A Review of Artificial Intelligence Algorithms Used for Smart Machine Tools

Chih-Wen Chang * iD, Hau-Wei Lee and Chein-Hung Liu *

Department of Mechanical Engineering, National Chung Hsing University, Taichung 40227, Taiwan; orsino.lee@gmail.com

* Correspondence: chihwen.chang8@gmail.com (C.-W.C.); carus@dragon.nchu.edu.tw (C.-H.L.);
  Tel.: +886-4-2284-0433 (ext. 411) (C.-W.C.)

Abstract: This paper offers a review of the artificial intelligence (AI) algorithms and applications presently being used for smart machine tools. These AI methods can be classified as learning algorithms (deep, meta-, unsupervised, supervised, and reinforcement learning) for diagnosis and detection of faults in mechanical components and AI technique applications in smart machine tools including intelligent manufacturing, cyber-physical systems, mechanical components prognosis, and smart sensors. A diagram of the architecture of AI schemes used for smart machine tools has been included. The respective strengths and weaknesses of the methods, as well as the challenges and future trends in AI schemes, are discussed. In the future, we will propose several AI approaches to tackle mechanical components as well as addressing different AI algorithms to deal with smart machine tools and the acquisition of accurate results.

Keywords: artificial intelligence; smart machine tools; learning algorithms; intelligent manufacturing; fault diagnosis and prognosis

---

1. Brief Introduction

We believe that a new epoch of the "Industrial Internet of Things (IIoT) plus artificial intelligence (AI)", characterized by big machinery data, data-driven techniques, ubiquitous networks, mass innovation, automatic intelligence, cross-border integration, and shared services, has arrived [1–3]. The fast development and combination of new AI and energy technologies, for materials, bioscience, the Internet, and new-generation information exchange, is a fundamental part of this new epoch. This will, in turn, permit game-changing transformation of models, ecosystems, and means in the light of their application to national security, well-being, and the economy. The main objective is a review and summary of recent achievement in data-based techniques, especially for complicated industrial applications, offering reference for further study from both an academic and practical point of view. Yin et al. [1] describes a brief evolutionary overview of data-based techniques over the last two decades. Recent development of modern industrial applications is presented mainly from the perspectives of monitoring and control. Their methodology, based on process measurements and model–data integrated techniques, will be introduced in the next study. Jeschke et al. [2] developed the core system science needed to enable the development of complex IIoT/manufacturing cyber-physical systems (CPS). Moreover, readers can learn the current state of IIoT and the concept of cybermanufacturing from this book. In 2014, Lund et al. [3] described the central issues contributing to, and characterizing, the worldwide and regional growth of the IoT. Besides, researchers can utilize the trend analysis of IoT their region markets in the future.

There are many AI algorithms for machine health monitoring and other machine tool applications: The second-order recurrent neural networks (RNN) method for the learning and extraction of finite

state automata [4], the continuous time RNN approach to dynamical systems [5], the RNN scheme for long short-term memory (LSTM) [6,7], the echo state network (ESN) approach to RNN training [8], the RNN algorithm for the learning of precise timing [9], the RNN encoder-decoder for learning phrase representations [10], the gated RNN method for sequence modeling [11], an overview of deep learning (DL) methods [12], the RNN scheme for machine health monitoring [13], machine health monitoring using convolutional bi-directional LSTM networks [14], the convolutional neural networks (CNN) method for handwritten digit recognition [15], the gradient-based learning approach to document recognition [16], the CNN scheme for object recognition [17], the CNN algorithm for large-scale hierarchical image databases [18], the CNN method for house number digit classification [19], the deep CNN approach to Imagenet classification [20], a CNN scheme for a hybrid nn-hmm model for speech recognition [21], the CNN approach to sentence classification [22], a deep residual learning algorithm for image recognition [23], the DL method for imbalanced multimedia data classification [24], a region-based CNN scheme for real-time object detection [25], a deep CNN based regression scheme for the estimation of remaining useful life [26], a deep CNN approach to automated feature extraction in industrial inspection [27], a CNN method for imbalanced classification [28], a deep neural networks (DNN) algorithm for natural language processing [29], a t-stochastic neighbor embedding method (t-SNE) for the visualization of high-dimensional data [30], a DNN method for acoustic modeling in speech recognition [31], a DNN approach to deep visualization [32], a DNN scheme for fault diagnosis [33], a restricted Boltzmann machine (RBM) method for failure diagnosis [34], an RBM approach to the regularization of prognosis and health assessment [35], an RBM scheme for the estimation of remaining useful life [36], a fast learning algorithm for deep belief nets [37], a deep multi-layer NNs for deep architecture [38], a deep Boltzmann machine (DBM) method for three-dimensional (3-D) object recognition [39], the introduction of a sparse auto-encoder learning algorithm [40], a review and new perspectives of representative learning methods [41], the introduction of extreme learning machine (ELM) methods [42], a deep auto-encoder (AE) approach to anomaly detection and fault disambiguation in large flight data [43], fault diagnosis using a denoising stacked auto-encoder [44], a continuous sparse auto-encoder (CSAE) approach to transformer fault diagnosis [45], a survey of transfer learning methods [46], the DL approach to tissue-regulated splicing code [47], the methods and applications of DL algorithms [48], the introduction of DL methods [49], a survey of the application of DL to machine health monitoring [50], an overview of DL approaches [51], an introduction to the learning of multiple layers of representation [52], a denoising AE for the extraction and composition of robust features [53], a large-scale deep unsupervised learning (UL) scheme for graphics processors and the building of high-level features [54,55], unsupervised learning of video representations using LSTM [56], the introduction of a constructive meta-learning (ML) method [57], an ML approach to automatic kernel selection [58], the ML method and search technique used to select parameters [59], an ML approach to the Bayesian optimization of hyperparameters [60], a clustering algorithm for new distance-based problems [61], Taxonomy and empirical analysis in a supervised learning (SL) scheme [62], a weakly SL algorithm and high-level feature learning for object detection in remotely sensed optical images [63], the introduction of an off-policy reinforcement learning (RL) method [64], a deep RL method for the augmentation of these models to exploit game feature information [65], a deep RL algorithm for robots to be learned directly from camera inputs in the real world [66], and the introduction of Q-learning approach [67]. Besides, our motivation is to organize and analyze those literature and find out the future research of smart machine tools. Detailed descriptions of these methods follow in Section 2.

　　Many studies have been made on the diagnosis and detection of mechanical components over the last few decades: a neural network (NN) algorithm for motor rolling bearing fault diagnosis [68], the artificial neural network (ANN) method for rolling element bearing fault diagnosis [69], the Pca-based feature selection scheme for machine defect classification [70], the support vector machine (SVM) approach to machine condition monitoring and fault diagnosis [71], the NN method for induction machine condition monitoring [72], the ANN approach and SVM scheme for fault detection

