et al., 2016), providing suggestions to users on what they might want to watch. Similarly, DNNs have been combined with generalized linear models to improve the recommendation system for apps in the Google Play mobile app store (Cheng et al., 2016). These successes have the potential to be transferred to a PHM context. For example, there are some parallels between recommendation systems and decision support systems, i.e. in the context of suggesting actions for an operator, in order to benefit from both the expertise of experienced humans and artificial intelligence. However, decision support systems are not only used to recommend the optimal actions but they can also suggest system operational set points, i.e. to prolong the remaining useful life. Although these scenarios seem to have a lot of commonalities, there are also challenges. For example, the actions used in PHM applications are often continuous instead of discrete, making it hard to directly adapt classical recommendation system algorithms to PHM applications. Also, recommendation systems generally rely on an abundance of data which is rarely present in the PHM context.

### 2.3. Interpretability

The inherent characteristics of deep neural networks, particularly the non-linear computations that are distributed over single nodes and layers, make them on the one hand powerful in terms of computational performance and on the other hand they do not allow for any interpretability. Therefore, particularly the users and the domain experts in the specific fields have been concerned about the interpretability, explainability, transparency and understanding of the models and the results. Also, the term “trust” has been recently increasingly raised in the context of deep learning. Whereby “trust” can be related to different topics, including bias, lacks of representativeness of the training datasets and adversarial attacks. Trust can be developed or enhanced by improving the transparency, the interpretability and the understanding of the algorithms (Lipton, 2016). The focus of this section is particularly on the term of interpretability which is also used as a synonym to explainability. The goal is to understand and explain what the model predicts. One approach could be to build interpretability into the model, e.g. by fusing physical models and machine learning, or learning the underlying physics explicitly, as described in Section 5.5. Here, however, we focus on an alternative, post-hoc approach to interpretability (Montavon et al., 2018) that tries to explain given models and their output. This approach seeks to provide interpretability on different levels (Samek and Müller, 2019):

- Explaining learned representations by analysing the internal representation of (e.g.) neural networks
- Explaining individual predictions
- Explaining model behaviour.

Furthermore, different approaches have been proposed to explain the behaviour of machine learning approaches post-hoc, these include (Samek and Müller, 2019):

<sup>1</sup> <https://www.ldc.upenn.edu/>.

<sup>2</sup> As is the case of speech, text data for different tasks are published by LDC and machine translation data can be found from <http://statmt.org/>.



- Explaining with Surrogates
- Explaining with local perturbations
- Propagation-Based Approaches (Leveraging Structure)
- Meta-explanations.

Particularly surrogate approaches have been gaining popularity where simple surrogate functions are used to explain the behaviour of more complex models. One of such models is the Local Interpretable Model-agnostic Explanations (LIME) (Ribeiro et al., 2016). Most of the approaches proposed to improve the interpretability are particularly applicable for computer vision tasks where the interpretability and understanding can be gained by different types of visualizations. While some of the approaches can be also transferred to the tasks in prognostics and health management, the explainability may not be sufficient for physical systems.

### 3. DL PHM applications in natural language processing and computer vision

#### 3.1. Natural language processing for maintenance reports

While sensory data is perhaps what most people would first think of when hearing the term “PHM data”, there are other kinds of information that can be as important and are often available in PHM. Failure/incident reports, maintenance records and event logs can help to identify frequency, causes and remedies of failures but require different processing than (numerical) sensor data. These data may be used as sources for both input data and labels/targets. However, in most cases, failure records and maintenance reports are entered manually by technicians as free form text, which introduces a certain degree of variability and sometimes even ambiguity. NLP techniques can help to automate this process and get more insights from these data, e.g. to identify the root causes of failures.

Current DL models for NLP have two major components, i.e., representation learning and RNNs. Representation learning in NLP is also called word embedding (Mikolov et al., 2013; Peters et al., 2018). Learning of word embedding can be done by either adding an embedding layer to a deep neural network or outside the deep neural network by unsupervised learning. Word embedding performs automatic feature engineering that can capture the correlations between consecutive words as well as between words and the wider context they appear in. These representations typically are more compact than the other widely used 1-of- $N$  representations of words. Long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997) is one of the most popular RNNs used in NLP. With the memory mechanism, it can capture not only short term but also long term relationships within the text. Integration of word embedding and RNNs enables end-to-end learning and makes the adoption of NLP in PHM a viable proposition. With the help of NLP, we can identify keywords in maintenance records or failure reports and predict the type of failure a maintenance record remedied. Recently, NLP techniques have been used in PHM for various tasks. For example, Su et al. (2019) uses LSTM models to provide fault diagnosis for storage devices based on self-monitoring logs and Ravi et al. (2019) conduct substation transformer failure analysis based on event logs.

#### 3.2. Computer vision for images and videos from inspections and quality control

Many PHM applications based on DL on image data have been proposed in recent years, typically with the motivation of automating the processing of these data, e.g. classifying or deriving quantitative descriptors for the content of these images. Gibert et al. (2015) built a multi-task deep CNN using image data collected for the purpose of monitoring railway infrastructure. The use of dedicated equipment allowed a substantial amount of data to be collected along railway track, which were manually labelled using purpose-built software. The multi-task CNN was capable of recognizing parts of railway infrastructure as

well as detecting damage such as cracked concrete ties and missing or broken rail fasteners. Liu et al. (2018) used image data collected with drones to train a CNN to detect coating breakdown and corrosion (CBC) in ballast tanks of marine and offshore structures. They employed a transfer learning approach using a VGG19 model (Simonyan and Zisserman, 2014) that was re-trained to classify different kinds of CBC. Similar image monitoring schemes have also been proposed in the field of aviation. Two different ImageNet-trained CNNs are compared in Malekzadeh et al. (2017) as feature detectors, the output of which is combined using an SVM to detect defects in aircraft fuselage. Svensén et al. (2018) use VGG16-based CNN models (Simonyan and Zisserman, 2014) to process images of jet engines and jet engine parts. In a first step, a CNN is used to separate conventional camera photographs from images collected during borescope inspections of engines. The latter are subsequently analysed for part detection using another CNN. For the first task, a substantial proportion of the available images could be reliably labelled using auxiliary labels or reliable heuristics whereas for the second task, images were manually labelled. Jet engine parts, in particular, turbine blades are also the focus of Bian et al. (2016), who used a fully convolutional neural network to process images of turbine blades that were collected in a dedicated imaging rig, following removal from the engine. This more elaborate data collection process enabled accurate segmentation of regions where the protective coating of the blade had been (partially) lost, which can be used as an indicator of the condition of the blade.

### 4. DL PHM applications for sensor condition monitoring data

#### 4.1. Supervised learning algorithms for PHM applications

Supervised learning is the machine learning task of learning a function that maps an input to an output based on example input-output pairs (Friedman et al., 2001; Stuart et al., 2003; Goodfellow et al., 2016). Classic supervised tasks in other fields include image classification (Krizhevsky et al., 2012; He et al., 2016), and text categorization (Cavnar et al., 1994; Sebastiani, 2002; Johnson and Zhang, 2017). To tackle these tasks, researchers in other fields have developed various kinds of deep learning structures. Some of them are particularly useful for PHM applications.

**Recurrent structures.** Recurrent neural networks (Hopfield, 1982; Rumelhart et al., 1988; Graves et al., 2013; Graves, 2013, especially LSTM (Gers, 1999; Graves and Schmidhuber, 2005) networks have been widely used for PHM applications (Yuan et al., 2016; Malhotra et al., 2016; Zhao et al., 2016, 2017; Wu et al., 2018). After the invention of LSTM, this line of work has been heavily developed in other fields, especially natural language processing. The method was further improved by bidirectional LSTM (Graves and Schmidhuber, 2005), and Gated Recurrent Unit (GRU) (Heck and Salem, 2017; Dey and Salemt, 2017). In recent years, the *attention mechanism* (Zhou et al., 2016) has improved the performance of LSTM networks on time series data significantly, especially on language tasks. Networks trained using the attention mechanism (Vaswani et al., 2017; Devlin et al., 2018; Radford et al., 2019) currently hold the state-of-the-art performance on a series of language tasks. This recent progress, however, is not yet fully exploited for PHM applications.

**Time series to image encoding.** Motivated by the success of CNNs on image representation learning (Deng et al., 2009; He et al., 2016), there has been a rising trend on understanding time series by translating them into images. By doing so, existing knowledge on image understanding and image representation learning can be directly exploited for PHM applications. An intuitive approach (Krummenacher et al., 2017) is to simply use the natural plot of 1-D time series data, signal vs time, as two-dimensional image. Alternatively, as shown in Fig. 2, Gramian Angular Fields (GAF), Markov Transition Fields (MTF) and Recurrence Plots (RP) have been introduced in Wang and Oates (2015), Debayle

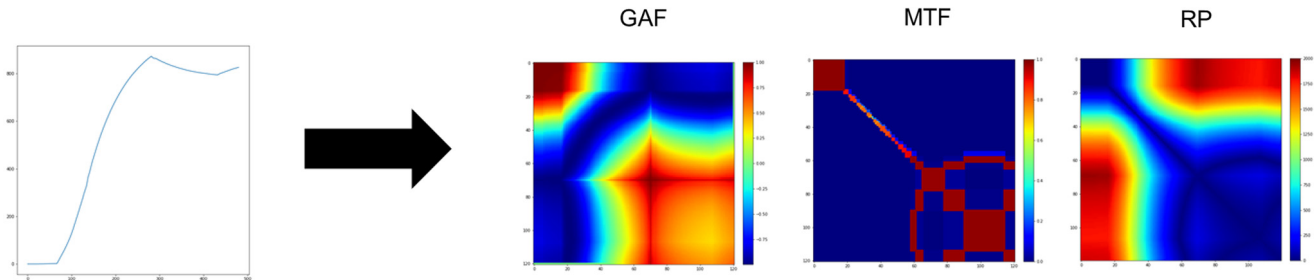


Fig. 2. Examples of 2D-encodings of time series with GAF, MTF and RP.

et al. (2018) as encoding approaches to translate signals to images. These encoding methods were adopted by Gecgel et al. (2019) for vibration data failure detection. Furthermore, time–frequency analysis also results in two-dimensional signal representations. The evaluation of the benefits of the different time series to image encodings and the comparison of their performance for different types of time series data and different fault types is a currently an open research question (Rodriguez Garcia et al., 2020).

