Acoustics Australia
https://doi.org/10.1007/s40857-021-00222-9

REVIEW PAPER

Recent Advancement of Deep Learning Applications to Machine Condition Monitoring Part 1: A Critical Review

Wenyi Wang[1] · John Taylor[1,2] · Robert J. Rees[1]

Received: 9 December 2020 / Accepted: 16 February 2021

Abstract

With the huge success of applying deep learning (DL) methodologies to image recognition and natural language processing in recent years, researchers are now keen to use them in the machine condition monitoring (MCM) context. There are numerous papers in applying various DL techniques, such as auto-encoder, restricted Boltzmann machine, convolutional neural network and recurrent neural network, etc., to MCM problems ranging from component-level condition monitoring (machine tool wear prediction, bearing fault diagnosis and classification and hydraulic pump fault diagnosis) to system-level health management (aircraft and spacecraft diagnosis). In this paper, we give a brief overview in the area of DL for MCM with a focus on reviewing the most recent papers published since 2019. In Part 1, we present some critical views regarding whether any breakthrough has been achieved from an MCM domain expert perspective, with the main conclusion that DL has great potential for MCM applications, and a major breakthrough could come soon since the shortfalls lie more in data than in the DL methodologies. Our overall impression is that (a) DL models are not really showing their great potentials with only a small training data; (b) faulty-condition data is hard to come by for training DL, but normal condition data is abundant, so anomaly detection makes more sense; (c) applying DL only to the Case Western Reserve University (CWRU) bearing fault dataset is not sufficient for real world industrial applications as it was from a very simple test rig, and applying DL to data from complex systems like helicopter gearbox data may deliver much more convincing results. In Part 2, we enhance the main conclusion of the critical review with supplement views and a case study on analysing Bell-206B helicopter main gearbox planet bearing failure data using some traditional MCM techniques in contrast to applying the long short-term memory (LSTM) DL method. We can conclude from the case study that the DL-based methods are not necessarily always superior to the traditional MCM techniques for dataset from moderately complex machinery.

Keywords Deep Learning · LSTM · Machine condition monitoring · Bearing fault diagnosis · Helicopter planet bearing

1 Introduction

Deep learning (DL) methodologies have attracted enormous attention with their success in applications to image recognition and natural language processing since 2012. Many researchers have been keen to use them in machine condition monitoring (MCM). Over the last few years, research into applying DL to machine health monitoring and fault diagnostics has been growing exponentially with hundreds of research papers and a number of recent review papers. Zhao et al. [1] provided a thorough review in this area in 2019, which has attracted many citations (763 as of 28 Jan 2021 by Google Scholar). This paper reviewed a wide range of papers applying DL techniques, such as auto-encoder (AE), restricted Boltzmann machine (RBM), convolutional network (CNN), recurrent neural network (RNN), etc., to MCM problems ranging from component-level condition monitoring (machine tool wear prediction, bearing fault diagnosis and classification and hydraulic pump fault diagnosis) to system-level health management (aircraft and spacecraft diagnosis). For DL-based bearing fault analysis alone, we have found four review papers by Zhang et al. [2], Hoang and Kang [3], Neupane and Soek [4] and Waziralilah et al. [5]. The study in [4] presented a review on recent works of applying DL to bearing fault monitoring

✉ Wenyi Wang
Wenyi.Wang@dst.defence.gov.au

[1] Defence Science and Technology Group, Melbourne, Australia

[2] Data61, CSIRO, Canberra, Australia

Published online: 13 April 2021

with the Case Western Reserve University (CWRU) bearing dataset (referred to as DL + CWRU hereafter) that has been widely used as a standard reference for validating the DL models. Waziralilah et al. [5] reviewed applying CNN to bearing fault diagnosis with comparison of three time–frequency representations as input to the CNN. They found the spectrogram and scalogram can deliver better classification accuracy than the Hilbert–Huang transform (HHT). There is also one review paper for machine tool condition monitoring by Serin et al. [6]. Fink et al. [7] provided a good evaluation of DL applications to the prognostics and health management (PHM) field with an emphasis on current developments, drivers, challenges, potential solutions and future research needs. Tang et al. [8] also gave a review on DL-based fault diagnosis approaches for rotating machinery components with focus primarily on bearings, gear/gearbox and pumps, and discussed the existing challenges and possible future research directions. Rezaeianjouybari and Shang [9] reviewed the PHM applications of the DL models in three categories, i.e. generative, discriminative and hybrid. They also discussed the transfer learning and domain adaptation in the context of PHM and challenges and future research directions. Zhao et al. [10] attempted to design an open-source benchmark study to allow fair and unified comparison between different DL models for MCM using some open-source datasets and codes and discussed some potential future directions in this field. The Tang et al. [11] review emphasised different approaches to generating image data for CNN in MCM, i.e. fast Fourier transform, wavelet transform, S-transform and cyclic spectral analysis for 2D-CNN. They also reviewed data augmentation methods for 1D-CNN.

We can identify at least 11 literature review papers covering the area of DL-based MCM since 2019. Although this demonstrates the popularity of this research area, it also begs the rather uncomfortable question whether it really deserves this much attention and whether these research works have really delivered on the promise of breakthrough performance in comparison to more traditional approaches to MCM problems. In order to provide guidance in determining the future direction of the field, our review is unique in emphasising the question as to whether a significant breakthrough has been achieved. This assessment will be conducted from the perspective of real industrial applications based on our own experience of applying DL to MCM, as subject domain experts in the MCM field. Importantly, to aid the reader in the assessment of the literature, our critical review provides counter arguments against key claims in the reviewed papers. This approach differentiates our review from previous review papers and represents the main contribution of this review.

Due to the page limitation, we divided the paper into two parts. In Part 1, we will first give a brief overview in the area of DL for MCM undertaken prior to 2019 in reference to the reviews in [1]. This will be followed by a critical review of some selected papers published since 2019. In Part 2, we will give some supplement critical views and present a case of comparative study using helicopter main gearbox planet bearing failure data between selected traditional MCM techniques and a DL-based methodology.

2 A Brief Summary of Applying DL to MCM up to 2019

We summarize the number of applications in Table 1 based on the reviews conducted by Zhao et al. [1], where each column in the table contains the applications of one category of DL methodologies. Out of the 156 papers cited in [1], there were 94 cases of applying DL to various MCM fields with bearing fault diagnosis (44 cases) by far the most popular. The review highlighted some impressive results with high accuracy fault classification, e.g. ref #54 with up to 100% accuracy in classifying common mechanical faults in an induction motor using stacked auto-encoder (SAE) and ref #108 with 99.5% accuracy in classifying gearbox faults using a 2D-CNN model with wavelet images as inputs.

3 Most Recent DL Applications to MCM since 2019

Modern mechanical systems are often too sophisticated to be modelled based on our understanding of the system physics; hence, data-based DL methods could be a viable option in modelling system behaviours. That could be the reason why DL applications to MCM is becoming so popular with an exponential growth in the numbers of research papers. Here is a brief overview on the scale of research effort in applying DL to MCM since 2019. We used Google Scholar as of 15 November 2020 to search for the number of research papers under seven most widely used keyword combinations: (a) auto-encoder "machine fault"; (b) "convolutional neural network" "machine fault"; (c) "recurrent neural network" "machine fault"; (d) auto-encoder "bearing fault"; (e) "convolutional neural network" "bearing fault"; (f) "recurrent neural network" "bearing fault" and (g) "deep learning" "case western reserve university".