in ball-bearings [73], the stacked auto-encoder (SAE) and softmax regression approach for bearing fault diagnosis [74], the wavelet transform and SAE method for roller bearing fault diagnosis [75], the DNN scheme for rolling bearing fault diagnosis [76], a transfer learning-based approach for bearing fault diagnosis [77], the DNN for fault characteristic mining and the intelligent diagnosis of rotating machinery with massive data [78], a rapid Fourier transform (STFT)-deep learning scheme for rolling bearing fault diagnosis [79], a new bearing condition recognition method based on multi-feature extraction and DNN for intelligent bearing condition monitoring [80], an SVM and ANN approach to bearing health [81], the Weibull distribution and deep belief network (DBN) method for the assessment of bearing degradation [82], the multivibration signals and DBN scheme for bearing fault diagnosis [83], a hierarchical diagnosis network (HDN) approach for fault pattern recognition in rolling element bearings [84], an unsupervised feature extraction scheme for the diagnosis of journal bearing rotor systems [85], a CNN method for fault detection in rotating machinery [86], a hierarchical adaptive deep CNN approach to bearing fault diagnosis [87], a stacked denoising auto-encoder (SDA) method for fault diagnosis in rotary machine components [88], a sparse auto-encoder and DBN scheme for bearing fault diagnosis [89], a wavelet packet energy (WPE) image and deep convolutional network (ConvNet) for spindle bearing fault diagnosis [90], an extreme learning machine for online sequential prediction of bearing imbalance fault diagnosis [91], an ANN scheme for intelligent condition monitoring of gearboxes [92], an intelligent fault diagnosis and prognosis approach to rotating machinery that integrates wavelet transformation and principal component analysis [93], a multimodal deep support vector classification (MDSVC) approach to gearbox fault diagnosis [94], a multi-layer NN scheme for gearbox fault diagnosis [95], a CNN method for gearbox fault identification [96], a model for deep statistical feature learning from vibration measurements of rotating machines for fault diagnosis [97], a deep random forest fusion (DRFF) technique for gearbox fault diagnosis [98], a transfer component analysis (TCA) for gearbox fault diagnosis [99], a DBN method for structural health diagnosis [100], a DBN algorithm based state classification for engineering health diagnosis applications [101], a DL method for signal recognition and the diagnosis of spacecraft [102], long short-term memory neural network (LSTMNN) for fault diagnosis and estimation of the remaining useful life of aero engines [103], a physics-based approach to the diagnosis and prognosis of cracked rotor shafts [104], a multi-scale CNN scheme for rotor systems [105], a new support vector data description method for machinery fault diagnosis [106], a sparse auto-encoder-based DNN approach to induction motor fault classification [107], a DBN-based approach to induction motor fault diagnosis [108], a CNN method for real-time motor fault detection [109], an ANN and fuzzy logic (FL) scheme for the intelligent diagnosis of turbine blade faults [110], an analog-circuit fault diagnostic system based on the back propagation of neural networks using wavelet decomposition, principal component analysis, and data normalization [111], a logistic regression based prognostic method for the on-line assessment and classification of degradation and failure modes [112], naïve Bayes and Bayes net classifier algorithms for fault diagnosis in monoblock centrifugal pumps [113], sparse auto-encoders for the monitoring of rotating machines [114], an SAE for fault diagnosis in hydraulic pumps [115], a probabilistic kernel factor analysis (PKFA) method for the prediction of tool condition [116], a DNN approach to the diagnosis of tidal turbine vibration data [117], diagnosis methods based on deep learning for the early detection of small faults in front-end controlled wind generators (FSCWG) [118], and the CNN approach to real-time vibration-based structural damage detection [119]. Details of these methods are given in Section 3.

There are many AI technique applications for smart machine tools: the ANN scheme for intelligent manufacturing [120], an integrated AI computer-aided process planning system [121], the application of AI to manufacturing systems [122], an NN analyzer for the mechanical properties of rolled steel bar [123], integrated artificial intelligent techniques for shape rolling sequences [124], the DBN method to the cutting state monitoring [125], multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing [126], survey control for intelligent manufacturing [127], a review of AI in intelligent manufacturing [128], an ANN scheme and fuzzy modeling system

for intelligent tool wear estimation [129], optical image scattering and hybrid artificial intelligence techniques for intelligent tool wear identification [130], a pattern recognition method for identifying the designed strength of concrete by evidence accumulation [131], adaptive neuro-fuzzy inference and hybrid systems, in an ANN approach, to the modeling and optimization of ultrasonic welding (USW) process parameters [132], an NN method for the prediction of cutting forces [133], an ANN and regression analysis scheme for the modeling and prediction of cutting forces [134], an ANN method for the intelligent control of a three-joint robotic manipulator [135], an ANN approach using the genetic algorithm for intelligent fixture design [136], a survey of CPS architecture used in Industry 4.0-based manufacturing systems [137], a review of CPS in advanced manufacturing [138], hierarchical hybrid models for face-detection and face-recognition [139], machine learning for the design and evaluation of obstacle detection in a transportation CPS [140], machine learning algorithms for CPS, decision sciences and data products [141], an ANN scheme for diagnosis, classification and prognosis of rotating machines [142], an ANN method for the prediction of the remaining useful life in rotating machines [143], wavelet packet decomposition, Fourier transform and ANN methods used in the classification of faults and the prediction of the degradation of components and machines in a manufacturing system [144], a recursive least-squares (RLS) approach to the behavior of rolling element bearings [145], an NN method for the prognosis of remaining bearing life [146], a model-based, data-driven (or both), prediction method for remaining useful life (RUL) [147], condition-based machinery maintenance (CBM), diagnostics and prognostics [148], a dynamic fault isolation scheme for the prognosis of hybrid systems [149], an LSTM based encoder-decoder for multi-sensor prognostics [150], an NN method for intelligent pressure sensors [151], a functional link ANN approach for the modeling of an intelligent pressure sensor [152], a rough set NN method for intelligent pressure sensors [153], an ANN scheme for the temperature compensation of Piezo-resistive micro-machined porous silicon pressure sensors [154], an ANN-based technique for smart sensors operating in harsh environments [155], an ELM method calibration algorithm for pressure drift [156], an optical sensor for the real-time measurement of axial thermal elongation in a high-speed machine tool spindle [157], optical sensors for on-line real-time measurement of straightness and angular errors in a linear stage [158], 3D optical sensors used to measure rotational errors in a stage or spindle based on deviations of the optical axis [159], an optimization technique and ANN model for hot-wire thermal flow sensors [160], an ANN method for self-calibration and determination of optimal response of intelligent sensor designs [161], a distributed system for the analysis and monitoring (DSAM) of vibrations and acoustic noise [162], and a distributed smart plug sensor network in an intelligent power monitoring and analysis system [163]. Details of these methods are given in Section 4.

The remaining parts of this article are as follows: Section 2 interprets the AI algorithms involved in machine health monitoring and machine tool inverse heat conduction problems. Section 3 is about the diagnosis and detection of mechanical components. AI technique applications in smart machine tools are explained in Section 4. In Section 5, some conclusions about the current challenges and future work are made. In addition, we have provided an explanatory tree diagram of this study (see Figure 1), and the organized literature is listed in Table 1.

Table 1. Many previous reports of AI approaches for smart machine tools.

Various Learning Algorithms (67 Literature)	Diagnosis and Detection of Mechanical Components (52 Literature)	AI Technique Applications in Smart Machine Tools (44 Literature)
DL (51): RNN [4–14], CNN [15–28], DNN [29–33], RBM [34–36], others [1–3,37–51]	Bearings (24): [68–91]	Intelligent Manufacturing (22): [120–141]
UL (5): [52–56]	Gear boxes (8): [92–99]	Prognosis of Mechanical Components (9): [142–150]
ML (5): [57–61]	Aircraft (4): [100–103]	Smart Sensors (13): [151–163]
SL (2): [62,63]	Rotors (3): [104–106]	
RL (2): [64–67]	Others (13): [107–119]	