**1-D CNN.** An alternative approach is to accept the fact that sensor data are often 1-D data by nature instead of 2-D images. In order to still benefit from recent progress in convolutional operations, 1-D CNN kernels, instead of 2-D kernels, can be used for time series classification, anomaly detection in time series or RUL prediction. The same idea has been widely adopted for motor fault diagnosis (Ince et al., 2016), and broader PHM applications (Babu et al., 2016; Jing et al., 2017; Li et al., 2018; Wang et al., 2019b).

**Combinations of LSTM and CNN.** Both RNNs and CNNs have shown their applicability in understanding time series for PHM applications. It is, thus, natural to leverage the power of both RNNs and CNNs. Existing works (Canizo et al., 2019; Zheng et al., 2019) in PHM mainly adopt a sequential approach, which first extracts local features using CNNs and then feeds them to LSTMs for temporal understanding. An alternative solution (Xingjian et al., 2015; Stollenga et al., 2015) is to embed convolution within the LSTM cell, such that convolutional structures are in both the input-to-state and state-to-state transitions. This approach has been widely adopted in video understanding (Lotter et al., 2016; Byeon et al., 2018) and volumetric medical data analysis (Chen et al., 2016; Litjens et al., 2017). Recently, Zhang et al. (2019b) adopted this idea and proposed to use ConvLSTM for anomaly detection and fault diagnosis on multivariate time series data.

#### 4.2. Unsupervised and semi-supervised learning for PHM applications

Supervised learning requires labels or supervision for every training sample. However, this can be a restrictive requirement, as labels can be expensive or even infeasible to collect. In reality, we would also like to learn partly or fully from unlabelled data. This leads us to the field of unsupervised and semi-supervised learning. In PHM applications, a label can constitute either a health indicator (for fault detection tasks), a specific fault type (for fault classification tasks), or the RUL at each time step of measurements (for prognostics tasks). Particularly in safety critical systems or in systems with high availability requirements, faults are rare and components are replaced or refurbished preventively before reaching their end of life. Thereby, the lifetime of the components is truncated and the true lifetime remains unknown. Furthermore, assessment of the true system health is often too expensive or may be impracticable. This results in a lack of labels for many PHM applications.

Therefore, labels cannot be used to learn the relevant patterns. Other approaches need to be developed to tackle these PHM challenges. These are mostly unsupervised and semi-supervised learning approaches.

One of the most popular unsupervised learning approaches applied in PHM is **signal reconstruction**, partly also referred to as residual-based approaches. The general idea behind the signal reconstruction is to define a model that is able to learn the normal system behaviour and to distinguish it afterwards from system states that are dissimilar to those observed under normal operating conditions. This technique is also known as novelty or anomaly detection. Generally, the implementation of signal reconstruction can be performed either with physics-based but also increasingly with data-driven approaches. In this context, *autoencoders* (Hinton and Zemel, 1993) are typically applied. Particularly for the data-driven signal reconstruction approaches, it is important to identify representative system conditions and to use condition monitoring data from these conditions to train the algorithms and learn the underlying functional relationships. The residuals are then used to detect the abnormal system conditions. Deep autoencoders (Bengio et al., 2006) can potentially capture more complex relationships between the condition monitoring signals and may hence be able to detect more subtle deviations from the representative system conditions. Since the autoencoders are an unsupervised learning architecture, they do not require large amounts of data to be labelled.

Clustering, adaptive dictionary learning (Lu et al., 2019) or causal maps combined with multivariate statistics (Chiang et al., 2015) are some examples of other unsupervised learning approaches applied for fault detection but also for fault isolation. Traditionally, clustering was applied directly to the preprocessed condition monitoring data or the handcrafted features (Detroja et al., 2006; Hu et al., 2012; Li and Zhou, 2014; Du et al., 2014; Costa et al., 2015; Zhu et al., 2017). Clustering approaches, particularly adaptive density clustering algorithms, can distinguish well the different fault types if the feature space (i.e. either the raw CM data or manually extracted features) is separable in fault types. However, distinguishing between fault types becomes increasingly difficult when there is a large number of CM signals, i.e. when the dimensionality of the input space increases and when these signals are additionally noisy and correlated. In such cases, it is often difficult to interpret distances in the feature space and, therefore, to define a valid proximity or similarity metric. As a result, some fault types can become inseparable, leading to clusters with mixed fault types. Therefore, recently clustering was applied on the more informative latent feature representation (Yoon et al., 2017; Arias Chao et al., 2019). In this case, again autoencoders are first applied to compress the high-dimensional condition monitoring data into lower dimensional representation of the input space and then clustering is performed on this informative latent space.

When working with large data sets where only small subsets of the samples are labelled, **semi-supervised learning** approaches have been applied (Oliver et al., 2018; Zhu et al., 2003; Chapelle et al., 2006; Zhu and Goldberg, 2009). Semi-supervised learning leverages the available unlabelled data to improve the performance of the supervised learning task for which different concepts have been proposed. These include (deep) generative models (Kingma et al., 0000), graph-based methods (Zhao et al., 2015) and transductive methods (Bruzzone et al., 2006), to name a few. A further possibility to distinguish the different semi-supervised learning approaches is to differentiate between

those based on consistency regularization (Lee, 2013; Tarvainen and Valpola, 2017; Miyato et al., 2018), entropy minimization and the traditional regularization (Zhu et al., 2003; Chapelle et al., 2006). We refer interested readers to the review paper (Oliver et al., 2018).

Recently, motivated by its successful use in supervised learning, the MixUp (Zhang et al., 2017a) method, mixing labelled and unlabelled data, has been found to be another promising approach to semi-supervised learning (Berthelot et al., 2019; Verma et al., 2019; Wang et al., 2019a).

Some of the semi-supervised approaches have also been applied to PHM applications (Shi, 2018; Listou Ellefsen et al., 2019), including self-training (Yoon et al., 2017), graph-based methods (Zhao et al., 2015) and co-training methods (Hu et al., 2011). Most of the contributions on semi-supervised learning for PHM applications have been focusing on the prediction of the RUL (Yoon et al., 2017; Listou Ellefsen et al., 2019).

In situations where unlabelled data is abundant but acquiring corresponding labels is difficult, expensive or time-consuming, the concept of active learning has been flourishing. Active learning (Settles, 2009) enables machine learning algorithms to achieve a higher accuracy with less labelled data, provided the algorithms are allowed to choose the training samples to be labelled out of the large pool of unlabelled data. Chosen data are labelled by “querying an oracle”, which might be a human expert or another method that affords labelling only a subset of the available data (Sener and Savarese, 2017).

Fault detection can also be defined as a **one-class classification** problem (Michau et al., 2017). However, differently to previously proposed one-class formulations (Schölkopf et al., 2000), the problem is defined in Michau et al. (2017) as a regression problem where the healthy system conditions during training are mapped to a constant target value  $T$ . If presented with operating conditions that are dissimilar to those observed during training, the output of the trained network will deviate from the target value. The problem formulation can be also considered as an unbounded similarity score to the training dataset with known healthy system conditions.

#### 4.3. Uncertainty quantification and decision support

In PHM, especially prognostics, it is of paramount importance to take into account uncertainty inherent in models and data. Predicting RUL as a single number, for instance, is illusory and can be deceptive. Decisions made on the basis of a deterministic model for variables that contain randomness are bound to be flawed.

Sources of uncertainty are several. First, the uncertainty inherent in the models, or epistemic uncertainty: physics-based degradation models are approximations of reality and usually contain unknown parameters. Second, any data acquired through sensors are affected by measurement errors: this is measurement uncertainty. And, last but not least, the future operating profile and loading of equipment being monitored is not known with certainty, yet it is key in predicting the evolution of degradations and how soon they can lead to failures.

As a result, the RUL, in order to be meaningful, must at the very least be accompanied by confidence intervals and, which is even better, by a description through probability distributions if at all possible, or by fuzzy representations. The case is made in Sankararaman and Goebel (2015) that a sharp distinction must be operated between test-based health management and condition-based health management: the former, being based on a population, can resort to more classical (frequentist) statistical techniques, while the latter, focused on one single asset, can only rely on Bayesian techniques. This is because, in the case of one single asset, variability among specimens does not occur.

The challenge is to propagate over time the uncertainties on both model and asset initial state. The distribution over future states and RUL are the outcomes of uncertainty propagation and should not be assumed to have a certain form (for instance, assumed to be Gaussian

without any justification). In most cases, the only practical method to estimate the uncertainty in the RUL is through Monte Carlo simulation and uncertainty propagation because, most of the time, the RUL is a nonlinear function of the initial state, and therefore, will not follow a Gaussian distribution even if the initial state does. Prognostics is only useful if it can aid decision making. According to what precedes, one has to deal with decision making under uncertainty.