Due to the large numbers in the search results, we attempted to narrow down the search results by using Google's metrics of timeframe (since 2019) and relevance. The "*relevance*" is a general metric by Google Scholar search engine, it ranks search results by the level of relevance, irrespective of where the papers were published. However, the narrowed search still resulted in a large number of papers, where we knew that not every result is relevant and yet it

Table 1 Number of DL applications to MCM for the papers reviewed in [1]

MCM and fault diagnosis fields of application	Auto-encoder (AE)	Restricted Boltzmann machine (RBM)	Convolutional neural network (CNN)	Recurrent neural network (RNN) & LSTM
Electric motor	1	2	3	
Bearing	18	9	16	1
Gear	4	4	7	1
Hydraulic system	1	1	1	
Tidal turbine	1			
Spacecraft	1			
Air compressor	2			
Aircraft systems	1	4	1	2
Cutting tool	1	1		4
Transformer	1	1		
Rotor system	1	2	1	
Ball Screw		1		

One paper can contain multiple application cases

would be too time-consuming to manually verify every single one of them. Hence, we further narrowed them down to the "top 25-percentile relevance", and followed by a manual verification.

For example, we had 1090 and 2040 results under combinations (b) and (e) since 2019, in which 273 and 510, respectively, were in the top 25-percentile relevance. By manually verifying the keywords in the title, abstract, keywords list and sometimes the full text for those in the top 25-percentile relevance, we finally identified 148 and 287 papers, respectively, as relevant. We repeated this process for combinations (a)–(f). For combination (g), i.e. DL + CWRU, we observed that the level of relevance beyond the top 25-percentile remained too high to ignore so that we changed to the "diminishing relevance" method. Under the diminishing relevance, we only stopped the manual verification when the level of relevance had reduced by 90 percent, i.e. the number of relevant papers fell below 5 per 50 or 10 percent relevance.

Under the 7 keyword combinations, we present the identified numbers of relevant research papers in a *bubble chart* as shown in Fig. 1, where one bubble represents one combination. The number inside each bubble is the number of relevant research papers published since 2019. Figure 1 gives an overall picture of the studies in DL-based MCM since 2019. As can be seen, the CNN is certainly the most popular DL method and the CWRU bearing data the most popular dataset in applying DL to MCM. The green bubble indicates the large number of the DL + CWRU studies.

As we were invited by the journal to write a critical review paper on DL-based MCM with some constraints on page limits, we felt that the sheer number of papers

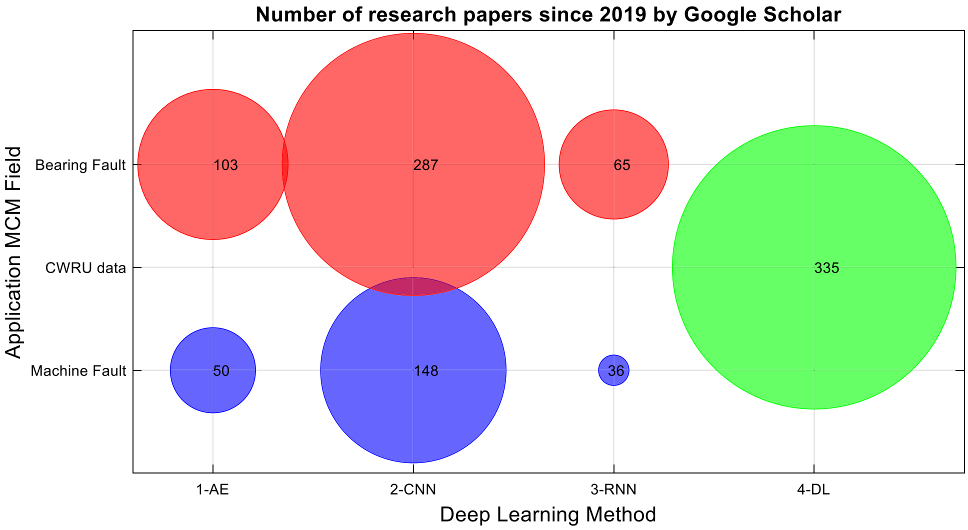

Fig. 1 An overview of the number of relevant research papers in DL-based MCM by Google Scholar (as of 15 November 2020). The result was obtained by using Google Scholar's functionality of "Since 2019" and "Sort by relevance" plus the criterion of "Top 25-percentile relevance" due to the sheer numbers in search results, except for the DL + CWRU studies (green bubble) that was based on "Diminishing relevance"

published since 2019 makes it very challenging to write an ordinary review paper. In addition, we do not intend to write a comprehensive review paper which has been done in previous review papers. As a result, we defined a metric of selecting papers for this critical review, i.e. “top 25-percentile relevance” plus “random selection”. As discussed above, the “top 25-percentile relevance” is used for manual verification to identify a boundary as shown in Fig. 1, and the “random selection” is used within this boundary (i.e. the 690 identified papers). In order to cover papers published in journals, conferences and arXiv preprints, we randomly selected 5% within those 690 identified papers (i.e. 34 papers) in several MCM application fields for this critical review. The rational for the random selection is that some of the highest cited DL papers were conference and arXiv papers, such as the original papers of the CNN (*AlexNet*) for ImageNet in 2012 and the generative adversarial networks (GAN) in 2014. Moreover, other criteria like the “top 25-percentile citation” would probably favour the papers published in 2019, and those published in 2020 would be less likely to be selected.

3.1 DL Applications to Condition Monitoring of Mechanical Components

One of the main applications to MCM is wind turbine gearbox health monitoring. Wind turbines are increasing in number rapidly around the world. They normally have a preinstalled MCM system which accumulates large quantities of data. There are many studies conducted on applying DL to wind turbine CM. He et al. [12] claimed to have achieved promising performance in wind turbine planetary gearbox fault diagnosis by learning the fault features directly from raw vibration data using retrieval neural network with dictionary learning. With no supervised fine-tuning, the dictionary learning network generates a sparse representation of the raw vibration data for the retrieval network classifier to obtain fault diagnosis. However, the seeded fault laboratory test data may not be representative of the unexpected fault types that are occurring so often, which makes it hard for DL models to classify correctly when trained on laboratory data.

In the study by Fu et al. [13], they first used the adaptive elastic network to select the variables from the temperature data of gearbox bearings and then established the logical relationship between observed variables by combining the CNN and the long short-term memory (LSTM) network. Using experimental data, they claim to be able to efficiently achieve over-temperature fault warning for the high-speed bearing in the wind turbine gearbox. However, the detection of bearing over-temperature ($< 1\%$ of life remaining) may not give long enough warning time for operators, and an auto shutdown process needs to be incorporated to prevent catastrophic failure.

Helbing and Ritter [14] studied the application of DL to SCADA (Supervisory Control and Data Acquisition) data where DL application was impeded by relatively low dimensionality in the data and suggested ways of working with higher-dimensional SCADA data. They also conducted an overview of recent applications of artificial neural networks and DL and found that most approaches in the literature are unsupervised. As for supervised approaches, they concluded that applications with high-dimensional input data are quite successful; nevertheless, there are issues with operational data in terms of quality, availability, dimensionality, labels and class imbalance, which limit the applicability of the DL models.

Zhao et al. [15] proposed a method to apply RBM-based deep auto-encoder (DAE) to wind turbine SCADA data for anomaly detection and component health monitoring. They used an abnormality-sensitive condition index “Re” to detect faults in turbine components and employed an adaptive threshold method on the basis of the extreme theory to reduce false alarm rates under gusty wind conditions where wind speed is changing abruptly. The DAE residuals were used for component fault diagnosis and isolation. No comparison was made against any traditional MCM technique to justify the use of DL-based method.