Artificial intelligence = AI; Deep learning = DL; Meta-learning = ML; Unsupervised learning = UL; Supervised learning = SL; Reinforcement learning = RL; Convolutional neural networks = CNN; Deep neural networks = DNN; Restricted boltzmann machine = RBM; Recurrent neural networks = RNN; (Number) = Number of literature.
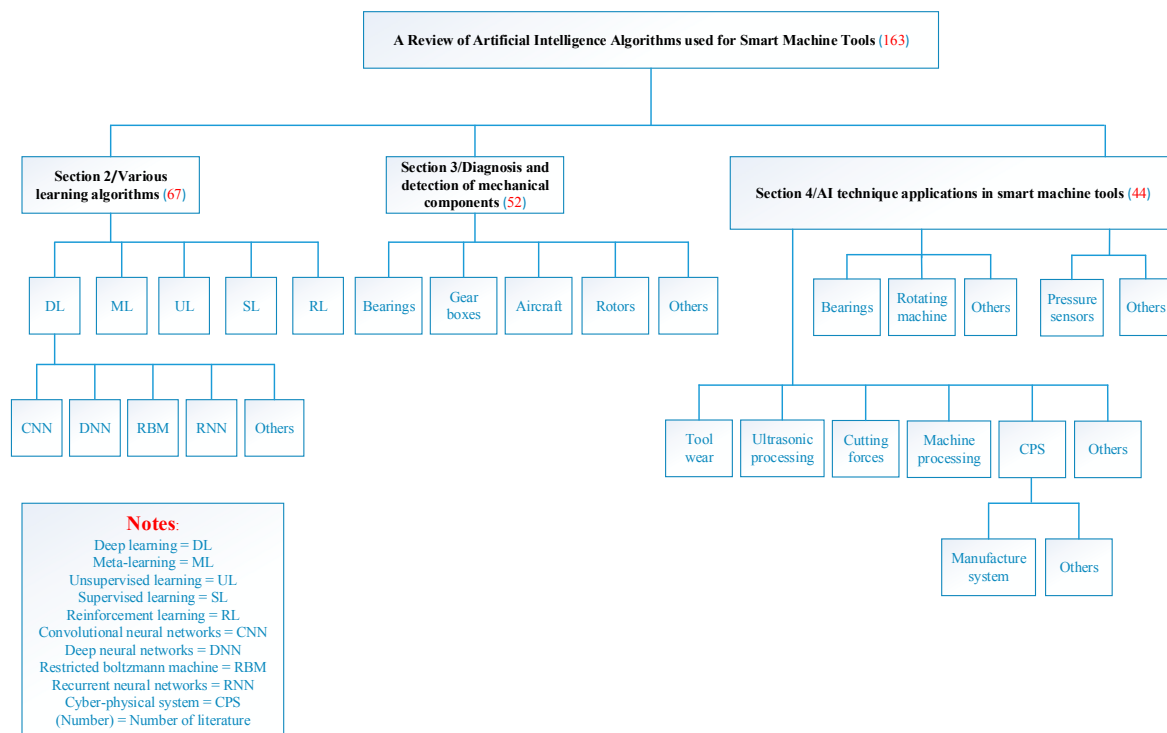

Figure 1. The architecture of artificial intelligence (AI) schemes used for smart machine tools.

2. Various Learning Algorithms

2.1. Recurrent Neural Networks (RNNs)

Giles et al. [4] showed that a recurrent, second-order NN method using a real-time, forward training algorithm readily learned to infer small regular grammar samples from the positive and negative strings used for training. All simulations were performed with random initial weight strength and usually converged after about a hundred training epochs. They also discussed a quantization algorithm for dynamically extracting finite state automata during and after training. Some of the extracted automata outperformed the trained NN scheme for the classification of unseen strings. Later, Funahashi and Nakamura [5] proved that any finite time trajectory of a given n-dimensional dynamical system can be approximated by the internal state of the output units of a continuous time

RNN approach with n output units, some hidden units, and an appropriate initial condition. They also demonstrated that any continuous curve can be approximated by the output of an RNN method. Hochreiter and Schmidhuber [6] addressed a novel, efficient, gradient-based method called long short-term memory (LSTM) to solve the complex and artificial long time-lag tasks. The LSTM is local in space and time; its computational complexity per time step and weight is $O(1)$. Their experiments with artificial data involved local, distributed, real-valued, and noisy pattern representations. LSTM had more successful runs, and learns much faster, than real-time recurrent learning, back propagation through time, recurrent cascade correlation, Elman nets, and neural sequence chunking. However, Gers et al. [7] identified a weakness in LSTM networks processing continual input streams that were not a priori segmented into subsequences with explicitly marked ends where the internal state of the network could be reset. Without resets, the state could grow indefinitely and eventually cause the network to break down. Their remedy was a novel, adaptive "forget gate" that enabled an LSTM cell to learn to reset itself at appropriate times to release internal resources. They also reviewed the illustrative benchmark problems in which standard LSTM outperforms other RNN algorithms.

In 2002, Jaeger [8] interpreted several traditional training methods including the RNNs, covering back-propagation through time (BPTT), real-time recurrent learning (RTRL), and extended Kalman filtering approaches (EKF). They were intended to be useful as a stand-alone tutorial for the echo state network (ESN) scheme for RNN training. Gers et al. [9] found that LSTM augmented by "peephole connections" from internal cells to multiplicative gates can learn the fine distinction between sequences of spikes spaced either 50 or 49 time steps apart without the need of help from any short training exemplars. Their LSTM variant also learned to generate stable streams of precisely timed spikes and other highly nonlinear periodic patterns without any external resets or teacher forcing. This made LSTM a promising method for tasks that required accurate measurement or the generation of time intervals. Cho et al. [10] proposed a novel RNN Encoder–Decoder that consisted of two RNNs. One RNN encoded a sequence of symbols into a fixed-length vector representation, and the other decoded the representation into another sequence of symbols. The encoder and decoder were jointly trained to maximize the conditional probability of a target sequence, given a source sequence. Moreover, they demonstrated that the proposed model learned a semantically and syntactically meaningful representation of linguistic phrases.

In 2014, some different types of RNNs were used to evaluate the tasks of polyphonic music and speech signal modeling [11]. This study revealed that these advanced recurrent units were indeed better than more traditional recurrent units such as tanh units. Recently, the deep ANNs (including recurrent ones) have won numerous contests in pattern recognition and machine learning [12]. This historical survey compactly summarized relevant work, much of it from the previous millennium. The author reviewed the deep supervised learning (also recapitulating the history of backpropagation), UL method, RL schemes and evolutionary computation, and indirect search for short programs encoding deep and large networks. After that, with the development of deep learning methods in the last few years [13], the representation learning from raw data has been redefined. Among DL models, the LSTMs are able to capture long-term dependencies and model sequential data. Thus, the LSTMs were able to work on the sensor data about machine condition. The basic and deep LSTMs were designed to predict the actual tool wear on the basis of raw sensor data. Experimental results show that these models, especially deep LSTMs, could outperform several state-of-the-art baseline schemes. Zhao et al. [14] redefined representation learning from raw data, and a DNN structure and convolutional bi-directional long short-term memory networks (CBLSTM), were designed to address raw sensory data. The LSTMs could capture long-term dependencies and model sequential data, and the bi-directional structure enables the capture of past and future context. Stacked, fully-connected layers and the linear regression layer are built on top of bi-directional LSTMs to predict the target value. Furthermore, a real-life tool-wear test was introduced, and the proposed CBLSTM could predict the actual tool wear based on raw sensor data. Furthermore, these abovementioned references were represented the problem-solving revolutions of mechanical components rather than smart machine tools.