Beyond uncertainty propagation, the need therefore arises for uncertainty management. For instance, if the variance of the RUL is very large, how can one deal with the uncertainty in input conditions to reduce the uncertainty on the RUL.

A survey of uncertainty propagation techniques in DL has been published in Abdelaziz et al. (2015). Applications involve image processing and natural language processing.

Not surprisingly, given the nonlinear nature of DNNs, Monte Carlo simulation is the preferred method for propagating uncertainty; but, in view of the computational cost of simulations, approximations are being sought, such as responses surface techniques (Ji et al., 2019) which rely on the property that DNN outputs are influenced by just a few inputs and the sensitivity to various inputs can be assessed by evaluating gradients.

An interesting approach, introduced by MIT in this context (Titenksy et al., 2018), is the use of the extended Kalman filter (EKF): if the layer in the deep network is used as the time step and the value as the state, it is possible to propagate uncertainty in the initial input but also process noise (the latter results from errors in biases and weights of the pre-trained network). In spite of what has been stated above in terms of frequentist versus Bayesian approaches, useful insights can be derived from consideration of the mean residual life (MRL) as the expectation of RUL (the expectation can be seen as Bayesian but possibly also as frequentist if a large number of similar scenarios are considered or if many identical assets in a fleet are considered). Recent work (Huynh et al., 2017) has made progress towards including a probabilistic description of RUL in maintenance optimization approaches. Also, in some categories of problems (Dersin, 2018), it has been shown that the variability of RUL, measured by a coefficient of variation, decreases with the rate of ageing (i.e. the rate at which the MRL decreases as a function of time), and useful methods of traditional reliability engineering can be brought to bear on uncertainty quantification of the RUL (Dersin, 2019). A typical example of decision support problem that relies on PHM results is that of dynamic maintenance planning: deciding when preventive maintenance operations should be performed on assets of a fleet (such as trains or aircraft) on the basis of RUL predictions communicated by the various assets and on logistic and operational constraints (such as spare part management policies and schedule adherence constraints). See Herr et al. (2017) for a railway application that involves multi-agent modelling.

## 5. Promising DL directions for PHM applications

### 5.1. Transfer learning

Most DL methods assume that training and test data are drawn from the same distribution. However, the difference in operating conditions often leads to a considerable distribution discrepancy. Consequently, applying a model trained on one machine to another similar machine that is operated differently may result in a performance drop. Therefore, to apply a DL model for a fleet of machines, collecting, labelling, and re-training models for each machine is a series of necessary but tedious steps. Unsupervised domain adaptation techniques, as a subtopic in transfer learning, provide a promising solution to this problem by transferring knowledge from one well-understood machine to another in the fleet, without the need of labels from target machines.

In addition, we are facing more challenges when adopting transfer learning or domain adaptation methods from a general setup to the industrial settings. For example, in the ideal general setup for domain



adaptation, the source and target tasks have the same input space and same output space. This is particularly not practical in industrial setups, where the set of input sensor signals is likely to be different across different machines, and the set of output labels (types of faults, RUL range) may also be different between different machines. Limited number of existing research works (Wang et al., 2020; Jiao et al., 2019) focus on the case where the output label space is not identical in source and target.

Before the wide adaptation of deep learning approaches, domain adaptation was mainly tackled by learning invariant features across domains (Saenko et al., 2010; Pan et al., 2010; Duan et al., 2012; Zhang et al., 2013), or assigning weights to source samples based on their relevance to the target (Gong et al., 2012). Before the rise of adversarial training, distribution alignment approaches by statistics matching (Long et al., 2015; Yang et al., 2016; Carlucci et al., 2017) focused on using statistical moments to mitigate the domain difference. In recent years, adversarial approaches (Ganin and Lempitsky, 2015; Tzeng et al., 2017; Ganin et al., 2016), where a separate discriminator is used to align the distributions, are yielding superior performance on many different tasks.

Domain adaptation methods have recently also been introduced to PHM applications, especially on fault diagnosis problems. AdaBN (Zhang et al., 2017b), adversarial training (Zhang et al., 2018; Wang et al., 2019b), and MMD-minimization (Li et al., 2018, 2019) were used to align the full source and target distributions for rotating machines. In addition to these domain adaptation approaches, there is a recent trend of direct transfer learning without explicit distribution alignment in PHM applications. These works (Shao et al., 2018; Cao et al., 2018) convert raw signals into graphical images as input, and then fine tune ImageNet pre-trained image classification models for their own fault diagnosis tasks. Despite its simplicity, this approach leads to promising results on fault diagnosis for gearbox, and induction motors.

In summary, many domain adaptation and transfer learning methods have been proposed in recent years. However, the setups used by these methods need to be refined in order to meet the needs of PHM applications. How do we deal with the case where input and output space of the source and target domain are not identical? How can we efficiently encode the input data to make them more transferable? How do we deal with the quantity imbalance between healthy and fault data? These are the questions which require more research.

## 5.2. DL for fleet approaches

Failures occurring in critical systems, such as power or railway systems, are rare and the fault patterns are often unique to the specific system configurations and specific operating conditions. Therefore, they cannot be directly transferred to a different system of the same fleet operated under different conditions. There are different perspectives on the fleets of complex systems. The most commonly used perspective is the one where a fleet is defined as a set of homogeneous systems with corresponding characteristics and features, operated under different conditions, not necessarily by a single operator (Leone et al., 2017). A good example is a fleet of gas turbines or a fleet of cars produced by one manufacturer with different system configurations, operated under different conditions in different parts of the world. The biggest challenge in the fleet PHM is the high variability of the system configurations and the dissimilarity of the operating conditions. Due to the limited number of occurring faults, the fault patterns from dissimilar operating conditions within the fleet need to be used to learn and extract relevant informative patterns across the entire fleet. Several approaches have been proposed for fleet PHM (Michau and Fink, 2019a). For RUL predictions, the existing approaches rely on the assumption that systems that are parts of a fleet all have similar operating conditions and that selected parts of degradation trajectories can be directly transferred from one system of the fleet to another. However, the life time of systems within a fleet depends on many

different factors, including operating and environmental conditions, material properties and variability in manufacturing. Therefore, this assumption is not valid for the majority of complex industrial assets.

This poses a significant challenge on transferring a model developed on one unit to other units of the fleet. Most of the research, has been focusing so far on this perspective, which is already challenging in terms of variability of configurations and operating conditions (Michau et al., 2018). For the monitoring and diagnosis of such fleets, different approaches have been proposed (organized by increasing complexity) (Michau and Fink, 2019a):

1. Identifying some relevant operating or design parameters of the units (e.g. average operating regimes) in order to find sub-fleets, possibly with clustering, defined by similar characteristics based on the selected parameters; subsequently using the data of each of the sub-fleets to train the algorithms (Zio and Di Maio, 2010).
2. Using the entire time series of condition monitoring signals to perform time series cluster analysis of units (Liu, 2018).
3. Developing models for the functional behaviour of the units and identifying similar units following this learned functional behaviour (Michau et al., 2018). Contrary to the time series clustering, the approaches do not depend on the length of the observation time period.
4. Performing domain alignment in the feature space of the different units to compensate for the distribution shift between different units of a fleet (Michau and Fink, 2019a).

While the methods used at each complexity level of the fleet approaches in the list above overcome some of the limitations of the methods used at the previous level, the complexity of the proposed solutions increases. The main limitations of each of these approaches are (Michau and Fink, 2019b) :

1. Aggregated parameters used for comparison may not cover all the relevant conditions or the aggregated parameters may not be representative of the unit specificities.
2. Comparing the distances between time series is affected by the curse of dimensionality. Time series cluster analysis becomes even more challenging when operating conditions evolve over time.
3. Even though approaches that learn the functional behaviour of units are more robust to variations in the behaviour of the system, one of the underlying requirements is that the units experience a sufficient similarity in their operating regimes. If the units are operated in a dissimilar way, large fleets may be required to find units with a sufficient similarity (since the similarity is defined at the unit level).
4. Since the alignment is performed in an unsupervised way and the performance depends on the assumption that the future operating conditions of the unit of interest will be representative to the aligned operating conditions, no guarantees can be made that the system of interest will be behaving in a similar way in the future. However, this limitation is in fact true for all the fleet PHM approaches since the past experience of other fleet units is transferred to the unit of interest.

Therefore, the challenges for designing a PHM system for fleets from the manufacturer's perspective are extensive (Michau and Fink, 2019a) due to the variability of the system configurations and the operating conditions. However, designing a PHM system for a fleet from the perspective of an operator is even more challenging. In this case, a fleet is defined as a set of systems fulfilling the same function not necessarily originating from the same manufacturer and typically not monitored by the same set of sensors. A good example of such a fleet is a fleet of turbines in a hydro-power plant that may have been taken into operation at different points in time and that were produced by different manufacturers. The units of a fleet share in this case a



similar behaviour, fulfil a similar functionality but may be monitored by a different set of sensors, at different locations, different sampling frequencies, etc. This line of research is particularly interesting for the industry. However, this problem has not yet been addressed in data-driven PHM research. Solving this problem will provide a major leap forward for many industrial applications.

### 5.3. Deep generative algorithms

Within the last years, Deep Generative Models have emerged as a novel way to learn a probabilistic distribution without any assumption of the induced family distribution. Indeed, unlike Gaussian mixtures that assume that the training data are sampled from a mixture of Gaussian distributions whose number of modes and covariance matrices are known beforehand, deep generative models use a neural network, usually denoted as a decoder or a generator, that maps a sample from a noise distribution to a sample from the ground-truth distribution. The deep generative models vary in their implementation mainly from the induced metric distribution used in their optimization scheme.