Resendiz-Ochoa et al. [16] conducted research into uniform gear tooth wear classification by artificial neural network (ANN) with reduced dimension statistical time domain features from infrared imaging. They used linear discriminant analysis for dimensionality reduction and compared the results by the proposed method to those by classical vibration condition monitoring approaches. A laboratory seeded fault testing dataset was used for supervised training of the ANN. However, the question is how to obtain the supervised training data in reality and whether or not the ANN trained with seeded data is transferable to operational conditions. Further investigation of the effect of network depths on the classification performance is warranted.

Wang et al. [17] employed a deep machine learning method, i.e. the long short-term memory-based recurrent neural network (LSTM-RNN), as a signal processing tool to analyse vibration data generated in a run-to-failure gear rig test. They claimed that gear tooth crack can be detected by detecting anomalies or changes induced by the gear faults in the LSTM prediction error signal (LSTM-residual) and provided a detailed study on how to apply the LSTM to monitoring gear tooth crack growth at various stages of the fault development in terms of detectability and robustness. They also compared the fault detectability of LSTM-Residual with that of other commonly used residual signals in gear fault diagnostics. They concluded that the LSTM-residual method can be a powerful tool with superior fault detectability provided that the LSTM network is properly trained. They also pointed out that the proposed method can have potential pitfalls especially with the problem of false detection or lack of robustness when data contains any minor perturbations.

Singh et al. [18] made a good contribution to gearbox health condition classification using automatic feature generation and domain adaptation based on a DL method. This is a particularly useful generalization solution when the training dataset has a different distribution from the target (or test) dataset under variable speed and changing health conditions. The CNN training was formulated based on an architecture using two different loss functions simultaneously for domain adaptation, i.e. minimum sum of cross-entropy loss between the labelled source data and maximum mean discrepancy loss between the labelled source and unlabelled target datasets. Evaluation using experimental data from a gearbox under variable speeds and multiple health conditions demonstrated that the proposed method is superior to an appropriate benchmarking with both traditional machine learning and other DL methods. It would be more informative if a comparison was made against traditional MCM techniques to justify the use of DL-based method.

Mallikarjuna et al. [19] explored two deep learning models (LSTM and Bi-LSTM) for feature generation to classify the aircraft gearbox health conditions into good or bad using a publicly available aircraft gearbox vibration dataset and a number of traditional classifiers. They claimed that the proposed DL models have achieved above 98.38 and 99.75% of classification accuracy for LSTM and Bi-LSTM models, respectively, and superior performance as compared to hand-crafted input features by traditional MCM techniques. The performance improvement is very significant from 79.75% by traditional techniques to 99.75 by the Bi-LSTM model with 100, 50 and 30 LSTM hidden cells in the three LSTM layers. They also found that results from time domain data are more accurate and reliable than those from the frequency domain data. The more than 25% performance boost by the DL method demonstrated a clear advantage over the traditional MCM techniques that are challenged by fault diagnosis for complex aircraft gearboxes. In situations where you may not be sure whether the gearbox is bad but you are certain that it is different based on the observed behaviour, a more general question to ask would be whether we can do clustering between normal and abnormal health conditions by generalising the proposed Bi-LSTM method.

Bearing fault analysis and classification using DL has attracted significant attention, where many studies seem to focus on applying CNN to time–frequency domain image data derived from raw vibration signals. There are a number of DL + CWRU studies reviewed here. Li et al. [20] proposed to use CNN with multi-scale features extracted from time–frequency domain data for estimating the remaining useful life (RUL) of rolling element bearings. By validating their technique with experimental data from the PRONOSTIA experimental platform used for the bearing health prognostic challenge at the IEEE PHM 2012 conference, they demonstrated a superior performance in comparisons with other approaches and showed that their method is promising for industrial applications. However, it would be difficult to obtain data in real-world applications for supervised training; hence, it would be essential to be able to transfer learning if it is impossible to source the RUL-labelled data in practice.

Galati et al. [21] used the LSTM network to model the healthy-state pre-processed vibration signals of the Bell-206B helicopter main gearbox and a simpler single-stage spur gearbox. They applied the trained LSTM model to predict the future-state vibration signals, and they used the prediction error signals to trend the health progression of these gearboxes. With data from bearing damage in the Bell-206B planetary stage and gear tooth damage in the spur gearbox, they demonstrated that the LSTM method can effectively trend the changes generated by progressive damages of the planet bearing and the spur gear with a comparable performance to traditional MCM techniques. Application to the gear tooth crack data showed that an LSTM was able to clearly reveal the fault signature as well as produce a monotonic trend for the kurtosis of the prediction error. The application to the helicopter planet-gear bearing spalling fault data showed that whilst fault features were able to be detected by an LSTM model, the trending of the prediction error RMS was not monotonic. They employed some pre-processing steps such as synchronous signal resampling or averaging and band pass filtering, where the information about the gearbox is required.

In the case of bearing RUL prediction, Chen et al. [22] pre-processed vibration data into spectral sub-band energy as selected features for an RNN based on gated recurrent units (GRU) encoder–decoder architecture with attention mechanism to produce the health indicator (HI). They then used linear regression to calculate RUL. By validating the method with the PRONOSTIA dataset, they were able to demonstrate a better performance in comparison to other novel approaches. However, it is not clear why they used a linear equation to fit all the HI's from RNN to predict the bearing RUL or how RNN generated HI's can be linear. In comparison with the results of Li et al. [20], a commonality is that both methods did not use DL to automatically generate features. The PRONOSTIA datasets were from machinery too unsophisticated (where traditional techniques can work just fine) to demonstrate that the DL-based method is really advantageous.

Udmale et al. [23] suggested that a deep sequential model with kurtograms as the input sequence data can deliver an improved performance for bearing fault classification and believed that conversion of raw signals to kurtograms can facilitate strong feature representations of bearing vibration. They used RNN, LSTM and GRU models to process sequential kurtograms and compared the results with the CNN, ELM (extreme learning machine), MLP (multi-level perceptron) and SVM (support vector machines) classifiers. Also, they compared findings with the non-kurtogram features to prove the advantages of using kurtogram as the input. Validated using two sets of bearing vibration data, i.e. the CWRU datasets and a machinery fault simulator (MFS) data (the MFS testbed is a product of SpectraQuest Inc.), they demonstrated that the proposed method has a promising performance with high fault classification accuracy in comparison with other methods. For the MFS data, the highest accuracy of 99.47% was achieved by the RNN model with much shorter training time than the LSTM, GRU and CNN models, whereas LSTM delivered the highest classification accuracy of 98.3% for the CWRU data. The comparative study is quite thorough; however, the data they used were not representative of a sophisticated class of machinery.

Xu et al. [24] proposed a method to apply CNN and random forest (RF) ensemble learning to diagnose bearing faults. They used continuous wavelet transform of the raw time domain vibration signals to generate image data for a CNN model which extracted multi-level features representing the bearing faults. They then employed an ensemble of multiple RF classifiers to diagnose bearing faults. Using two sets of bearing fault data collected from a reliance electric motor (i.e. CWRU dataset) and a rolling mill at BaoSteel, they validated the effectiveness of the method and concluded that the proposed method delivered a good classification accuracy (up to 99.73%) for bearing faults and is superior to traditional methods and standard machine learning methods (BPNN—back propagation neural network, SVM, DAE and DBN). It is noteworthy that the method was validated by both the benchmark CWRU data and data from real-world industrial applications.