2.2. Convolutional Neural Networks (CNNs)

Cun et al. [15] used a back-propagation network to recognize handwritten digits and minimal preprocessing of the data was required. However, the architecture of the network needed to be constrained and specifically designed for the task. The input of the network consisted of normalized images of isolated digits. The approach had a 1% error rate and about a 9% reject rate on zipcode digits. Later, LeCun et al. [16] mentioned that the various methods were applied to handwritten character recognition and compared them using a standard handwritten digit recognition task. The CNNs, which were specifically designed to deal with the variability of two dimensional (2D) shapes, outperformed all the other techniques. Jarrett et al. [17] proposed three issues: (1) How do non-linearities that follow the filter banks influence recognition accuracy? (2) Does filter bank learning in an unsupervised or supervised manner affect the performance or make it better than that of random or hardwired filters? (3) Is there any advantage in using an architecture with two stages of feature extraction, rather than one? They demonstrated that using non-linearities that included rectification and local contrast normalization was the single most important ingredient for good accuracy on object recognition benchmarks.

Deng et al. [18] then constructed a new database called "ImageNet", a largescale ontology of images built upon the backbone of the WordNet structure. ImageNet aims to populate the majority of the 80,000 synsets of WordNet with an average of 500–1000 clean and full resolution images. This led to tens of millions of annotated images organized by the semantic hierarchy of WordNet. This study provided a detailed analysis of ImageNet in its current state: 12 subtrees with 5247 synsets and 3.2 million images in total. This showed that ImageNet was much larger in scale and diversity and much more accurate than current image datasets. Sermanet et al. [19] classified the digits of real-world house numbers using convolutional neural networks (ConvNets). The ConvNets are hierarchical feature learning NNs whose structure is biologically inspired. Unlike many popular vision schemes that are hand-designed, the ConvNets can automatically learn a unique set of features optimized for a given work.

Later, Krizhevsky et al. [20] trained a large, deep CNN method to classify the 1.2 million high-resolution images in the ImageNet LSVRC-2010 contest into 1000 different classes as presented in Figure 2. To make training faster, they used non-saturating neurons and a very efficient graphics processing unit for the convolution. To reduce over-fitting in the fully-connected layers, they employed a recently-developed regularization method called "dropout" that proved to be very effective. Abdel-Hamid et al. [21] proposed a CNN approach to speech recognition within the framework of hybrid neural network Hidden Markov Models (NNHMM) and used local filtering and max-pooling in the frequency domain to normalize speaker variance. In their method, a pair of local filtering layers and a max-pooling layer was added at the lowest end of the NN to normalize spectral variations of speech signals. In their experiments, the proposed CNN architecture was evaluated in a speaker independent speech recognition task using the standard DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT) data sets. Experimental results showed that the proposed CNN method can achieve more than 10% relative error reduction in the core TIMIT test sets when compared with a regular NN utilizing the same number of hidden layers and weights. Their results also showed that the best results of the presented CNN model were better than previously published results using the same TIMIT test sets that employed a pre-trained deep NN model.

Kim [22] conducted a series of experiments with a trained CNN on top of pre-trained word vectors for sentence-level classification tasks. They also showed that a simple CNN with little hyperparameter tuning and static vectors could achieve excellent results on multiple benchmarks. Fine-tuning of the learning task-specific vectors offers further gains in performance. The CNN models discussed here showed improvement over the state-of-the-art on four out of seven tasks, which included sentiment analysis and question classification. After that, because the DNNs are more difficult to train, He et al. [23] presented a residual learning framework to ease the training of networks that were substantially deeper than those previously utilized. They explicitly reformulated the layers as learning residual

functions with reference to input, instead of learning unreferenced functions. In addition, they provided comprehensive empirical evidence that showed these residual networks were easier to optimize and could gain accuracy from their considerably increased depth.

Yan et al. [24] addressed an extended DL method to achieve promising performance in classifying skewed multimedia datasets. Specifically, they studied the integration of bootstrapping schemes and a state-of-the-art DL CNN method in an extensive empirical study. Considering the fact that DL algorithms such as CNNs were usually computationally expensive, they used the low-level features of CNNs and achieved promising performance while saving a lot of training time. The experimental results displayed the effectiveness of the framework in classifying severely imbalanced data in the TRECVID data set. Ren et al. [25] interpreted a Region Proposal Network (RPN) that shared full-image convolutional features with the detection network, therefore enabling nearly cost-free regional proposals. An RPN algorithm is a fully convolutional network that can foresee object bounds and objectness scores at each position simultaneously. The RPN method is trained end-to-end to generate high-quality region proposals, which are used by fast R-CNN schemes for detection. They also further merged the RPN approach and fast R-CNN algorithm into a single network by sharing their convolutional features—using the recently popular terminology of neural networks with "attention" mechanisms.
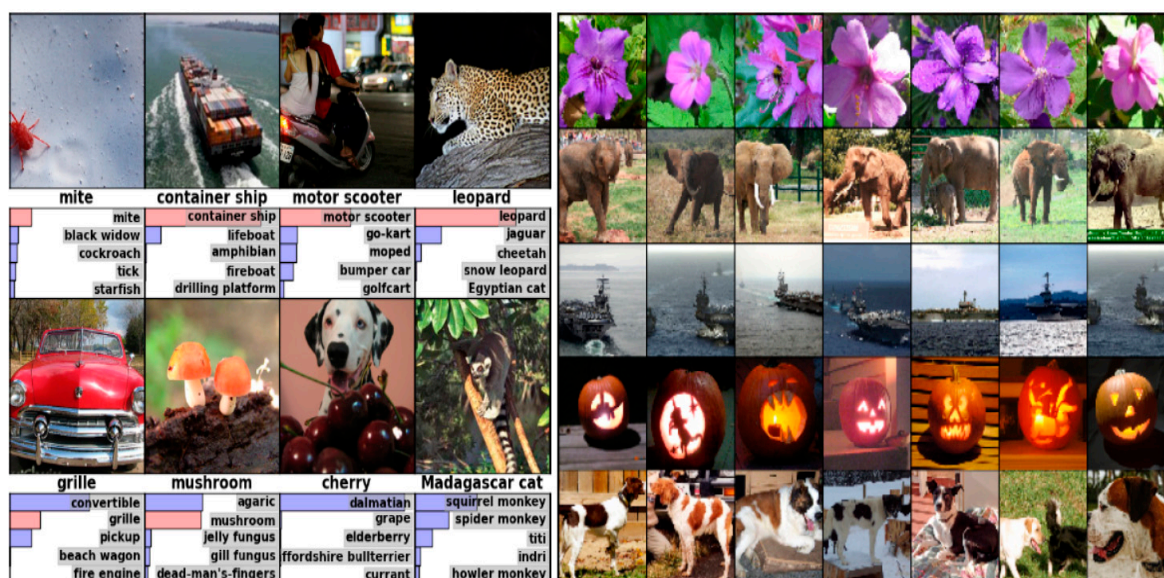

Figure 2. (**Left**) Eight ILSVRC-2010 test images and the five labels considered most probable by our model. The correct label is written under each image, and the probability assigned to the correct label is also shown with a red bar (if it happens to be in the top 5); (**Right**) Five ILSVRC-2010 test images in the first column. The remaining columns show the six training images that produce feature vectors in the last hidden layer with the smallest Euclidean distance from the feature vector for the test image [20].

Later, Babu et al. [26] compared several state-of-the-art schemes on two publicly available data sets to evaluate the effectiveness of this current method. The encouraging results showed that they presented the deep CNN based regression algorithm for RUL estimation was more efficient and more accurate. Weimer et al. [27] established a new paradigm from machine learning, namely deep machine learning by examining the design configuration of deep CNN and the impact of different hyper-parameter settings on the accuracy of defect detection results. In contrast to manually designed image processing solutions, the deep CNN automatically generates powerful features by hierarchical learning strategies from massive amounts of training data with a minimum of human interaction or expert process knowledge. They also claimed that this approach showed excellent defect detection results with low false alarm rates. After that, Huang et al. [28] proposed extensive

and systematic experiments to validate the effectiveness of these classic schemes for representation learning on class-imbalanced data. They further showed that more discriminative deep representation could be achieved by forcing a deep network to maintain both inter-cluster and inter-class margins. This tighter constraint effectively reduced the class imbalance inherent in the local data neighborhood. They showed that the margins could be easily deployed in a standard deep learning framework through quintuplet instance sampling and the associated triple-header hinge loss. Furthermore, they found significant improvement over existing methods in both high- and low-level vision classification tasks that exhibited imbalanced class distribution. Moreover, these above-mentioned references were demonstrated the solving revolutions of image processing rather than using an efficient scheme.