The most famous example of deep generative models are Generative Adversarial Networks (GANs) (from Goodfellow et al. (2014)). The training of the Generator  $G$  is based on an adversarial loss. The training of the generator is combined with the learning of a second network, the Discriminator  $D$  whose goal is to distinguish correctly between real data and generated data. The training will alternate the optimization for both the Discriminator and the Generator until the quality of the generated samples is sufficient. Despite their success, GANs have suffered from several limitations, mainly due to their training scheme: the loss used to train the generator depends on the result of classification of the generated data (does each generated data point fool the discriminator or not?). Thus, once the algorithm discovered how to generate examples that fool the discriminator, the loss leads the training process to generate similar examples and not discover new examples. This leads to mode collapse (generated data only represents a few number of modes from the original data).

To address this problem, recent research focused on using a metric distribution called the 1-Wasserstein distance to assess the global similarity between generated and original data. The Wasserstein distance is based on the theory of optimal transport to compare data distributions with wide applications in image processing, computer vision, and machine learning. Wasserstein Generative Adversarial Networks (W-GAN, see in Arjovsky et al. (2017)) were first introduced as a solution to the mode collapse problem. Indeed, since the Wasserstein distance is continuous and is a global function, it forces the network not to focus on a subset of the distribution. Wasserstein GAN learns a generator on a noise distribution (generally Gaussian) so that the output distribution matches the ground-truth distribution. In its primal form, the Wasserstein distance requires to measure the expectation of the distances between two continuous distributions which may not be tractable in high dimension. Instead, the formulation of W-GAN relies on the dual expression of the 1-Wasserstein distance, which allows better optimization properties.

Several works apply generative deep modelling, either variational autoencoders (VAEs) or GANs for anomaly detection (An and Cho, 2015; Schlegl et al., 2017; Deecke et al., 2018). The principle is as follows: they learn the induced distribution and then assert whether a sample was part of this distribution, by mapping it to the closest sample in the generated distribution. They tackled this mapping, either by adapting an autoencoder learnt during the training like VAEs or DCAEs (An and Cho, 2015; Masci et al., 2011; Donahue et al., 2016), or by doing optimization scheme in the latent dimension directly (Schlegl et al., 2017; Deecke et al., 2018).

Schlegl et al. (2017) and Deecke et al. (2018), whose methods are respectively denoted as AnoGAN and AD-GAN, optimize the latent dimension with gradient descent to approximate at best a given input. If they do succeed in reconstructing the data up to a certain threshold,

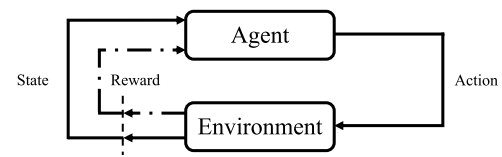


Fig. 3. Agent-environment interaction, based on Sutton and Barto (1998).

then the data is recognized as normal. Otherwise, it is detected as anomalous.

When it comes to generating time series data, previous works proposed to use Recurrent Neural Networks for both the generator  $G$  and the discriminator  $D$  to take into account the sequential nature of the input data. Eventually, they generated fake time series with sequences from a random noise space. Note that those works still require the duration of the input signal to be fixed.

None of the research studies in Schlegl et al. (2017) and Deecke et al. (2018) yield an efficient way of estimating the latent value that generated a given input data. Moreover, training GANs is a challenging optimization task, even more, when considering non-image data, so that adding the encoder during the training stage as in Donahue et al. (2016) may harness the generated samples quality.

Recent works by Schlegl et al. (2019) and Ducoffe et al. (2019) combine W-GAN and an encoder to perform anomaly detection on medical images and time series respectively.

### 5.4. Deep reinforcement learning

Due to the dynamic and complex nature of corrective and predictive maintenance tasks, it is challenging to react to maintenance requests effectively within a short time. In operations research, scheduling and planning problems are mostly solved by mathematical modelling if the (simplified) problem can be properly formulated, or by heuristic search if a locally optimal solution is acceptable. However, neither of these approaches takes into account previous solutions to similar problem instances. That is, given any problem instance, solutions need to be searched and found from scratch. Deep reinforcement learning (Mnih et al., 2015) provides potential to adjust to new problems quickly by utilizing experience and knowledge gained from solving old problems.

Reinforcement learning (RL) systems (Sutton and Barto, 1998) learn a mapping from situations to actions by trial-and-error interactions with a dynamic environment in order to maximize an expected future reward. After performing an action in a given state the RL agent will receive some reward from the environment in the form of a scalar value and it learns to perform actions that will maximize the sum of the rewards received when starting from some initial state and proceeding to a terminal state. By choosing actions, the environment moves from state to state. This interaction is visualized in Fig. 3. Many RL systems have been proposed including Q-Learning (a model-free reinforcement learning algorithm), State-Action-Reward-State-Action (SARSA), and actor-critic learning (Sutton and Barto, 1998).

However, for problems with a very large state space, regular RL falls short in learning strategies since the mapping between state-action pairs and rewards is stored in a lookup table. Recent developments in the Deep Reinforcement Learning (DRL), where DL is used as a function approximator for RL, have shown that strategies can be derived for difficult and complex problems. For instance, deep Q-network (DQN), the DRL model that mastered the game of Go (Silver et al., 2016) uses CNNs as a function approximator for the optimal action-value function to replace the Q-table in Q-learning.

DRL can be for example used in PHM to decide whether to take a maintenance action as shown in Knowles et al. (2010). The age and condition can be used to represent the state of a component and the possible actions will be “maintenance” or “no maintenance”. A profit will be returned as a reward if the component does not fail when “no

|  |   |
|--|---|
| Training Data                                      |   |
| Hypothesis Space                                   |   |
| Deterministic: Model Structure,<br>Hyperparameters | Probabilistic: Graphical Struct.,<br>Probability Distribution |
| Training Algorithm                                 |   |
| Deterministic: Loss Function,<br>Regularization    | Probabilistic: Bayesian Priors                                |
| Final Hypothesis                                   |   |

Fig. 4. Different ways of integrating prior knowledge in machine learning, derived from von Rueden et al. (2019).

maintenance” is chosen as action. If the component does fail then a repair cost will be deducted from the profit. If “maintenance” is chosen as the action, a maintenance cost is deducted from the profit but there will be no repair cost since the component will not fail. Typically, the maintenance cost is considerably lower than the failure cost. Thus at each time step, the DRL model must decide based on the status of the component between a moderate reward by performing maintenance or risking no maintenance which could incur either a high reward in the event of no failure or a low reward if the component fails.

Another application of DRL in PHM could be using DRL to plan the order of tasks to be handled on assets as given in Peer et al. (2018) to derive a feasible schedule so that all assigned tasks can be finished with the right order in time.

The major challenge of applying DRL in PHM lies in reward shaping (Marek, 2017). Reward shaping is to design a reward function that encourages the DRL model to learn certain behaviours. It is usually hand-crafted and needs to be carefully designed by experienced human experts. However, it is not always easy to shape the rewards properly and improperly shaped rewards can bias learning and can lead to behaviours that do not match expectations. Moreover, reinforcement learning does not learn on fixed datasets and ground truth targets, and therefore could be unstable in its performance.

### 5.5. Physics-induced machine learning

A promising approach of inducing interpretability in the machine learning models, particularly for applications outside the image processing domain, where visualizations may not be readily derived, is physics-induced machine learning. Prior knowledge is integrated in the models, delivering improvements in terms of performance as well as interpretability. Several studies have shown that integrating prior knowledge in the learning process of neural networks does not only help to significantly reduce the amount of required data samples but also to improve the performance of the learning algorithms (Karpatne et al., 2017; Tipireddy et al., 2019).

Different approaches have been proposed for integrating domain knowledge in the learning process, including physics-informed neural networks (Raissi et al., 2019; Raissi, 2018), physics-guided neural networks (Karpatne et al., 2017), semantic-based regularization (Diligenti et al., 2017) and integration of logic rules (Hu et al., 2016a). As illustrated in Fig. 4, integration of the domain knowledge can be performed at different levels of the learning process, such as in the training data, the hypothesis space, the training algorithm and the final hypothesis. In addition, the way how the knowledge is represented and transformed differs quite significantly between the different approaches. It includes simulation, statistical relations, symmetries and constraints (von Rueden et al., 2019). In the field of fault diagnosis and prognosis, the direction of hybrid modelling and physics-informed approaches has only recently been explored with different directions (Nascimento and Viana, 2019; Arias Chao et al., 2019; Arias Chao et al., 2020). The underlying physical models are integrated to enhance the input space,

providing additional information from the physical models to the learning algorithms. Integrating physical constraints and physical models in the AI algorithms will change the model interpretability and also the extrapolation abilities of the algorithms and is subject to further research.

## 6. Future research needs

Deep learning provides several promising directions for PHM applications. In terms of methodological advancements, the research directions listed in the previous section: transfer learning, fleet approaches, generative models, reinforcement learning, and physics-induced machine learning hold great potential, but also highlight the need for more research.

The main focus of domain adaptation research has been on homogeneous cases when source and target input space contain the same features. However, in real applications, complex systems will not only experience a domain shift but will also have different features. In the context of complex industrial systems, heterogeneous domains comprise the case of similar systems produced by different manufacturers and equipped with different sets of condition monitoring sensors (not only sensor locations differ but also sensor types and number of sensors). The research on heterogeneous unsupervised domain adaptation, particularly applied to complex physical systems has been very limited but has a large potential to be impactful, particularly for industrial applications. Another perspective to address the challenges is to make use of simulation environment and adapt from simulation to real-life applications. This direction is particularly interesting as the data will more likely be sufficient in the source domain.