Zhao et al. [25] developed an approach to using CNN with time–frequency domain input data for fault classification of planet bearings. A new fault classification algorithm was proposed to detect fault type of the planet bearing. Using the Hilbert transform and the synchro-squeezing transform, they first convert the vibration signal to an enveloped time–frequency representation, using the synchro-squeezing transform which is an extension of the continuous wavelet transform, as the input data to the neural network. They then train the CNN to learn key bearing fault features for fault classification. Their analysis results demonstrated the effectiveness of the approach with an impressive accuracy of over 98% of correctly classifying planet bearing faults which is one of the most difficult problems in traditional MCM. Comparisons were made to other pre-processing methods, such as the short time Fourier transform and traditional wavelet transform, with an average performance improvement of 15% under the same CNN structure. However, the improvement seems to be mainly related to the enhanced pre-processing rather than to the enhanced CNN. Although the data were not from a sophisticated planetary gearbox, it was still a challenging diagnosis task because the gear mesh harmonics can dominate the measured vibration signals, which heavily mask the planet bearing fault information.

Hoang and Kang [26] analysed the CWRU bearing datasets with CNN using greyscale image input derived from raw vibration signals. They extracted bearing fault features directly from raw vibration data via a pre-processing of time–frequency transform and claimed to have achieved very high accuracy of classification (100% with a convolutional layer kernel size of 30) and robustness under noisy environments. By adding noise to the CWRU data, they compared the proposed CNN model with the SAE and 1D-CNN models, where their vibration image VI-CNN method achieved an accuracy of 97.75% versus 95.5% and 90.75% for SAE and 1D-CNN, respectively, at − 10 dB signal-to-noise ratio. However, what is unclear is how the improvement of VI-CNN over the standard 2D-CNN's (used in the ImageNet competitions) was achieved.

Similar to the approach in [26], Chen et al. [27] used another pre-processing technique (cyclic spectral coherence) to generate the image data (termed 2D CSCoh maps) as the input to their proposed CNN (with group normalization embedded in each convolutional and fully connected layer) for analysing the CWRU bearing data and another experimental dataset. They injected a group normalization process to the data to account for data distribution discrepancy between data acquired under different health conditions. They concluded that the proposed method is able to provide superior discriminative feature representations and to obtain improved classification performance (average accuracy of 99.02%) for bearing health conditions under various operations with a comparison to traditional benchmarks and other CNN architectures (averaged at 96% accuracy). They also mentioned fact that traditional MCM techniques have difficulty in detecting the ball fault in the CWRU dataset, despite the CWRU data being acquired from relatively simple machinery. We will give more comments on the CWRU dataset in Part 2.

Zhuang et al. [28] proposed an approach of applying a less common DL architecture—stacked residual dilated convolutional neural network (SRD-CNN) to real-time bearing fault classification using the CWRU data. The SRD-CNN combines the dilated convolution, the input gate structure of long short-term memory network (LSTM) and the residual network. This combination allowed the SRD-CNN model to exponentially increase the receptive field of convolution kernel and extract features from the sample with more points, alleviating the influence of randomness via the dilated convolution and to effectively remove noise and control the entry of information contained in the input sample by using the input gate structure of LSTM. Additionally, introducing the residual network made the SRD-CNN to overcome the problem of vanishing gradients caused by the deeper structure of the neural network. They claimed that the proposed SRD-CNN model has a higher denoising ability and better workload adaptability, hence is able to deliver better overall classification accuracy than three other DL models. However, The SRD-CNN model's prediction accuracy of more than 84% under various operating conditions and more than 95% prediction accuracy under different noise conditions don't appear to be as advantageous compared to some other methods reviewed here. [26, 28] are some DL + CWRU studies where only the CWRU data were used to validate the DL models.

Cipollini et al. [29] used an induction motor stator current signal as the input data to a DL framework for bearing condition monitoring. They exploited a deep learning architecture with a series of seeded bearing fault data from an inverter-fed motor and seemed to be able to deliver an effective yet simple bearing fault detection system. Interestingly, the seeded faults are obtained by drilling a hole on the outer race of the test bearings, and their test rig has a similar level of sophistication to the CWRU rig. The question is how sensitive the motor current signals are to real-world incipient fault conditions. It appears that the DL architecture used is like a SAE, but the paper did not provide a clear description of the method. Additionally, the method may be just narrowly applicable to simple cases like the induction motor bearing fault. An advantage though is that no extra sensor is required.

Precision manufacturing requires constant monitoring of tool wear conditions to ensure product quality. Shi et al. [30] presented a DL framework with fused multi-stacked sparse autoencoders (FMSSAE) for tool condition monitoring. The DL model is trained by the lower-level features in multiple parallel feature spaces, which are then fused to the higher-level features associated with tool wear conditions. They also employed a modified loss function to enable this learning structure that can enhance the feature extraction and classification. Using real-world machine tool vibration data, they demonstrated the good performance of the proposed framework with over 96% classification accuracy, which outperforms traditional machine learning methodologies such as the BPNN (86% highest) and SVM (highest of 91%). Martínez-Arellano et al. [31] applied an off-the-shelf CNN model on image input converted from raw vibration and force data acquired on a CNC (computer numerical control) machine for tool wear classification. Using experimental validation, they presented a classification accuracy above 90% in some cases, which perhaps does not compare favourably with other methods reviewed here.

3.2 More General Applications of DL in Mechanical Systems and Deep Transfer Learning

Looking at applications of DL to more general engineering fields like smart manufacturing, Lee et al. [32] evaluated DL models like LSTM and CNN with experimental time series data and time–frequency domain data from a motor test bed for MCM problems such as imbalance under variable speed operations. They focused on studying DL model invariance to changing operation conditions such as speed variations and data pre-processing like data scaling/smoothing and continuous wavelet transform (CWT). Under different combinations of pre-processing methods and DL models, their evaluation results demonstrated that an imbalance condition can be well classified when tested on data with the same RPM as its training set. The CWT-CNN combination has the best overall performance (95.5% average accuracy) when it is trained at 360 rpm in classifying imbalance conditions at various speeds in the test data. In comparison with a physics-based technique, imbalance under variable speed may not be something that DL method would have an obvious advantage as an imbalance index may be reliably defined as proportional to the speed squared.

Arellano-Espitia et al. [33] presented a DL methodology for fault diagnosis in electromechanical systems based on an unsupervised SAE and a supervised discriminant analysis. For validation and testing of the SAE model, they sourced data from multiple domains such as mechanical vibration and stator current signals of an electromechanical system consisting of two servo-motors, a gearbox and several actuators. They demonstrated that the SAE can classify five health conditions including one healthy state and four faulty states (gear and bearing faults, eccentricity fault and demagnetization fault) with good accuracy of 92.03 against seven other methods in which the highest accuracy of 88.63 was achieved by the LSTM method. They also claimed that the proposed methodology is easy to implement and highly adaptable to available data.

Li et al. [34] applied a DL approach (a five-layer BPNN with tanh activation function) to classifying mechanical faults and assessing system degradation with features extracted by the wavelet packet decomposition (WPD) of vibration data. After comprehensive comparison with some traditional machine learning (ML) techniques (SVM, deep belief network, back propagation neural network and *k*-nearest neighbour), they showed that DL enjoys superior performance in degradation assessment using rotor vibration data from the Bently Nevada Rotor Kit RK3 containing three types of simulated mechanical faults—bearing looseness, main spindle friction and load imbalance. With the same features generated by WPD, DL's performance in fault classification (99.87% accuracy) is actually comparable to the traditional ML methods with the worst accuracy at 99.77%.