2.3. Deep Neural Networks (DNNs)

Collobert and Weston [29] presented a single DNN architecture that, given a sentence, outputs a host of language processing predictions: part-of-speech tags, chunks, named entity tags, semantic roles, semantically similar words, and the likelihood that the sentence makes sense (grammatically and semantically) using a language model. The entire network was trained jointly on all these tasks using weight-sharing, an instance of multitask learning. They also demonstrated how both multitask learning and semi-supervised learning improve the generalization of shared tasks, resulting in state-of-the-art performance. Later, Maaten and Hinton [30] developed a new technique called "t-SNE" that visualized high-dimensional data by giving each datapoint a location in a two- or three-dimensional map. Furthermore, the t-SNE is better than existing techniques at creating a single map that reveals structure on many different scales. They used t-SNE to visualize the structure of very large data sets by employing random walks on neighborhood graphs to allow the implicit structure of all the data to influence the way a subset was displayed. They also showed how t-SNE performed on a wide variety of data sets and compared it with many other non-parametric visualization techniques, including Sammon mapping, isomap, and locally linear embedding.

In [31], Hinton et al. mentioned that most speech recognition systems used HMMs to tackle the temporal variability of speech and Gaussian mixture models (GMMs) to determine how well each state of an HMM fitted a frame or a short window of frames of coefficients that represented acoustic input. An alternative way to evaluate the fit was to use a feed-forward NN that took several frames of coefficients as input and produced posterior probabilities over HMM states as output. The DNNs have many hidden layers, are trained using new methods, and have been shown to outperform GMMs on a range of speech recognition benchmarks. They offered an overview of this progress that represented the shared views of four research groups that had shown recent success in using DNNs for acoustic modeling in speech recognition.

Yosinski et al. [32] proposed two tools: They first visualized the activity in each layer of a trained convnet as it processes an image or video (such as a live webcam stream). They found that looking at live activations that changed in response to user input helped to build valuable intuition about how convnets work. The second tool enabled features to be visualized at each layer of a DNN using regularized optimization in image space. Later, Lu et al. [33] presented a novel deep NN model with domain adaptation for fault diagnosis. Two main conclusions were reached by comparisons with previous work: first, the proposed model can use domain adaptation while strengthening the representative information of the original data to achieve high classification accuracy in the target domain; second, several strategies were addressed to investigate the optimal hyperparameters of the model. Zhang et al. [34] investigated the diagnosis and prognosis of learning methods and used multi-objective optimization to tackle failure diagnosis. Liao et al. [35] used an enhanced restricted Boltzmann machine to deal with prognosis and health assessment, and Zhang et al. [36] used DBNs to solve remaining useful life estimation in prognostics. In addition, these abovementioned references were presented the problem-solving by using many hidden layers rather than time-saving schemes.

2.4. The Other Methods

There are a number of different learning algorithms: Hinton et al. [37] showed how to use "complementary priors" to eliminate the explaining away effects that made inference difficult in densely connected belief nets that have many hidden layers. They used complementary priors to derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, the top two layers being offered in the form of undirected associative memory. Bengio et al. [38] offered deep empiric multi-layered NNs and explored variants to better understand their success. They extended their use to cases where the input was continuous, or where the structure of the input distribution did not reveal enough about the variables to be predictive in a supervised task. Their experiments also showed that greedy, layer-wise, unsupervised training strategy mainly helped optimization by initializing weight in a region near a good local minimum. This gave rise to internal distributed representations that were high-level abstractions of the input, which gave better generalization.

Salakhutdinov and Hinton [39] demonstrated a new learning algorithm for Boltzmann machines that contained many layers of hidden variables as presented in Figure 3. The data-dependent expectations were estimated employing a variational approximation that tended to focus on a single mode, and data-independent expectations that were approximated by persistent Markov chains. Two quite different techniques for estimating the two types of expectation that enter into the gradient of the log-likelihood made it practical for Boltzmann machines to learn with multiple hidden layers and millions of parameters. Ng [40] introduced the SAE learning algorithm for the automatic learning of features from unlabeled data. In some domains, such as computer vision, this approach is not by itself competitive with the best hand-engineered features, but the features it can learn do turn out to be useful for a range of problems (including some in audio, text and so forth). Moreover, there are more sophisticated versions of the SAE (not described in these notes) that do surprisingly well, and in many cases are competitive with, or superior to, the best of hand-engineered representations.
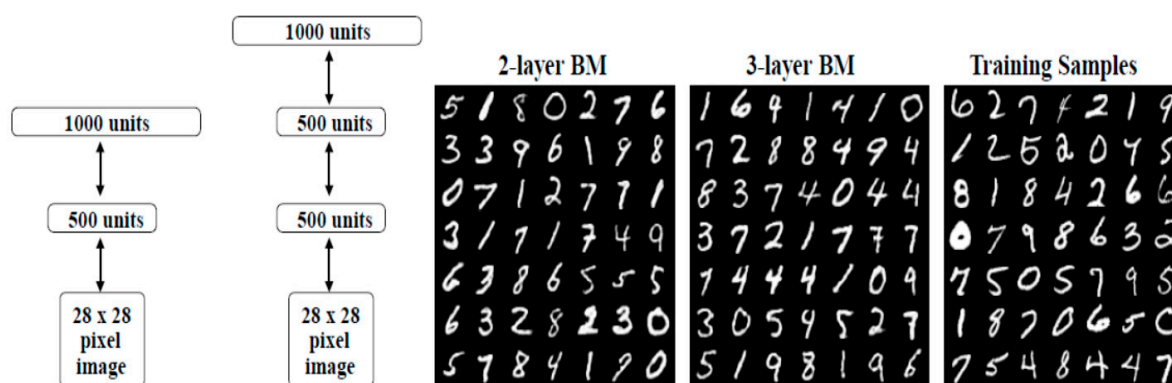

Figure 3. (**Left**): Two deep Boltzmann machines used in experiments; (**Right**): Random samples from the training set, and samples generated from the two deep Boltzmann machines by running the Gibbs sampler for 100,000 steps. The images shown are the probabilities of the binary visible units given the binary states of the hidden units [39].

Bengio et al. [41] reviewed recent work done in the area of unsupervised feature learning and DL, covering advances in probabilistic models, auto-encoders, manifold learning, and deep networks. This motivated investigation of longer term unanswered questions about the appropriate objectives for learning good representations, for computing representations (inference), and geometrical connections between representation learning, manifold learning, and density estimation. Later, Cambria et al. [42] interpreted the ELM method as an emerging learning technique that provides efficient unified solutions to generalized feed-forward networks, including, but not limited to (single- and multi-hidden-layer) NNs, radial basis function networks, and kernel learning.