Generative models have been mainly applied to computer vision tasks. The evaluation of the plausibility of the generated samples in terms of physical system coherence is an open research question. Also, controlling the generation of the relevant data samples from the perspective of the physical processes is similarly an open research question.

Reinforcement learning has been particularly thriving for applications with extensive and detailed simulation environment. However, this condition cannot be fulfilled for many complex real applications, particularly in the context of PHM. Additionally, the levels of complexity of the problems to which RL has been typically applied in the context of PHM have been still comparably limited. The extension to more complex problems requires additional research.

Combination of DL with expert knowledge is a potentially fertile research area as models can be enriched dynamically with acquired data, thus leading to effective digital twins that can provide support to maintenance decision making. While several directions are currently pursued in physics-induced machine learning, there is neither a consensus nor a consolidation on the different directions and how they could be transferred to industrial applications. Further research is required to develop and consolidate these approaches, which may also result in improvements in the interpretability of the developed models and methods.

Representative datasets have been key drivers of research and innovation in many leading fields in DL, including computer vision and natural language processing. DL methods require large and representative data sets. However, in the context of PHM, the lack of representative datasets has been obstructing a broad usage and adaptation of DL approaches in industrial applications. There are several solutions to this challenge: data augmentation, data generation, application of physics-induced machine learning models or rather from the organization point of view: sharing the data across companies' borders.

If insufficient data are available, ML models tend not to generalize well and there is an increasing risk of overfitting. In DL, one technique that mitigates that risk is data augmentation (Perez and Wang, 2017), which has been shown to improve the performance of DL models in image classification. Recently, also automatic approaches for data

augmentation have been proposed, such as AutoAugment (Cubuk et al., 2019). However, research for data augmentation on time series data has been limited. An investigation into how data augmentation can be applied in PHM contexts, particularly for time series data, is one of the potential research directions.

Recently, several approaches have been proposed to generate faulty samples or fault features with generative neural networks (Yu et al., 2019; Zhou et al., 2020). However, most of the studies focused on vibration data and signals that were pre-processed to make them image-like. An interesting research direction is to evaluate the transferability of such approaches to more complex datasets and time-series data. A further evaluation of the physical plausibility of the generated samples and the effects of the different generated faults on the performance of the algorithms is also required.

An additional challenge that needs to be addressed in future research is effective and efficient composition and selection of the training datasets. This is particularly relevant in evolving environments with highly varying operating conditions where the training dataset is not fully representative of the full range of the expected operating conditions. A decision needs to be taken continuously if the newly measured data should be included in the training dataset and the algorithms should be updated or if the information is redundant and already contained in the dataset used for training the algorithms. This becomes even more complex if the decision on including the newly measured data in the training dataset is taken on the fleet level and not only on the single unit level. The research also goes in the direction of active learning: selecting the observations that will have the largest improvement on the algorithm's performance.

Uncertainty propagation should also be addressed more comprehensively, in particular by developing or enhancing Monte-Carlo simulation methods that guarantee sufficient accuracy with reasonable computation load in presence of nonlinear, non-Gaussian, non-stationary stochastic processes. Furthermore, synergies with traditional reliability engineering techniques may be exploited more effectively.

DL provides many promising directions in PHM with the potential to improve the availability, safety and cost efficiency of complex industrial assets. However, there are also several requirements for the industrial stakeholders that need to be fulfilled before a significant progress can be realized. These requirements include automation and standardization of data collection, including particularly maintenance and inspection reports and implementation of data sharing across several stakeholders; as well as perhaps generally accepted ways of assessing data quality. An additional hurdle for a broad implementation of DL in PHM is one of the general hurdles in PHM: development of suitable business models and underlying legal frameworks that foster collaboration between stakeholders and provide benefits for all the involved partners.

## Declaration of competing interest

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

## CRediT authorship contribution statement

**Olga Fink:** Conceptualization, Writing - review & editing. **Qin Wang:** Conceptualization, Writing - review & editing. **Markus Svensén:** Conceptualization, Writing - review & editing. **Pierre Dersin:** Conceptualization, Writing - review & editing. **Wan-Jui Lee:** Conceptualization, Writing - review & editing. **Melanie Ducoffe:** Conceptualization, Writing - review & editing.

## Acknowledgements

The contributions of Olga Fink and Qin Wang were funded by the Swiss National Science Foundation (SNSF) Grant no. PP00P2\_176878.

## References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P.A., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zhang, X., 2016. TensorFlow: A system for large-scale machine learning. In: OSDI.
- Abdelaziz, A.H., Watanabe, S., Hershey, J.R., Vincent, E., Kolossa, D., Un, D.K., Hussen Abdelaziz, A., 0000. Uncertainty propagation through deep neural networks. Tech. rep.
- Adomavicius, G., Tuzhilin, A., 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* 17, 734–749.
- Al-Rfou', R., Alain, G., Almahairi, A., Angermüller, C., Bahdanau, D., Ballas, N., Bastien, F., Bayer, J., Belikov, A., Belopolsky, A., Bengio, Y., Bergeron, A., Bergstra, J., Bisson, V., Snyder, J.B., Bouchard, N., Boulanger-Lewandowski, N., Bouthillier, X., de Brébisson, A., Breuleux, O., Carrier, P.L., Cho, K., Chorowski, J., Christiano, P.F., Cooijmans, T., Côté, M.-A., Côté, M., Courville, A.C., Dauphin, Y., Delalleau, O., Demouth, J., Desjardins, G., Dieleman, S., Dinh, L., Ducoffe, M., Dumoulin, V., Kahou, S.E., Erhan, D., Fan, Z., Firat, O., Germain, M., Glorot, X., Goodfellow, I.J., Graham, M., Gülçehre, C., Hamel, P., Harlouchet, I., Heng, J.-P., Hidasi, B., Honari, S., Jain, A., Jean, S., Jia, K., Korobov, M.V., Kulkarni, V., Lamb, A., Lamblin, P., Larsen, E., Laurent, C., Lee, S., cois, S.L., Lemieux, S., Léonard, N., Lin, Z., Livezey, J.A., Lorenz, C., Lowin, J., Ma, Q., Manzagol, P.-A., Mastropietro, O., McGibbon, R., Memisevic, R., van Merriënboer, B., Michalski, V., Mirza, M., Orlandi, A., Pal, C.J., Pascanu, R., Pezeshki, M., Raffel, C., Renshaw, D., Rocklin, M., Romero, A., Roth, M., Sadowski, P., Salvatier, J., Savard, F., Schlüter, J., Schulman, J., Schwartz, G., Serban, I., Serdyuk, D., Shabanian, S., Simon, É., Spieckermann, S., Subramanyam, S.R., Sygnowski, J., Tanguay, J., van Tulder, G., Turian, J.P., Urban, S., Vincent, P., Visin, F., de Vries, H., Warde-Farley, D., Webb, D.J., Willson, M., Xu, K., Xue, L., Yao, L., Zhang, S., Zhang, Y., 2016. Theano: A python framework for fast computation of mathematical expressions. *ArXiv abs/1605.02688*.
- Amodei, D., Anantharayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., Casper, J., Catanzaro, B., Chen, J., Ch�ranowski, M., Coates, A., Diamos, G., Elsen, E., Engel, J., Fan, L., Fougner, C., Hannun, A.Y., Jun, B., Han, T., LeGresley, P., Li, X., Lin, L., Narang, S., Ng, A.Y., Ozair, S., Prenger, R., Qian, S., Raiman, J., Satheesh, S., Seetapun, D., Sengupta, S., Sriram, A., Wang, C., Wang, Y., Wang, Z., Xiao, B., Xie, Y., Yagatama, D., Zhan, J., Zhu, Z., 2015. Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin. In: ICML.
- An, J., Cho, S., 2015. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lect. IE 2*, 1–18.
- Arias Chao, M., Adey, B.T., Fink, O., 2019. Knowledge-induced learning with adaptive sampling variational autoencoders for open set fault diagnostics. *arXiv, arXiv:1912.00000*.
- Arias Chao, M., Kulkarni, C., Goebel, K., Fink, O., 2019. Hybrid deep fault detection and isolation: Combining deep neural networks and system performance models. *Int. J. Progn. Health Manag.* 10 (033).
- Arias Chao, M., Kulkarni, C., Goebel, K., Fink, O., 2020. Fusing physics-based and deep learning models for prognostics. *arXiv preprint arXiv:2003.00732*.
- Arjovsky, M., Chintala, S., Bottou, L., 2017. Wasserstein gan. *arXiv preprint arXiv:1701.07875*.
- Atamuradov, V., Medjaher, K., Lamoureux, B., Dersin, P., Zerhouni, N., 0000. Fault detection by Segment evaluation based on Inferential Statistics for Asset monitoring. In: Annual PHM Society Conference, St-Petersburg, Florida.
- Babu, G.S., Zhao, P., Li, X.-L., 2016. Deep convolutional neural network based regression approach for estimation of remaining useful life. In: International Conference on Database Systems for Advanced Applications. Springer, pp. 214–228.
- Banjevic, D., 2009. Remaining useful life in theory and practice. *Metroika* 69 (2–3), 337–349.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (8), 1798–1828.
- Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., 2006. Greedy layer-wise training of deep networks. In: Schölkopf, B., Platt, J.C., Hofmann, T. (Eds.), *Advances in Neural Information Processing Systems 19: Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, December 4–7, 2006. MIT Press, pp. 153–160.
- Bennett, J., Lanning, S., 2007. The Netflix prize. In: In KDD Cup and Workshop in Conjunction with KDD. ACM.
- Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C., 2019. Mixmatch: A holistic approach to semi-supervised learning. *arXiv preprint arXiv:1905.02249*.
- Bian, X., Lim, S.-N., Zhou, N., 2016. Multiscale fully convolutional network with application to industrial inspection. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1–8.
- Bishop, C.M., 2007. *Pattern Recognition and Machine Learning*, fifth ed. In: *Information Science and Statistics*, Springer.
- Bruzzone, L., Chi, M., Marconcini, M., 2006. A novel transductive SVM for semisupervised classification of remote-sensing images. *IEEE Trans. Geosci. Remote Sens.* 44 (11), 3363–3373.
- Byeon, W., Wang, Q., Kumar Srivastava, R., Koumoutsakos, P., 2018. ContextVP: Fully Context-Aware Video Prediction. In: The European Conference on Computer Vision (ECCV).