If traditional ML provides such high accuracy for fault classification, would people still consider DL in practice?

As machine health conditions evolve with time, there is a common problem in applying DL approach to MCM cases where the testing data often has different distributions from the training data, e.g. training is more likely to be based on the healthy-state of the machine, whereas testing data can contain rare but possible faulty-conditions. Han et al. [35] addressed this real problem—lack of fault data for supervised training of DL models. They rightly pointed out that "*the success of supervised deep models is largely attributed to a mass of typically labelled data, while it is often limited in real diagnosis tasks. In addition, the diagnostic model trained with data from limited conditions may generalize poorly for conditions not observed during training.*" Taking on these challenges, they proposed a novel DL framework—the deep adversarial convolutional neural network (DACNN), where adversarial learning as a regularization was introduced and a discriminative classifier (two-class classifier) was added into the convolutional neural network (CNN). Two data sources were used to evaluate the DACNN method: (a) data from a direct-drive wind turbine test bench under six different wind speeds with ten health conditions—one healthy, three front bearing fault conditions, two back bearing pedestal loosening, two misalignment and two blade fault conditions; (b) PHM2009 Challenge Data—gearbox fault classification. They conducted comparison studies between other more conventional DL models and the DACNN and found that DACNN is more applicable and superior in terms of robust feature representation, boosting the generalization ability of the trained model as well as avoiding overfitting with a small size of labelled samples. The performance improvement by DACNN is significant: for dataset (a) with limited labelled samples, an 11.7% improvement under seen conditions and a 15.5% improvement under unseen conditions over the best performing traditional CNN architecture with base accuracy of 84.3% and 71.8%, respectively; for dataset (b), the seen/unseen accuracies are 96.9/88.6% versus 84.7/57.9% by LSTM that is the best performing among DL methods proposed in other studies.

Li et al. [36] addressed the data continuity problem of domain adaptation in MCM where the training and test data are acquired on different machines (or different locations of the same machine) with different data distributions. They proposed a DL method (a CNN-based domain adversarial network) with domain adaptation by introducing adversarial training for marginal domain fusion and exploring unsupervised parallel data to achieve conditional distribution alignments with respect to different machine health conditions. Using two rotating machinery datasets (CWRU bearing dataset and a rotor health dataset with nine health conditions) in frequency-domain for testing, they suggested that the proposed method has an enhanced applicability in dealing with training and test data coming from different locations within the same machines. Generally, for CWRU data over 80% accuracies can be achieved if the parallel data contain more than three classes, and for rotor health data up to 100% accuracy is achievable where the parallel data cover more than half the concerned classes.

Han et al. [37] used a transfer learning framework to handle a similar scenario where the assumption that the same distribution holds for both the training data and test data is broken. They can facilitate the diagnosis of a new yet similar fault using a pre-trained CNN on the training data distributed differently from the test data. In the framework, the architecture and weights of the pre-trained CNN are fine-tuned to achieve transfer learning. They discussed three fine-tuning strategies with a comparison to evaluate their feature transferability and were able to show by two case studies of gearbox health data that the proposed framework can effectively transfer the pre-learnt features to a new CNN to handle previously unseen or new operating and health conditions. For example, using CNN with 2 convolution layers in the PHM 2009 dataset and using a small quantity of target samples of 100 and 50, they still managed to demonstrate satisfactory diagnostic accuracy rates of 93.7% and 83.6% respectively.

Li et al. [38] proposed to use data augmentation techniques to enhance small original training datasets by artificially adding extra data samples for better training of DL models. This is because accurately labelled data (faulty-condition data in particular) are usually difficult to obtain in real applications. They investigated various augmentation techniques such as additional Gaussian noise, masking tonal noise, signal translation, amplitude shifting and time stretching. Using two popular rolling bearing datasets (CWRU dataset and a dataset from University of Cincinnati's Center of Intelligent Maintenance Systems—IMS dataset), they managed to deliver high diagnosis accuracy of up to 99.9% with only limited training data, which is favourably compared with other recent DL methods applied on the same datasets. The superior performance of the proposed method was extensively evaluated in terms of data augmentation strength and network depth. However, both datasets were acquired on unsophisticated machinery, which is the only aspect of their work that makes the proposed method less convincing.

Zhao and Jia [39] put forth an approach to tackle the MCM problem of sparsity of labelled data for supervised learning, which makes unsupervised learning more meaningful and significant. Using a sparse filtering as the feature extractor (two layers) and the weighted Euclidean affinity propagation (WEAP) as the clustering extractor (one layer), they constructed a span-new unsupervised DL network for intelligent fault diagnosis of rotating machinery. Utilizing two rolling bearing fault datasets (CWRU dataset and a dataset from the accelerated bearing life tester—ABLT-1A), they were able to demonstrate the superiority of the proposed algorithms by comparing to other clustering methods where the proposed method can reach a recognition rate of 100% from the training/testing sample ratio of 40/60, and better performance at all other sample ratios.

Garcia et al. [40] attended to the fact that there are very limited application cases of unsupervised DL-based (AE & CNN) anomaly detection. They firstly evaluated six image encoding strategies such as Gramian angular field, Markov transition field, recurrence plot, grey scale encoding, spectrogram and scalogram to transform the raw time series data into images for a convolutional auto-encoder (CAE). Then, they defined a more robust encoding method by modifying each of these six existing algorithms. Training the DL model only on healthy condition data, they extracted the 99$^{\text{th}}$%ile in the distribution of the residuals of all sub-series to define the detection threshold τ and then monitor the maximum residual over the sub-series (for the detection of local anomalies) and compared it to the threshold τ beyond which an anomaly is detected. Using the real-world flight test vibration data—Airbus Helicopters Accelerometer Dataset (https://doi.org/10.3929/ethz-b-000415151)—a sensor fault dataset, they conducted a comprehensive study comparing the modified and the existing encoding methods and showed an improved performance by using the encoded images against using the raw time series. All the modified versions were observed to perform better than their un-modified counterparts, in which the scalogram indicated the best performance with an F1 score of 0.91 and AUC (area under the curve) score of 0.92. However, the use of DL methods is not sufficiently justified against traditional techniques, which we will discuss more in Part 2.

3.3 Reinforcement Learning Applications to CM Decision-Making

Reinforcement learning (RL) is gathering momentum in various MCM and machine health management applications, and the MCM decision-making process in particular. Xanthopoulos et al. [41] investigated an RL-based approach to obtaining optimal or near-optimal joint production/maintenance control policies. They addressed the problem of finding the optimal trade-off between maintaining a high service level and carrying as low inventory as possible or finding the right balance between condition-based maintenance and periodic maintenance. The optimization objective of the reward scheme is to minimise the total costs associated with the inventory and backorder. Using simulation experiments, they compared the proposed approach with the traditional parametric policies and found that the proposed approach clearly outperformed other maintenance control policies where the best is simulation case-3 with at least 3 times the cost savings under the RL approach.