Thirukovalluru et al. [43] showed a brief survey of traditional handcrafted features and later exhibited a short analysis of handcrafted features v/s features learned by DNN, for fault diagnosis. The DNN based features were generated in 3 phases: (1) extracted handcrafted features using traditional techniques; (2) initialization of the weights of DNN by learning de-noising sparse auto-encoders with handcrafted features in an unsupervised fashion and (3) the application of two generic fine-tuning heuristics that tailored DNN weight to give good classification performance. After that, Reddy et al. [44] addressed the existing leveraging unsupervised learning methods based on deep auto-encoders (DAE) on raw time series data from multiple sensors to build a robust model for anomaly detection. Their anomaly detection algorithm analyzes the reconstruction error of a DAE trained on nominal data scenarios. The reconstruction errors of individual sensors were examined to perform fault disambiguation. Training and validation were conducted in a laboratory setting for various operating conditions. They mentioned that the current framework did not need any hand-crafted features and employed raw time series data.

In addition, Wang et al. [45] constructed a novel CSAE that can be used for unsupervised feature learning. The CSAE adds a Gaussian stochastic unit into the activation function to extract features of nonlinear data. The CSAE was applied to solve the problem of transformer fault recognition. Comparative experiments clearly showed that CSAE can extract features from the original data and achieved a superior correct differentiation rate on transformer fault diagnosis. In several publications [46–51], DL algorithms were used in tissue-regulated splicing code, methods and applications, and machine health monitoring. Later, ML methods were utilized in nonlinear distributed representations [52] and some test images, auto-encoders [53], a flurry of machine learning applications [54], problems in building high-level and class-specific feature detectors using unlabeled data [55], and video representations [56]. After that, ML schemes were employed in selective meta-learning [57], automatic kernel selection [58], kernel and regularization parameters [59], model selection and hyperparameter optimization [60], and distance-based problem characterization [61].

García et al. [62] developed a taxonomy on the basis of the main properties pointed out in previous research, unifying the notation and including all the known methods up to date. They conducted an empirical experimental survey in supervised classification involving the most representative and newest discretizers, different types of classifiers, and a large number of data sets. Their findings, measured in terms of accuracy, number of intervals, and consistency have been verified by means of nonparametric statistical tests. Han et al. [63] proposed a novel and effective geospatial object detection framework by combining weakly supervised learning (WSL) with high-level feature learning. Furthermore, this WSL scheme was subjected to object detection where the training sets require only binary labels indicating whether an image contains the target object or not. On the basis of the learnt high-level features, it jointly integrates saliency, intraclass compactness, and interclass separability in a Bayesian framework to initialize a set of training examples from weakly labeled images and start iterative learning from the object detector.

Later, Munos et al. [64] evaluated the contractive nature of related operators for off-policy, return-based reinforcement learning and control settings and derived online sample-based algorithms. Lample and Chaplot [65] established the first architecture to tackle 3-D environments in first-person shooter games, that involved partially observable states. Representatively, the deep reinforcement learning (RL) methods use only visual input for training. The authors also demonstrated an algorithm to augment these models and developed game feature information such as the presence of enemies or objects during the training phase. Arulkumaran et al. [66] highlighted the unique advantages of deep neural networks, focusing on visual understanding via reinforcement learning. Watkins and Dayan [67] demonstrated that Q-learning converged to the optimum action-values with probability 1 so long as all actions were repeatedly sampled in all states and the action-values were represented discretely. However, these above-mentioned references were represented the complex problem-solving processes rather than the simple and efficient algorithms.

3. Diagnosis and Detection of Mechanical Components

3.1. Bearings

Bearings play a significant role in all motors and other rotating systems. Many issues arising in motor operations are linked to bearing behavior. In many cases, the accuracy of the devices and instruments used to control and monitor the motor system is dependent on the dynamic performance of the motor bearings. This makes fault diagnosis a vital part of the management of all machines containing bearings. Bearing vibration frequency for motor bearing fault diagnosis is discussed by Li et al. [68]. They proposed a method for motor rolling bearing fault diagnosis using time/frequency-domain bearing vibration analysis and NNs. Later, Samanta and Al-Balushi [69] investigated fault diagnosis of rolling element bearings using ANN. The characteristic features of time-domain vibration signals from rotating machinery with normal and defective bearings were used as inputs to ANN consisting of input, hidden, and output layers. ANN was trained using a back propagation algorithm with a subset of experimental data from known machine conditions. The results displayed the effectiveness of ANN for the diagnosis of bearing condition. The proposed procedure needed only a few features extracted from the measured vibration data either directly or with simple preprocessing.

Later, Malhi and Gao [70] used a feature selection scheme based on a principal component analysis (PCA) approach. The effectiveness of the scheme was verified experimentally on a bearing test bed, using both supervised and unsupervised defect classification algorithms. The objective of the investigation was to identify the severity level of bearing defects, where no a priori knowledge about the defect condition was available. Widodo and Yang [71] conducted a survey of fault diagnosis and machine condition monitoring using the SVM method. They made an attempt to summarize and review recent research and development of SVM schemes in machine condition monitoring and diagnosis. Numerous methods have been developed based on intelligent systems such as ANN, fuzzy expert systems (ES), condition-based reasoning, random forest, and so forth. However, an SVM approach to fault diagnosis and machine monitoring is rarely used. They stated that the SVM method had excellent general performance and was capable of producing highly accurate results in machine condition monitoring and diagnosis.

Su and Chong [72] suggested that model-based schemes were efficient monitoring systems that could provide warning and prediction in the early stages of certain kinds of faults. However, conventional methods have to work with particular models of motor and cannot be utilized effectively for vibration signal diagnosis because adaption to the random nature of the vibration signals is not possible. They proposed an analytical redundancy algorithm employing NN modeling of the induction motor vibration spectrum for diagnosis and fault detection. Faults were detected from changes in the expected vibration spectrum model. Later, Seryasat et al. [73] used ANN and SVM on experimental vibration data from a rotating machine. They studied the role and effects of different vibration signals as well as signal preprocessing techniques. Recently, Tao et al. [74] and Junbo et al. [75] utilized SAE methods to deal with bearing faults and diagnosis in roller bearing systems. Lu et al. [76] established a feature extraction approach using DNN to extract a meaningful representation of bearing signals. The DNN approach is a new kind of machine learning tool with strong power of representation. It has been used successfully as a feature extractor in many practical applications.

Fei et al. [77] constructed a transfer learning-based approach for bearing fault diagnosis, where a transfer strategy was proposed that improved the diagnosis of bearing behavior under various operating conditions. The major concept of transfer learning was the use of selective auxiliary data to assist target data classification in which weight adjustment was made in a TrAdaBoost scheme to enhance diagnostic capability. Negative transfer was avoided through similarity judgment, which improved accuracy and relaxed the computational load of the current method. Later, Jia et al. [78] developed a novel intelligent scheme based on DNNs to overcome deficiencies of these intelligent diagnosis approaches. The effectiveness of the proposed method was validated using datasets from

rolling element bearings and planetary gearboxes. This data is comprised of massive sets of measured signals from bearings in different states of health under various operating conditions. Liu et al. [79] then proposed a new rolling bearing fault diagnosis approach that was based on short-time Fourier transform and stacked SAE. This scheme analyzed the surrounding signals. After spectrograms were obtained by short-time Fourier transform, the stacked SAE was utilized to automatically extract fault features, and softmax regression was used to classify the fault modes.

Guo et al. [80] established a new bearing condition recognition approach based on multifeature extraction and DNN. First, the scheme estimated time domain, frequency domain, and time-frequency domain features to represent the characteristics of the vibration signals. Then, a nonlinear dimension reduction algorithm based on DL was proposed to reduce redundancy. Finally, the top-layer classifier of deep NN outputs the bearing condition. Later, Mao et al. [81] applied an auto-encoder-ELM-based diagnosis scheme to diagnose bearing faults and overcome the earlier mentioned deficiencies. This study performed a comparative analysis of the current approach with some state-of-the-art algorithms, and the experimental results on rolling element bearing data sets displayed the effectiveness of the present method, with adaptive mining of the discriminative fault characteristics and high diagnosis speed.