- Canizo, M., Triguero, I., Conde, A., Onieva, E., 2019. Multi-head CNN-RNN for multi-time series anomaly detection: An industrial case study. *Neurocomputing* 363, 246–260.
- 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.
- Carlucci, F.M., Porzi, L., Caputo, B., Ricci, E., Bulò, S.R., 2017. AutoDIAL: Automatic domain alignment layers. In: *ICCV*, pp. 5077–5085.
- Cavnar, W.B., Trenkle, J.M., et al., 1994. N-gram-based text categorization. In: *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, vol. 161175. Citeseer.
- Chapelle, O., Schölkopf, B., Zien, A., 2006. Semi-supervised learning. MIT Press, p. 508.
- Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z., 2016. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In: *Advances in Neural Information Processing Systems*, pp. 3036–3044.
- Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G.S., Chai, W., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., Shah, H., 2016. Wide & Deep Learning for Recommender Systems. In: *DLRS@RecSys*.
- Chiang, L.H., Jiang, B., Zhu, X., Huang, D., Braatz, R.D., 2015. Diagnosis of multiple and unknown faults using the causal map and multivariate statistics. *J. Process Control* 28, 27–39.
- Coates, A., Huval, B., Wang, T., Wu, D.J., Catanzaro, B., Ng, A.Y., 2013. Deep learning with COTS HPC systems. In: *ICML*.
- Costa, B.S.J., Angelov, P.P., Guedes, L.A., 2015. Fully unsupervised fault detection and identification based on recursive density estimation and self-evolving cloud-based classifier. *Neurocomputing* 150, 289–303.
- Covington, P., Adams, J., Sargin, E., 2016. Deep Neural Networks for YouTube Recommendations. In: *RecSys*.
- Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V., 2019. Autoaugment: Learning augmentation strategies from data. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 113–123.
- Debayle, J., Hatami, N., Gavet, Y., 2018. Classification of time-series images using deep convolutional neural networks. In: Zhou, J., Radeva, P., Nikolaev, D., Verikas, A. (Eds.), *Tenth International Conference on Machine Vision (ICMV 2017)*, vol. 10696. SPIE, p. 23.
- Deecke, L., Vandermeulen, R., Ruff, L., Mandt, S., Kloft, M., 2018. Image anomaly detection with generative adversarial networks. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, pp. 3–17.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 248–255.
- Dersin, P., 2018. The class of life time distributions with a mean residual life linear in time. In: *Safety and Reliability – Safe Societies in a Changing World*. CRC Press, pp. 1093–1099.
- Dersin, P., 2019. New Investigations into mean residual life (MRL) and remaining useful life (RUL). In: *Proc. of Mathematical Methods in Reliability Conference (MMR 2019)*, Hong Kong, June.
- Detroja, K., Gudi, R., Patwardhan, S., 2006. A possibilistic clustering approach to novel fault detection and isolation. *J. Process Control* 16 (10), 1055–1073.
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Dey, R., Salemt, F.M., 2017. Gate-variants of gated recurrent unit (GRU) neural networks. In: *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, pp. 1597–1600.
- Diligenti, M., Gori, M., Sacca, C., 2017. Semantic-based regularization for learning and inference. *Artificial Intelligence* 244, 143–165.
- Donahue, J., Krähenbühl, P., Darrell, T., 2016. Adversarial feature learning. *arXiv preprint arXiv:1605.09782*.
- Du, Z., Fan, B., Jin, X., Chi, J., 2014. Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. *Build. Environ.* 73, 1–11.
- Duan, L., Tsang, I.W., Xu, D., 2012. Domain transfer multiple kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (3), 465–479.
- Ducoffe, M., Haloui, I., Sen Gupta, J., 2019. Anomaly detection on time series with wasserstein GAN applied to PHM. In: *PHM Applications of Deep Learning and Emerging Analytics*. International Journal of Prognostics and Health Management Reviewed (Special Issue).
- Fiedler, N., Bestmann, M., Hendrich, N., 2018. Imagetagger: An open source online platform for collaborative image labeling. In: *Robot World Cup*. Springer, pp. 162–169.
- Fink, O., 2020. Data-driven intelligent predictive maintenance of industrial assets. In: *Women in Industrial and Systems Engineering*. Springer, pp. 589–605.
- Forman, G., 2003. An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.* 3 (Mar), 1289–1305.
- Friedman, J., Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning, Vol. 1. Springer Series in Statistics New York.
- Ganin, Y., Lempitsky, V., 2015. Unsupervised domain adaptation by backpropagation. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37*. JMLR. org, pp. 1180–1189.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V., 2016. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.* 17 (1), 2096–2030.
- Garofolo, J.S., Lamel, L.F., Fisher, W.M., Fiscus, J.G., Pallett, D.S., Dahlgren, N.L., 1993. DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM. NIST.
- Gegel, O., Ekworo-Osire, S., Dias, J.P., Serwadda, A., Alemayehu, F.M., Nispel, A., 2019. Gearbox fault diagnostics using deep learning with simulated data. In: *2019 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, pp. 1–8.
- Gers, F., 1999. Learning to forget: continual prediction with LSTM. *IET Conf. Proc.* 850–855(5).
- Gibert, X., Patel, V.M., Chellappa, R., 2015. Deep multitask learning for railway track inspection. *IEEE Trans. Intell. Transp. Syst.* 18, 153–164.
- Godfrey, J.J., Holliman, E., 1997. Switchboard-1 Release 2. NIST.
- Gong, B., Shi, Y., Sha, F., Grauman, K., 2012. Geodesic flow kernel for unsupervised domain adaptation. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 2066–2073.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. In: *Advances in Neural Information Processing Systems*, pp. 2672–2680.
- Graves, A., 2013. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.
- Graves, A., Mohamed, A.-r., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, pp. 6645–6649.
- Graves, A., Schmidhuber, J., 2005. Framework phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* 18 (5–6), 602–610.
- Guo, L., Li, N., Jia, F., Lei, Y., Lin, J., 2017. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* 240, 98–109.
- Haensch, W., Gokmen, T., Puri, R., 2018. The next generation of deep learning hardware: analog computing. *Proc. IEEE* 107 (1), 108–122.
- He, K., Gkioxari, G., Dollár, P., Girshick, R.B., 2017. Mask R-CNN. *IEEE Trans. Pattern Anal. Mach. Intell.*
- He, Y., Sainath, T.N., Prabhavalkar, R., McGraw, I., Alvarez, R., Zhao, D., Rybach, D., Kannan, A., Wu, Y., Pang, R., Liang, Q., Bhatia, D., Shangguan, Y., Li, B., Pundak, G., Sim, K.C., Bagby, T., Chang, S.-Y., Rao, K., Gruenstein, A., 2018. Streaming end-to-end speech recognition for mobile devices. In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6381–6385.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition. *CoRR abs/1512.03385*, *arXiv:1512.03385*.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: *European Conference on Computer Vision*. Springer, pp. 630–645.
- Heck, J.C., Salem, F.M., 2017. Simplified minimal gated unit variations for recurrent neural networks. In: *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*. IEEE, pp. 1593–1596.
- Herr, N., Nicod, J.M., Varnier, C., Zerhouni, N., Dersin, P., 2017. Predictive maintenance of moving systems. In: *2017 Prognostics and System Health Management Conference, PHM-Harbin 2017 - Proceedings*. Institute of Electrical and Electronics Engineers Inc.
- Hieber, F., Domhan, T., Denkowski, M., Vilar, D., Sokolov, A., Clifton, A., Post, M., 2017. Sockeye: A toolkit for neural machine translation. *ArXiv abs/1712.05690*.
- Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* 29, 82–97.
- Hinton, G.E., Osindero, S., Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. *Neural Comput.* 18, 1527–1554.
- Hinton, G.E., Zemel, R.S., 1993. Autoencoders, Minimum Description Length and Helmholtz Free Energy. In: *NIPS*.
- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.* 79 (8), 2554–2558.
- Hu, Z., Ma, X., Liu, Z., Hovy, E., Xing, E.P., 2016a. Harnessing deep neural networks with logic rules. In: *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, vol. 4. Association for Computational Linguistics (ACL), pp. 2410–2420, *arXiv:1603.06318*.
- Hu, Y., Palmé, T., Fink, O., 2016b. Deep health indicator extraction: A method based on auto-encoders and extreme learning machines. In: *PHM 2016*, Denver, CO, USA, 3–6 October 2016. PMH Society, pp. 446–452.
- Hu, C., Youn, B.D., Kim, T., 2011. Semi-supervised learning with co-training for data-driven prognostics. In: *Volume 2: 31st Computers and Information in Engineering Conference, Parts a and B*. ASME, pp. 1297–1306.
- Hu, C., Youn, B.D., Wang, P., Taek Yoon, J., 2012. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliab. Eng. Syst. Saf.*
- Huynh, K.T., Grall, A., Bérenguer, C., 2017. Assessment of diagnostic and prognostic condition indices for efficient and robust maintenance decision-making of systems subject to stress corrosion cracking. *Reliab. Eng. Syst. Saf.* 159, 237–254.