Bellani et al. [42] addressed the core problem of PHM where health condition predictions drive the decisions on the equipment operation and maintenance (O&M). They proposed an approach to optimal sequential O&M decision-making based on sequential decision problem, artificial neural networks and RL. They applied the framework to a scaled-down case study concerning a real mechanical equipment equipped with PHM capabilities with a comparison of the proposed framework against some traditional methods. They applied the framework to a simulated operating environment of a pumping system at various loads with a shaft crack with the crack length being the state. The reward is operating profit, and the actions are to operate the pump at particular load level or to do maintenance in order to achieve a maximum operating profit. The simulation result showed that some RL-driven actions are quite counter-intuitive, which might make it harder to implement such strategies in practical engineering applications. Yousefi et al. [43] proposed a dynamic condition-based maintenance model for multi-component systems with repairable components using an RL method that is more time-efficient and cost-effective compared to the traditional maintenance optimization solutions. In this study, the maintenance problems are formulated as a Markov decision process and are solved by using a Q-learning algorithm and deep Q-learning with the goal of providing more practical and effective maintenance models to avoid failures while minimizing the maintenance cost.

Usually, it is difficult for traditional methods to automatically build appropriate models for different datasets. Wang et al. [44] proposed a RL-based method to automatic search for neural network architectures for fault diagnosis of rolling element bearings. The RL architecture contains a controller (or an agent) model and child models, where the controller is an RNN for generating a series of actions, each of which specifies a design choice to construct the child models (CNNs) for fault diagnosis. The reward is the accuracy of the child CNN models, which is to be maximized. Since the reward is not differentiable, they updated the parameters with the policy gradient method. By applying the method to several bearing datasets (CWRU dataset and a bigger dataset from locomotive bearings), they demonstrated that the proposed method can achieve an automatic design of neural network architecture to be able to handle different datasets. For the CWRU data, the optimized CNN model delivered a better performance with an average accuracy of 98.47% and a standard deviation of 0.61% against non-optimized CNN's with highest accuracy of 92.6% with a standard deviation of 3.93%. For the locomotive bearing data, the proposed method achieved an accuracy of 97.63 ± 1.36% versus the highest accuracy of 83.53 ± 4.17% by non-optimized CNNs. They also discussed some shortcomings of the proposed method, such as far too many candidate child models to choose from and the local optima problems.

Ding et al. [45] put forward a deep reinforcement learning method to build an end-to-end fault diagnosis neural network (NN) architecture to allow direct mapping of raw fault data to the corresponding fault modes. They employed deep RL to optimize the NN architecture and parameters with a reward mechanism that awards one point for the correct answer and minus one point for a wrong answer to the candidate NN and used SAE to learn fault features from vibration signals. Using two types of rotating machinery datasets (CWRU bearing data and a hydraulic pump dataset), they validated the proposed method with promising results in establishing a general fault diagnosis architecture for rotating machinery. The results from both datasets show that the proposed RL-based method (unsupervised) can achieve comparable accuracy to the supervised SAE-softmax method. Hence, they claimed that the proposed method can train a DL-based agent to diagnose the bearing faults solely on the raw vibration signals, which might have varying health and operating conditions. Because the learning process only depends on the replayed memories of the agent and the overall rewards, they claimed that the proposed method represents much weaker feedback than that obtained by the traditional supervised DL method, which facilitated the promising results obtained in the paper.

3.4 General Impressions of DL Applications in MCM

Now we attempt to address the question whether a major breakthrough is achieved in DL applications to MCM. We can assign a breakthrough index (BTI), e.g. a fractional number between 0 and 1, to the papers reviewed above to judge the performance of applying DL methodologies to MCM problems. Our intent is to give some kind of quantitative measure to a subjective qualitative judgement, which may help form a clearer view of ours on the current status. A BTI of 1 means the paper brings a true breakthrough with at least 20% improvement from non-DL-based traditional MCM methods; 0.75 is for 10% improvement; 0.5 is for no real improvement over or on par with the non-DL methods; 0.25 means the DL method is not as good as the non-DL methods but can deliver some level of detection to very simple MCM problems with complex DL solution; and a BTI of 0 is for no real detection to MCM problems commonly solvable with non-DL methods. The BTI = 1 for 20% improvement is based on the major breakthrough of applying CNN (or *AlexNet*) to image classification achieved in 2012 by Hinton's group from University of Toronto [46]. In this study, they managed to bring down the top-5 classification error rate from the previous benchmark of 25.7% to 15.3%, which is about 40% improvement. We could define half of that or 20% improvement as a major breakthrough to MCM problems for the purpose of this critical review paper.

We thought that it is probably not a good idea to give each paper a BTI score, and instead, we would prefer to give an overall judgement with further comments to a few really good ones in our view. Of the 34 selected papers reviewed above, our general impression is that the majority of the papers reviewed here can be scored at BTI = 0.5 ~ 0.65, while a high BTI score of above 0.8 can be given to [19, 35, 38], which claimed to have achieved more than 10% improvement in performance.

In [19], the application dealt with a practical vibration diagnosis problem of aircraft engine accessory gearboxes that often challenges the traditional MCM techniques due to their complexity. Such a gearbox would have many rotating components with tens of gears and bearings plus several pumps connected to a few shafts. The vibration signals are mostly dominated by multiple sets of gear mesh harmonics and their intermodulation contents. The more than 25% performance boost by the Bi-LSTM method compared to traditional methods qualifies the paper to be scored highly. One might argue that it only differentiates bad from good (or healthy), but often that is the most difficult issue at the front-line screening for aircraft engine systems. Operators would want to conduct maintenance activities as soon as possible once they find the gearbox is no longer behaving normally—there is very little margin for prognosis in aircraft critical systems due to safety concerns. However, if further diagnosis is needed, traditional MCM signal processing techniques may be employed, such as the synchronous signal averaging for isolating gear and pump problems and cyclo-stationary analysis for identifying bearing fault characteristics. In turn, the diagnosis results can feed into a database with known fault types for further enhancing the DL-based methods. We would strongly advocate more DL-based research effort into this kind of MCM problem—anomaly detection of complex systems with very limited (if any) labelled data and large amounts of healthy-state data, and there is limited diagnosis capability by traditional MCM techniques (either physics-based or signal processing based).

In [35], the focus was on a common challenge for supervised training of DL models—the lack of labelled fault data for MCM problems. For a new piece of machinery (x) under monitoring, the zero fault-history means DL models can only start to learn the normal behaviours of the machine except for transferring DL models learned from other machines (z) with fault-history. In the generative adversarial network (GAN) framework, the generative model G was previously trained on data from z, as real data from x comes in, G will try to represent the distribution of x and generate the fake samples $\hat{x}$, and the discriminative model D (like a two-class classifier) takes in both x and $\hat{x}$ to judge if $\hat{x}$ is close to or far from real. In every iteration, model G makes a better representation of the distribution of the real data x, and $\hat{x}$ becomes less fake (or more real). Hence the two DL-based models of G and D (both are CNNs in [35]) are trained in parallel in an adversarial style to gain transfer learning. Han et al. [35] deserves a high BTI score because of DACNN's significant performance improvement of 11.7% under seen conditions and 15.5% under unseen conditions over the best performing traditional CNN architecture. Due to the scarcity of labelled data in MCM, transfer learning will be a key characteristic for DL-based methods if they are to be applied effectively in real-world MCM fields.