Ma et al. [82] assessed the degradation of bearing performance with a new approach based on the Weibull distribution and DBN. A healthy state, and five states of degradation, were determined using Weibull distribution of fitted vibration feature. This was done to avoid areas of statistical parameter fluctuation. The DBN method can model nonlinear time series, so it was used to classify the different states of the bearing. Tao et al. [83] employed a novel fault diagnosis approach utilizing multivibration signals and the DBN method. The current scheme could adaptively fuse multifeature data and identify various bearing faults by using the learning ability of DBN. First, the multiple vibration signals are extracted from various faulty bearings. Then some time-domain characteristics are extracted from the original signals from each individual sensor, and finally, DBN was used with the features data of all the sensors to generate an appropriate classifier for complete fault diagnosis.

Later, Gan et al. [84] utilized a novel hierarchical diagnosis network (HDN) and collected the DBNs layer by layer for the hierarchical identification of mechanical systems. The deeper layers in an HDN scheme present a more detailed classification of the results generated from the last layer to offer representative features for different application. The experimental results showed HDN to be highly reliable for precise multi-stage diagnosis and it also overcame the overlap issue caused by noise and other disturbances. Oh [85] used an unsupervised approach to extract features from correlated vibration signals. First, the raw vibration signals from a pair of sensors were preprocessed by generating two-dimensional (2-D) images. Then, the vibration images were characterized with a HOG (histogram of original gradients) descriptor. Finally, the DL method was used to extract relevant features for journal bearing rotor system diagnosis.

Janssens et al. [86] and Guo et al. [87] applied CNN to bearing fault detection and diagnosis. Lu et al. [88] studied an effective and reliable DL scheme known as a stacked denoising auto-encoder (SDA), that was shown to be suitable for the identification of bearing state using signals containing ambient noise and working condition fluctuations. SDA has become a popular method to accomplish the promised merits of deep architecture-based robust feature representations. To improve the fault diagnosis reliability, a new multisensor data fusion technique was proposed by Chen and Li [89]. First, the time- and frequency-domain features were extracted from the different sensor signals, and input into multiple two-layer SAE NNs for feature fusion. The fused feature vectors could be regarded as machine health indicators and used to train a DBN method for further classification.

Ding and He [90] established a novel energy-fluctuated multiscale feature mining method using WPE images and deep ConvNet as bases for spindle bearing fault diagnosis. This was different from the vector characteristics applied in intelligent diagnosis of spindle bearings, in that a wavelet packet transform was first combined with phase space reconstruction to rebuild a 2D WPE image of the frequency subspaces. This special image could be used to reconstruct the local relationship of the WP

nodes and hold the energy fluctuation of the measured signal. The identifiable characteristics could then be further studied by the special architecture of the deep ConvNet. In addition to the traditional NN architecture, deep ConvNet combined the skipping layer with the last convolutional layer, as input to the multiscale layer, to simultaneously maintain global and local information. Mao et al. [91] later presented a new OS-ELM approach for solving the online sequential data imbalance issue. The acquired granules and principal curves were rebuilt online using the bearing data which arrived in sequence, and after an over- and under-sampling process, the balanced sample set was used to update the diagnosis model dynamically. The comparative results showed that the current algorithm was more reliable, accurate and effective than the others. Furthermore, these above-mentioned references were represented several problem-solving revolutions of bearings; however, researchers will develop some easy, stable and accurate methods in the future.

3.2. Gearboxes

Rafiee et al. [92] and Zhang et al. [93] utilized the ANN approach to deal with gear and bearing faults in a typical gearbox system and fault conditions in equipment and machine components. Li et al. [94] applied a multimodal deep support vector classification (MDSVC) scheme using separation–fusion, based on a DL method, for fault diagnosis in gearboxes. Because different modalities can describe the same object, multimodal homologous features of the gearbox vibration measurements were first separated in the time, frequency and wavelet modalities. A Gaussian-Bernoulli deep Boltzmann machine (GDBM) without final output was subsequently used to study the pattern representations for features in each modality. A support vector classifier was then applied to fuse GDBMs in different modalities to establish an MDSVC model. The "deep" representations from "wide" modalities improve fault diagnosis capacity. After that, Chen et al. [95] proposed multiple classifiers on the basis of multi-layer neural networks (MLNN) to implement vibration signals for fault diagnosis in gearboxes. They presented an MLNN-based learning architecture utilizing a deep belief network (MLNNDBN) for gearbox fault diagnosis.

Later, Chen et al. [96] established an implementation of CNN for gearbox fault identification and classification. Different combinations of condition patterns on the basis of some basic fault conditions were studied. Twenty test cases with different combinations of condition pattern were employed and each test included 12 combinations of different basic condition pattern. They claimed that their approach gave better results than (then current) peer methods for the detection of faults and fault patterns in gearbox and bearing systems. To tackle this issue, a model for deep statistical feature learning from vibration measurements of rotating machinery was presented by Li et al. [97] and the flowchart. Vibration sensor signals were collected from rotating mechanical systems in the time, frequency, and time-frequency domains, each of which was then used to generate a statistical feature set. For the learning statistical features, real-value Gaussian-Bernoulli restricted Boltzmann machines (GRBMs) were stacked to develop a Gaussian-Bernoulli deep Boltzmann machine.

Later, Li et al. [98] proposed a deep random forest fusion (DRFF) technique to improve fault diagnosis performance for gearboxes by utilizing measured simultaneous signals from an acoustic emission (AE) sensor and an accelerometer to monitor gearbox condition. Statistical parameters of wavelet packet transforms (WPT) were generated from the AE and the vibration signals. This method displayed better performance than other peer methods and deep learning fusion of acoustic and vibratory signals seemed to improve the diagnosis of faults in gearboxes. Xie et al. [99] used cross-domain feature extraction and fusion from the time and frequency domains with spectrum envelop preprocessing, as well as domain synchronization averaging using transfer component analysis (TCA), for gearbox fault diagnosis. Since TCA is based on kernel methods, the effect of different kernels, including the Gaussian, linear, polynomial, and polyplus kernels on the performance of TCA, was investigated in comprehensive experiments with a gearbox test-bed under various operating conditions. Besides, these above-mentioned references were shown several problem-solving revolutions of gearboxes. Nevertheless, there are no simple approaches.

3.3. Aircraft

Tamilselvan et al. [100] established a novel multi-sensor health diagnosis scheme utilizing DBN, a method that has recently become quite popular for use in machine learning. It has several advantages such as fast inference and the ability to encode rich, high order, network structures. DBN uses a hierarchical structure with multiple stacked Restricted Boltzmann Machines and works through a successive, layer by layer, learning process. The present multi-sensor health diagnosis methodology employing DBN based state classification can be structured in three consecutive stages: (1) a defining of health states and preprocessing the sensory data for DBN training and testing, (2) development of DBN based classification models for the diagnosis of predefined health states, and (3) validation of the DBN classification models by the testing of sensory datasets. Tamilselvan and Wang [101] used the same approach as before to deal with aircraft engine and electrical transformer health diagnosis.