- Ince, T., Kiranyaz, S., Eren, L., Askar, M., Gabbouj, M., 2016. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Ind. Electron.* 63 (11), 7067–7075.
- Ji, W., Ren, Z., Law, C.K., 2019. Uncertainty propagation in deep neural network using active subspace. *arXiv:1903.03989*.
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R.B., Guadarrama, S., Darrell, T., 2014. Caffe: Convolutional architecture for fast feature embedding. *CoRR abs/1408.5093*, *arXiv:1408.5093*.
- Jiao, J., Zhao, M., Lin, J., Ding, C., 2019. Classifier inconsistency based domain adaptation network for partial transfer intelligent diagnosis. *IEEE Trans. Ind. Inf.* 1.
- Jing, L., Zhao, M., Li, P., Xu, X., 2017. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement* 111, 1–10.
- Johnson, R., Zhang, T., 2017. Deep pyramid convolutional neural networks for text categorization. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 562–570.
- Kalchbrenner, N., Blunsom, P., 2013. Recurrent Continuous Translation Models. In: *EMNLP*.
- Karpathy, A., 2015. The unreasonable effectiveness of recurrent neural networks.
- Karpathy, A., Fei-Fei, L., 2015. Deep visual-semantic alignments for generating image descriptions. In: *CVPR*.
- Karpatne, A., Watkins, W., Read, J., Kumar, V., 2017. Physics-guided neural networks (PGNN): An application in lake temperature modeling. *arXiv:1710.11431*.
- Kingma, D.P., Rezende, D.J., Mohamed, S., Welling, M., 0000. Semi-supervised Learning with Deep Generative Models. *Tech. rep.*
- Knowles, M., Baglee, D., Wermter, S., 2010. Reinforcement learning for scheduling of maintenance. In: *The Thirtieth SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pp. 409–422.
- Krizhevsky, A., 2009. Learning multiple layers of features from tiny images.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: *Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25*. Curran Associates, Inc., pp. 1097–1105.
- Krummenacher, G., Ong, C.S., Koller, S., Kobayashi, S., Buhmann, J.M., 2017. Wheel defect detection with machine learning. *IEEE Trans. Intell. Transp. Syst.* 19 (4), 1176–1187.
- LeCun, Y., Boser, B.E., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W.E., Jackel, L.D., 1989. Handwritten Digit Recognition with a Back-Propagation Network. In: *NIPS*.
- Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. In: *Proceedings of the IEEE*, pp. 2278–2324.
- LeCun, Y., Cortes, C., 2005. The mnist database of handwritten digits.
- Lee, D.-H., 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: *Workshop on Challenges in Representation Learning, ICML*, vol. 3, p. 2.
- Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D., 2014. Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mech. Syst. Signal Process.* 42 (1–2), 314–334.
- Leone, G., Cristaldi, L., Turrin, S., 2017. A data-driven prognostic approach based on statistical similarity: An application to industrial circuit breakers. *Measurement* 108, 163–170.
- Li, F.-F., Fergus, R., Perona, P., 2004. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. In: *2004 Conference on Computer Vision and Pattern Recognition Workshop*, pp. 178–178.
- Li, X., Zhang, W., Ding, Q., 2018. Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks. *IEEE Trans. Ind. Electron.*
- Li, X., Zhang, W., Ding, Q., Sun, J.-Q., 2019. Multi-layer domain adaptation method for rolling bearing fault diagnosis. *Signal Process.* 157, 180–197.
- Li, C., Zhou, J., 2014. Semi-supervised weighted kernel clustering based on gravitational search for fault diagnosis. *ISA Trans.* 53 (5), 1534–1543.
- Lipton, Z.C., 2016. The mythos of model interpretability. *ArXiv abs/1606.03490*.
- Listou Ellefsen, A., Bjørlykhaug, E., Aesøy, V., Ushakov, S., Zhang, H., 2019. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliab. Eng. Syst. Saf.* 183, 240–251.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88.
- Liu, Z., 2018. Cyber-physical system augmented prognostics and health management for fleet-based systems (Ph.D. thesis). University of Cincinnati.
- Liu, L., Tan, E., Zhen, Y., Yin, X.J., Cai, Z.Q., 2018. AI-facilitated coating corrosion assessment system for productivity enhancement. In: *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 606–610.
- Long, M., Cao, Y., Wang, J., Jordan, M.I., 2015. Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791*.
- Lotter, W., Kreiman, G., Cox, D., 2016. Deep predictive coding networks for video prediction and unsupervised learning. *arXiv preprint arXiv:1605.08104*.
- Lu, Y., Xie, R., Liang, S.Y., 2019. Adaptive online dictionary learning for bearing fault diagnosis. *Int. J. Adv. Manuf. Technol.* 101 (1–4), 195–202.
- Malekzadeh, T., Abdollahzadeh, M., Nejati, H., Cheung, N.-M., 2017. Aircraft fuselage defect detection using deep neural networks. *ArXiv abs/1712.09213*.
- Malhotra, P., TV, V., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., Shroff, G., 2016. Multi-sensor prognostics using an unsupervised health index based on lstm encoder-decoder. *arXiv preprint arXiv:1608.06154*.
- Marek, G., 2017. Reward Shaping in Episodic Reinforcement Learning. In: *Proceedings of the 16th Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-17)*, pp. 565–573.
- Masci, J., Meier, U., Cireşan, D., Schmidhuber, J., 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. In: *International Conference on Artificial Neural Networks*. Springer, pp. 52–59.
- Michau, G., Fink, O., 0000. Unsupervised fault detection in varying operating conditions. In: *2019 IEEE International Conference on Prognostics and Health Management*.
- Michau, G., Fink, O., 2019b. Domain adaptation for one-class classification: monitoring the health of critical systems under limited information. *International Journal of Prognostics and Health Management* 10, 11.
- Michau, G., Hu, Y., Palmé, T., Fink, O., 2020. Feature learning for fault detection in high-dimensional condition monitoring signals. *Proc. Inst. Mech. Eng. Part O* 234 (1), 104–115.
- Michau, G., Palmé, T., Fink, O., 2017. Deep feature learning network for fault detection and isolation. In: *Conference of the PHM Society*.
- Michau, G., Palmé, T., Fink, O., Fleet PHM for critical systems: Bi-level deep learning approach for fault detection. In: *Proceedings of the European Conference of the PHM Society*, vol. 4.
- Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. Efficient estimation of word representations in vector space. In: *Proceedings of Workshop at ICLR*.
- Miyata, T., Maeda, S.-i., Koyama, M., Ishii, S., 2018. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 41 (8), 1979–1993.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M.A., Fidjeland, A., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D., 2015. Human-level control through deep reinforcement learning. *Nature* 518, 529–533.
- Montavon, G., Samek, W., Müller, K.R., 2018. Methods for interpreting and understanding deep neural networks. *Digital Signal Process. Rev. J.* 73, 1–15, *arXiv:1706.07979*.
- Nascimento, R.G., Viana, F.A., 2019. Fleet prognosis with physics-informed recurrent neural networks. In: *Structural Health Monitoring 2019: Enabling Intelligent Life-Cycle Health Management for Industry Internet of Things (IIOT) - Proceedings of the 12th International Workshop on Structural Health Monitoring*, vol. 2, pp. 1740–1747. *arXiv:1901.05512*.
- Oliver, A., Odena, A., Raffel, C., Cubuk, E.D., Goodfellow, I.J., 2018. Realistic evaluation of semi-supervised learning algorithms.
- Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q., 2010. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* 22 (2), 199–210.
- Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A., 2017. Automatic differentiation in PyTorch. In: *NIPS Autodiff Workshop*.
- Peer, E., Menkovski, V., Zhang, Y., Lee, W.-J., 2018. Shunting trains with deep reinforcement learning. In: *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE-SMC, pp. 3063–3068.
- Perez, L., Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L., 2018. Deep contextualized word representations. In: *Proc. of NAACL*.
- Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. In: *Proceedings of the IEEE*. IEEE, pp. 257–286.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., 2019. Language models are unsupervised multitask learners. *OpenAI Blog* 1 (8).
- Raissi, M., 2018. Deep hidden physics models: Deep learning of nonlinear partial differential equations. *J. Mach. Learn. Res.* 19, 1–24.
- Raissi, M., Perdikaris, P., Karniadakis, G., 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.
- Ravi, N.N., Drus, S.M., Krishnan, P.S., Ghani, N.L.A., 2019. Substation transformer failure analysis through text mining. In: *2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. IEEE, pp. 293–298.
- Ribeiro, M.T., Singh, S., Guestrin, C., 2016. “Why should i trust you?” Explaining the predictions of any classifier. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Vol. 13–17-Aug. ACM Press, New York, New York, USA, pp. 1135–1144.
- Ricci, F., Rokach, L., Shapira, B., 2011. *Introduction to recommender systems handbook*. In: *Recommender Systems Handbook*. Springer.
- Rodriguez García, G., Michau, G., Ducoffe, M., Gupta, J.S., Fink, O., 2020. Time series to images: Monitoring the condition of industrial assets with deep learning image processing algorithms. In: *European Prognostics and Health Management Conference*.