Instead of using transfer learning, [38] tackled the problem of labelled data scarcity by exploring data augmentation techniques to enhance the training of DL models. With various sample-based augmentation techniques over the limited CWRU bearing datasets, the authors found that the performance improvement from the data translation augmentation seems to be significant (from a baseline 55% accuracy to 95% for fault type classification) while the other four augmentation techniques can suffer from the overfitting problem. Further extending the training samples by augmentation to 400 resulted in an accuracy as high as 99.91%. Using dataset-based augmentation over the IMS bearing constant load run-to-failure test data from the University of Cincinnati, they managed to achieve high accuracy of 98% and 94% in classifying the fault severity with signal translation and time stretching augmentation techniques, respectively, which was more than 30% performance improvement from the baseline. The classification accuracy can be further improved to above 99% by optimizing the hyper parameters. While a challenging problem of planet bearing diagnosis is addressed in [25], an innovative way of training DL model using data augmentation is proposed in [38]. We believe that a major breakthrough with a BTI of 1 may be achievable if the approach in [38] can be successfully applied to a more challenging problem presented in [25], preferably to the planet bearing diagnosis for complex helicopter gearboxes (e.g. the Super Puma main gearbox) where traditional MCM methodologies are inefficient. It would be of huge benefit to the MCM and DL communities if helicopter manufacturers could design some datasets for a data challenge type of competition, similar to the famous ImageNet competition.

Therefore, our assessment is that a major breakthrough in applying DL to MCM problems is not quite there yet but it could be just around the corner. Major breakthroughs will happen soon provided that data from complex machinery becomes benchmarks for researchers to test a large reservoir of DL methodologies, and the fault types in the data are extremely difficult for traditional MCM techniques to handle. In other words, we believe that a major breakthrough is dependent more on data than on DL algorithms.

4 Concluding Remarks

Based on the above review, we observe that the DL application to MCM has gathered an accelerating momentum with varying degrees of success. The greatest success lies in those applicability investigations where multiple DL architectures can be successfully applied to various MCM problems such as fault detection, fault diagnosis and classification, and RUL predictions. In many cases, DL-based methods with automatic feature generation capability can outperform traditional ML methods with features manually generated by traditional statistical signal processing techniques. In a few cases, the margin of outperformance is approaching the prescribed breakthrough level (i.e. 20% or more) when applied to data from unsophisticated machinery with the exception of [19]. However, we believe a major breakthrough is still yet to come largely due to the lack of labelled fault data and lack of data from complex state-of-the-art machinery. We are hopeful that DL is highly capable of solving real-world MCM problems, and with more data of high fidelity from complex systems made available by industry stakeholders and more DL-based research efforts, major breakthroughs could be just around the corner. To put it in a quantitative way, we may arguably conclude is that the DL-based MCM is about 80% on its way to a major breakthrough.

From DL algorithm point of view, possible focuses for future research can be in (a) RL-based DL algorithm optimization using high-performance computing (HPC) and lots of data; (b) anomaly detection with adaptive capacity to classify faults progressively—adaptive transfer learning, where GAN is perhaps one of the most promising techniques [35]. From data point of view, it would be ideal for future benchmark datasets to come from complex real-world machinery where traditional MCM techniques lack of capability. Laboratory test data from simplistic test rigs would probably be insufficient to demonstrate the full potential of DL algorithms for MCM purposes. Out of the above review, we have made some more observations and can give further critical views that may help facilitate future research efforts into DL-based MCM. However, due to the page limit we can only present them with a case study in Part 2 of the paper.

Declarations

Ethical Approval The work is original and conducted by authors under the employment of the Commonwealth Government of Australia. The authors have no external financial support and no conflict of interest to declare. There is no confidential or proprietary information in any material discussed in this article.

References

1. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.: Deep learning and its applications to machine health monitoring. Mech. Syst. Signal Process. **115**, 213–237 (2019). https://doi.org/10.1016/j.ymssp.2018.05.050
2. Zhang, S., Zhang, S., Wang, B., Habetler, T.G.: Deep learning algorithms for bearing fault diagnostics: a comprehensive review. IEEE Access **8**, 29857–29881 (2020). https://doi.org/10.1109/ACCESS.2020.2972859
3. Hoang, D.T., Kang, H.J.: A survey on Deep Learning based bearing fault diagnosis. Neurocomputing **335**, 327–335 (2019). https://doi.org/10.1016/j.neucom.2018.06.078
4. Neupane, D., Seok, J.: Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: a review. IEEE Access **8**, 93155–93178 (2020). https://doi.org/10.1109/ACCESS.2020.2990528
5. Waziralilah, N.F., Abu, A., Lim, M.H., Quen, L.K., Elfakharany, A.: A review on convolutional neural network in bearing fault diagnosis. MATEC Web Conf. **255**, 06002 (2019). https://doi.org/10.1051/matecconf/201925506002
6. Serin, G., Sener, B., Ozbayoglu, A.M., Unver, H.O.: Review of tool condition monitoring in machining and opportunities for deep learning. Int. J. Adv. Manuf. Technol. **109**, 953–974 (2020). https://doi.org/10.1007/s00170-020-05449-w
7. Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.J., Ducoffe, M.: Potential, challenges and future directions for deep learning in prognostics and health management applications. Eng. Appl. Artif. Intell. **92**, 103678 (2020). https://doi.org/10.1016/j.engappai.2020.103678
8. Tang, S., Yuan, S., Zhu, Y.: Deep learning-based intelligent fault diagnosis methods toward rotating machinery. IEEE Access **8**, 9335–9346 (2020). https://doi.org/10.1109/ACCESS.2019.2963092
9. Rezaeianjouybari, B., Shang, Y.: Deep learning for prognostics and health management: state of the art, challenges, and opportunities. Measurement **163**, 107929 (2020). https://doi.org/10.1016/j.measurement.2020.107929
10. Zhao Z., Li T., Wu J., Sun C., Wang S., Yan R, Chen X.: Deep learning algorithms for rotating machinery intelligent diagnosis: an open source benchmark study. arXiv preprint (2020). https://arxiv.org/abs/2003.03315
11. Tang, S., Yuan, S., Zhu, Y.: Data preprocessing techniques in convolutional neural network based on fault diagnosis towards rotating machinery. IEEE Access **8**, 149487–149496 (2020). https://doi.org/10.1109/ACCESS.2020.3012182
12. He, M., He, D., Yoon, J., Nostrand, T.J., Zhu, J., Bechhoefer, E.: Wind turbine planetary gearbox feature extraction and fault diagnosis using a deep-learning-based approach. Proc. Inst. Mech. Eng. Part O J. Risk Reliab. **233**(3), 303–316 (2019). https://doi.org/10.1177/1748006X18768701
13. Fu, J., Chu, J., Guo, P., Chen, Z.: Condition monitoring of wind turbine gearbox bearing based on deep learning model. IEEE Access **7**, 57078–57087 (2019). https://doi.org/10.1109/ACCESS.2019.2912621
14. Helbing, G., Ritter, M.: Deep learning for fault detection in wind turbines. Renew. Sustain. Energy Rev. **98**, 189–198 (2018). https://doi.org/10.1016/j.rser.2018.09.012
15. Zhao, H., Liu, H., Hu, W., Yan, X.: Anomaly detection and fault analysis of wind turbine components based on deep learning network. Renew. Energy **127**, 825–834 (2018). https://doi.org/10.1016/j.renene.2018.05.024
16. Resendiz-Ochoa, E., Saucedo-Dorantes, J.J., Benitez-Rangel, J.P., Osornio-Rios, R.A., Morales-Hernandez, L.A.: Novel methodology for condition monitoring of gear wear using supervised learning and infrared thermography. Appl. Sci. **10**, 506 (2020). https://doi.org/10.3390/app10020506
17. Wang W., Galati F.A., Szibbo D.: LSTM residual signal for gear tooth crack diagnosis. Advances in Asset Management and Condition Monitoring, Smart Innovation, Systems and Technologies, vol. 166. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57745-2_89
18. Singh, J., Azamfar, M., Ainapure, A., Lee, J.: Deep learning-based cross-domain adaptation for gearbox fault diagnosis under variable speed conditions. Meas. Sci. Technol. **31**, 5 (2020). https://doi.org/10.1088/1361-6501/ab64aa
19. Mallikarjuna, P.B., Sreenatha, M., Manjunath, S., Kundur, N.: Aircraft gearbox fault diagnosis system: an approach based on deep learning techniques. J. Intell. Syst. **30**(1), 258–272 (2021). https://doi.org/10.1515/jisys-2019-0237
20. Li, X., Zhang, W., Ding, Q.: Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliab. Eng. Syst. Saf. **182**, 208–218 (2019). https://doi.org/10.1016/j.ress.2018.11.011
21. Galati F.A., Wang W., Bielenberg B.: Gear-bearing fault detection based on deep learning. In: Proceedings of the 11th DST international conference on Health and Usage Monitoring Systems (HUMS2019), 24–28 Feb 2019, Melbourne, Australia. http://www.humsconference.com.au/Papers2019/Peer_Reviewed/HUMS2019_Wang.pdf
22. Chen, Y., Peng, G., Zhu, Z., Li, S.: A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. Appl. Soft Comput. **86**, 105919 (2020). https://doi.org/10.1016/j.asoc.2019.105919
23. Udmale, S.S., Singh, S.K., Bhirud, S.G.: A bearing data analysis based on kurtogram and deep learning sequence models. Measurement **145**, 665–677 (2019). https://doi.org/10.1016/j.measurement.2019.05.039
24. Xu, G., Liu, M., Jiang, Z., Söffker, D., Shen, W.: Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning. Sensors **19**(5), 1088 (2019). https://doi.org/10.3390/s19051088
25. Zhao, D., Wang, T., Chu, F.: Deep convolutional neural network based planet bearing fault classification. Comput. Ind. **107**, 59–66 (2019). https://doi.org/10.1016/j.compind.2019.02.001
26. Hoang, D.T., Kang, H.J.: Rolling element bearing fault diagnosis using convolutional neural network and vibration image. Cogn. Syst. Res. **53**, 42–50 (2019). https://doi.org/10.1016/j.cogsys.2018.03.002
27. Chen, Z., Mauricio, A., Li, W.H., Gryllias, K.: A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks. Mech. Syst. Signal Process. **140**, 106683 (2020). https://doi.org/10.1016/j.ymssp.2020.106683
28. Zhuang, Z., Lv, H., Xu, J., Huang, Z., Qin, W.: A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions. Appl. Sci. **9**, 1823 (2019). https://doi.org/10.3390/app9091823
29. Cipollini, F., Oneto, L., Coraddu, A., Savio, S.: Unsupervised deep learning for induction motor bearings monitoring. Data Enabled Discov. Appl. **3**, 1 (2019). https://doi.org/10.1007/s41688-018-0025-2
30. Shi, C., Panoutsos, G., Luo, B., Liu, H., Li, B., Lin, X.: Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing. IEEE Trans. Ind. Electron. **66**(5), 3794–3803 (2019). https://doi.org/10.1109/TIE.2018.2856193
31. Martínez-Arellano, G., Terrazas, G., Ratchev, S.: Tool wear classification using time series imaging and deep learning. Int. J. Adv. Manuf. Technol. **104**, 3647–3662 (2019). https://doi.org/10.1007/s00170-019-04090-6

32. Lee, W.J., Xia, K., Denton, N.L., Ribeiro, B., Sutherland, J.W.: Development of a speed invariant deep learning model with application to condition monitoring of rotating machinery. J. Intell. Manuf. (2020). https://doi.org/10.1007/s10845-020-01578-x
33. Arellano-Espitia, F., Delgado-Prieto, M., Martinez-Viol, V., Saucedo-Dorantes, J.J., Osornio-Rios, R.A.: Deep-learning-based methodology for fault diagnosis in electromechanical systems. Sensors **20**, 3949 (2020). https://doi.org/10.3390/s20143949
34. Li, Z., Wang, Y., Wang, K.: A deep learning driven method for fault classification and degradation assessment in mechanical equipment. Comput. Ind. **104**, 1–10 (2019). https://doi.org/10.1016/j.compind.2018.07.002
35. Han, T., Liu, C., Yang, W., Jiang, D.: A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. Knowl. Based Syst. **165**, 474–487 (2019). https://doi.org/10.1016/j.knosys.2018.12.019
36. Li, X., Zhang, W., Xu, N., Ding, Q.: Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places. IEEE Trans. Ind. Electron. **67**(8), 6785–6794 (2020). https://doi.org/10.1109/TIE.2019.2935987
37. Han, T., Liu, C., Yang, W.G., Jiang, D.X.: Learning transferable features in deep convolutional neural networks for diagnosing unseen machine conditions. ISA Trans. **93**, 341–353 (2019). https://doi.org/10.1016/j.isatra.2019.03.017
38. Li, X., Zhang, W., Ding, Q., Sun, J.Q.: Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. J. Intell. Manuf. **31**, 433–452 (2020). https://doi.org/10.1007/s10845-018-1456-1
39. Zhao, X., Jia, M.: A novel unsupervised deep learning network for intelligent fault diagnosis of rotating machinery. Struct. Health Monit. **19**(6), 1745–1763 (2020). https://doi.org/10.1177/1475921719897317
40. Garcia G.R., Michau G., Ducoffe M., Gupta J.S., Fink O.: Time series to images: monitoring the condition of industrial assets with deep learning image processing algorithms. arXiv preprint (2020). https://arxiv.org/abs/2005.07031v2
41. Xanthopoulos, A.S., Kiatipis, A., Koulouriotis, D.E., Stieger, S.: Reinforcement learning-based and parametric production-maintenance control policies for a deteriorating manufacturing system. IEEE Access **6**, 576–588 (2018). https://doi.org/10.1109/ACCESS.2017.2771827
42. Bellani L., Compare M., Baraldi P., Zio E.: Towards developing a novel framework for practical PHM: a sequential decision problem solved by reinforcement learning and artificial neural networks. Int. J. Progn. Health Manag. 10, 031 (2019). https://www.phmsociety.org/node/2656
43. Yousefi, N., Tsianikas, S., Coit, D.W.: Reinforcement learning for dynamic condition-based maintenance of a system with individually repairable components. Qual. Eng. **32**(3), 388–408 (2020). https://doi.org/10.1080/08982112.2020.1766692
44. Wang, R., Jiang, H., Li, X., Liu, S.: A reinforcement neural architecture search method for rolling bearing fault diagnosis. Measurement **154**, 107417 (2020). https://doi.org/10.1016/j.measurement.2019.107417
45. Ding, Y., Ma, L., Ma, J., Suo, M., Tao, L., Cheng, Y., Lu, C.: Intelligent fault diagnosis for rotating machinery using deep Q-network based health state classification: a deep reinforcement learning approach. Adv. Eng. Inform. **42**, 100977 (2019). https://doi.org/10.1016/j.aei.2019.100977
46. Krizhevsky A., Sutskever I., Hinton G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of conference on advances in neural information processing systems, 1097–1105 (2012)