- von Rueden, L., Mayer, S., Garcke, J., Bauckhage, C., Schuecker, J., 2019. Informed machine learning - Towards a taxonomy of explicit integration of knowledge into machine learning. *arXiv:1903.12394*.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1988. Learning representations by back-propagating errors. *Cogn. Model.* 5, 3.
- Saenko, K., Kulis, B., Fritz, M., Darrell, T., 2010. Adapting visual category models to new domains. In: *European Conference on Computer Vision*. Springer, pp. 213–226.
- Samek, W., Müller, K.R., 2019. Towards explainable artificial intelligence. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. In: LNCS, vol. 11700, Springer Verlag, pp. 5–22, *arXiv:1909.12072*.
- Sankararaman, S., Goebel, K., 2015. Uncertainty in prognostics and systems health management. *Int. J. Progn. Health Manag.* 6.
- Saxena, A., Sankararaman, S., Goebel, K., 0000. Performance Evaluation for fleet-based and unit-based prognostic methods. In: *2nd European Conference of the Prognostics and Health Management Society*.
- Schlegl, T., Seeböck, P., Waldstein, S.M., Langs, G., Schmidt-Erfurth, U., 2019. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Med. Image Anal.* 54, 30–44.
- Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G., 2017. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: *International Conference on Information Processing in Medical Imaging*. Springer, pp. 146–157.
- Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J., 2000. *Support Vector Method for Novelty Detection*. MIT Press, pp. 582–588.
- Sebastiani, F., 2002. Machine learning in automated text categorization. *ACM Comput. Surv.* 34 (1), 1–47.
- Seide, F., Agarwal, A., 2016. CNTK: Microsoft's open-source deep-learning toolkit. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*. ACM, New York, NY, USA, p. 2135.
- Sener, O., Savarese, S., 2017. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*.
- Settles, B., 2009. *Active Learning Literature Survey*. University of Wisconsin-Madison.
- Shao, S., McAleer, S., Yan, R., Baldi, P., 2018. Highly accurate machine fault diagnosis using deep transfer learning. *IEEE Trans. Ind. Inf.* 15 (4), 2446–2455.
- Sharma, V., Parey, A., 2016. A review of gear fault diagnosis using various condition indicators. *Procedia Eng.* 144, 253–263.
- Shi, Z., 2018. Semi-supervised ensemble learning methods for enhanced prognostics and health management (Ph.D. thesis). University of Cincinnati.
- Si, X.-S., Wang, W., Hu, C.-H., Zhou, D.-H., 2011. Remaining useful life estimation—a review on the statistical data driven approaches. *European J. Oper. Res.* 213 (1), 1–14.
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T.P., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 484–489.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556*.
- Stollenga, M.F., Byeon, W., Liwicki, M., Schmidhuber, J., 2015. Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation. In: *Advances in Neural Information Processing Systems*. pp. 2998–3006.
- Stuart, R., Peter, N., et al., 2003. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, USA.
- Su, C.-J., Tsai, L.-C., Huang, S.-F., Li, Y., 2019. Deep learning-based real-time failure detection of storage devices. In: *International Conference on Applied Human Factors and Ergonomics*. Springer, pp. 160–168.
- Sutskever, I., Vinyals, O., Le, Q.V., 2014. Sequence to Sequence Learning with Neural Networks. In: *NIPS*.
- Sutton, R.S., Barto, A.G., 1998. *Introduction to Reinforcement Learning*, first ed. MIT Press, Cambridge, MA, USA.
- Svensén, M., Powrie, H., Hardwick, D., 2018. Deep neural networks analysis of borescope images. In: *Proceedings of the European Conference of the PHM Society*, vol. 4.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going deeper with convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9.
- Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: *Advances in Neural Information Processing Systems*. pp. 1195–1204.
- Tesauro, G., 1992. Practical issues in temporal difference learning. *Mach. Learn.* 8 (3–4), 257–277.
- Tipireddy, R., Perdikaris, P., Stinis, P., Tartakovsky, A., 2019. A comparative study of physics-informed neural network models for learning unknown dynamics and constitutive relations. *arXiv:1904.04058*.
- Titensky, J.S., Jananthan, H., Kepner, J., 2018. Uncertainty propagation in deep neural networks using extended kalman filtering. *arXiv:1809.06009*.
- Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7167–7176.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. In: *Advances in Neural Information Processing Systems*. pp. 5998–6008.
- Verma, V., Lamb, A., Kannala, J., Bengio, Y., Lopez-Paz, D., 2019. Interpolation consistency training for semi-supervised learning. *arXiv preprint arXiv:1903.03825*.
- Wang, Q., Li, W., Gool, L.V., 2019. Semi-Supervised Learning by Augmented Distribution Alignment. In: *The IEEE International Conference on Computer Vision (ICCV)*.
- Wang, Q., Michau, G., Fink, O., 2019. Domain adaptive transfer learning for fault diagnosis. In: *2019 Prognostics and System Health Management Conference*.
- Wang, Q., Michau, G., Fink, O., 2020. Missing-class-robust domain adaptation by unilateral alignment. *IEEE Trans. Ind. Electron.* 1.
- Wang, Z., Oates, T., 2015. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In: *Workshops At the Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- Wu, Y., Schuster, M., Chen, Z., Le, Q.V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G.S., Hughes, M., Dean, J., 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *ArXiv abs/1609.08144*.
- Wu, Y., Yuan, M., Dong, S., Lin, L., Liu, Y., 2018. Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing* 275, 167–179.
- Xingjian, S., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., Woo, W.-c., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: *Advances in Neural Information Processing Systems*. pp. 802–810.
- Yan, W., Yu, L., 2015. On accurate and reliable anomaly detection for gas turbine combustors: a deep learning approach. In: *Proceedings of the Annual Conference of the Prognostics and Health Management Society*.
- Yang, Z., Cohen, W.W., Salakhutdinov, R., 2016. Revisiting semi-supervised learning with graph embeddings. *arXiv preprint arXiv:1603.08861*.
- Yoon, A.S., Lee, T., Lim, Y., Jung, D., Kang, P., Kim, D., Park, K., Choi, Y., 2017. Semi-supervised learning with deep generative models for asset failure prediction. *arXiv:1709.00845*.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In: *NIPS*.
- Young, T., Hazarika, D., Poria, S., Cambria, E., 2017. Recent trends in deep learning based natural language processing [Review Article]. *IEEE Comput. Intell. Mag.* 13, 55–75.
- Yu, Y., Tang, B., Lin, R., Han, S., Tang, T., Chen, M., 2019. CWGAN: Conditional wasserstein generative adversarial nets for fault data generation. In: *2019 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, pp. 2713–2718.
- Yuan, M., Wu, Y., Lin, L., 2016. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. In: *2016 IEEE International Conference on Aircraft Utility Systems (AUS)*. IEEE, pp. 135–140.
- Zhai, Q., Ye, Z.-S., 0000. RUL prediction of deteriorating products using an adaptive wiener process model. *IEEE Trans. Ind. Inf.* 13.
- Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D., 2017a. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.
- Zhang, B., Li, W., Zhang, M., Tong, Z., 2018. Adversarial adaptive 1-d convolutional neural networks for bearing fault diagnosis under varying working condition. *arXiv preprint arXiv:1805.00778*.
- Zhang, A., Lipton, Z.C., Li, M., Smola, A.J., 2019a. Dive into deep learning.
- Zhang, W., Peng, G., Li, C., Chen, Y., Zhang, Z., 2017b. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. *Sensors* 17 (2), 425.
- Zhang, K., Schölkopf, B., Muandet, K., Wang, Z., 2013. Domain adaptation under target and conditional shift. In: *International Conference on Machine Learning*, pp. 819–827.
- Zhang, C., Song, D., Chen, Y., Feng, X., Lumezanu, C., Cheng, W., Ni, J., Zong, B., Chen, H., Chawla, N.V., 2019b. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 1409–1416.
- Zhang, S., Yao, L., Sun, A., Tay, Y., 2019c. Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Surv.* 52 (1), 5:1–5:38.
- Zhao, Y., Ball, R., Mosesian, J., de Palma, J.-F., Lehman, B., 2015. Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays. *IEEE Trans. Power Electron.* 30 (5), 2848–2858.
- Zhao, R., Wang, J., Yan, R., Mao, K., 2016. Machine health monitoring with LSTM networks. In: *2016 10th International Conference on Sensing Technology (ICST)*. IEEE, pp. 1–6.
- Zhao, R., Yan, R., Wang, J., Mao, K., 2017. Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors* 17 (2), 273.
- Zheng, L., Xue, W., Chen, F., Guo, P., Chen, J., Chen, B., Gao, H., 2019. A fault prediction of equipment based on CNN-LSTM network. In: *2019 IEEE International Conference on Energy Internet (ICEI)*. IEEE, pp. 537–541.

- Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B., 2016. Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 207–212.
- Zhou, F., Yang, S., Fujita, H., Chen, D., Wen, C., 2020. Deep learning fault diagnosis method based on global optimization GAN for unbalanced data. *Knowl.-Based Syst.* 187, 104837.
- Zhu, X., Ghahramani, Z., Lafferty, J.D., 2003. Semi-supervised learning using gaussian fields and harmonic functions. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 912–919.
- Zhu, X., Goldberg, A.B., 2009. Introduction to semi-supervised learning. *Synth. Lect. Artif. Intell. Mach. Learn.* 3 (1), 1–130.
- Zhu, K., Mei, F., Zheng, J., 2017. Adaptive fault diagnosis of HVCBs based on P-SVDD and P-KFCM. *Neurocomputing* 240, 127–136.
- Zio, E., Di Maio, F., 2010. A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. *Reliab. Eng. Syst. Saf.* 95 (1), 49–57.