Later, Li and Wang [102] derived a multi class classification scheme based on a DL approach. The method utilized SAE for initial weighting and offset of the multi-layer NN, and the parameters were monitored after initialization with a gradient descent approach. The scheme could overcome many of the weaknesses of SVM, such as complication and the fact that it occupied more space with large amounts of data or many categories. By studying measured data, expert knowledge can be offered for a spacecraft health management platform. Using LSTMNN, Yuan et al. [103] obtained good results for the diagnosis and prediction of performance in complex operations, hybrid faults and in the presence of intense noise. The whole proposition was studied and discussed before tests were carried out by examination of a monitored health dataset from aircraft turbofan engines provided by NASA. Performance of LSTM and some of its modifications were examined and compared. Moreover, these above-mentioned references were represented some problem-solving revolutions of aircraft; however, researchers will develop several parallel algorithms to accelerate the process in the future.

3.4. Rotors

Oppenheimer and Loparo [104] developed a physics-based scheme for diagnostics and prognostics using integrated observers and life models. Observers were filtered on the basis of physical models of machine-fault combinations and measured machine signatures were used to identify and characterize the state of a machine. Observers were adaptively deployed during machine wear and could be coupled with one another to handle interacting conditions and faults. The approach was detailed using a faulty (cracked) rotor shaft that interacted with gravity and was unbalanced. Observers for cracked shafts and imbalance were presented. The observers showed the machine condition and fault strengths and life models were used to determine remaining machine life. Later, Wang et al. [105] constructed a new automatic and intelligent fault diagnosis method based on the CNN method. Firstly, the vibration signal was processed by wavelet transform into a multi-scale spectrogram image that manifested the fault characteristics. The image was then fed directly into CNN to study the invariant representation for vibration signal and to determine the fault status for a diagnosis. During model construction, a rectifier neural activation function and a dropout layer were incorporated into the CNN scheme to improve the computational efficiency and model generalization.

After that, to tackle the unbalanced dataset issues in the machinery fault diagnosis area, Duan et al. [106] formulated a support vector data description system based on a machine learning model with a binary tree for multi-classification problems, and this was used for fault severity recognition and diagnosis. The binary tree structure of multiple clusters was first drawn on the basis of the order of cluster-to-cluster distances calculated by the Mahalanobis distance. The support vector data description model was then utilized for top to bottom classification using the binary tree structure. A particle swarm scheme was used to optimize the parameters of the support vector data description by using recognition accuracy as an objective function. The effectiveness of this approach was validated by rotor unbalance severity classification. Furthermore, these above-mentioned references were demonstrated some approaches of rotors, but there are no accurate schemes.

3.5. Other Mechanical Components

Sun et al. [107] proposed a DNN approach to Induction Motor fault diagnosis. The scheme employed SAE to study features. This was part of unsupervised feature learning that only needed unlabeled measurement data. Partial corruption was added to the input of the SAE with the aid of denoising code to improve the robustness of feature representation. Features learned from SAE were then used to train an NN classifier for identifying Induction Motor faults. To prevent overfitting during the training process, a recently developed regularization approach called "dropout" that proved to be very effective in NN, was also employed. An experiment performed on a machine fault simulator showed that the SAE-based DNN was better at feature learning and classification in Induction Motor fault diagnosis than traditional NN.

Shao et al. [108] established a DBN-based method for the automatic extraction of relevant features from vibration signals that characterize the working conditions of an Induction Motor. The DBN model utilized a structure with stacked RBMs and was trained by an efficient learning approach called greedy layer-wise training. Vibration signals were input to the DBN and fault diagnosis was provided by the output from the activation functions of the trained network. In contrast to the traditional feature extraction algorithms for Induction Motor fault diagnosis, such as a wavelet packet transform, this approach allowed the direct learning of features from the vibration signal. This gave comparable results with high classification accuracy. Ince et al. [109] presented a fast and accurate motor condition monitoring and early fault-detection system using one-dimensional CNNs that had an inherent adaptive design which fused the feature extraction and classification phases of motor fault detection into a single learning body. The method was directly applicable to the raw data (signal) which excluded the need for a separate feature extraction approach leading to more efficiency in terms of both hardware and speed.

A feature-extraction scheme and frequency analyzer were proposed by Kuo [110] in which features were formulated as the input of an ANN method that utilized back propagation. Fuzzy models were used to dynamically update the training parameters, training rate, momentum, and steepness of the activation function to increase training speed. Later, Aminian and Aminian [111] derived an analog-circuit fault diagnostic system based on back propagation NNs using wavelet decomposition, principal component analysis, and data normalization as preprocessors in a neural network. This system could detect and identify faulty components in an analog electronic circuit by analyzing its impulse response. Wavelet decomposition was used for preprocessing the impulse response and drastically reduced the number of inputs to the NN. This simplified the architecture and minimized both training and processing time. Fault diagnosis for other mechanical systems was addressed by others: the logistic regression method for a real elevator door system [112]; a Naïve Bayes scheme and Bayes net approach to monoblock centrifugal pump systems [113]; SAE for rotating machines and hydraulic pumps [114,115]; PKFA for machine tools [116] and the framework, DNN for tidal turbine generators [117], the DL method for wind generators [118], and CNNs for a structural damage detection system [119]. In addition, these above-mentioned references were displayed several problem-solving revolutions of mechanical components. However, there were some weaknesses so that researchers can propose many methods to solve them.

4. AI Applications in Smart Machine Tools

4.1. Intelligent Manufacturing

Monostori and Barschdorff [120] focused on ANNs or connectionist systems, which have the ability to integrate multiple sensor information, can function in real-time, have effective knowledge representation and can learn or adapt. A short survey of different ANN structures and learning algorithms was also given, as well as some common applications of NN techniques in fields other than intelligent manufacturing. The most popular back propagation learning procedures, and most important acceleration techniques, the competitive learning approach which has good prospects for

future applications, were all highlighted. In addition, they also surveyed the known NN applications and perspectives in intelligent manufacturing.

Chang and Chang [121] devised an integrated artificial intelligent (IAI) system for dynamic computer-aided process planning (CAPP). The system, IAI-CAPP, integrates fuzzy logic and ANN to perform dynamic workpiece recognition and adaptive-learning tasks as well as process plans. The concept included a pivotal feature for evaluating the suitability of existing process plans for incoming product design, the expert systems (ES) technique was also used. The system combined variant and generative CAPP and was able to generate plans suitable for new workpieces or workpieces that were similar to existing ones. The system was realized in a computer prototype program. The major aim of Sta et al. [122] was the real time analysis of the position of a product on a moving conveyor, the position data was sent to a Selective Compliance Assembly Robot Arm (SCARA) robot controller, which moved the product from the conveyor to a palletizing system. Real time data was acquired by three high resolution cameras. This reduced robot operation time, which was the slowest part of the system. An AI system was developed to reduce the time delay and ANNs were used in a pattern recognition system that has a very short time response.

The learning ability of NN was utilized by Hwang et al. [123] to learn the complex non-linear relationships between the many different mechanical properties of rolled steel bar. The billet composition and the fabrication parameters could be automatically controlled and developed. Lambiase [124] devised an ES that automatically designed and chose the rolling sequences for the production of square and round steel rod. The strategy was aimed at reducing the overall number of passes, assuming a series of process constraints such as the available roll cage power and torque, and the roll groove filling behavior. He pointed out that the ANN method was employed to foresee the technological requirements for achievement of the geometry of the semi-finished rolled product.

Fu et al. [125] used DBNs to analyze the vibration signals acquired from an end milling machine to fabricate feature space for the monitoring of cutting state. A greedy layer-wise strategy was adopted to pre-train the network and standard samples were utilized for fine-tuning by applying a back-propagation scheme. The results showed that the DL approach had similar ability to characterize the signal and cutting state monitoring as could be achieved manually. They also claimed that the modeling accuracy was much better than was possible with other more traditional modeling algorithms. Wang et al. [126] proposed a new virtual tool wear sensing technique that used a multisensory data fusion and artificial intelligence model for tool condition monitoring. Difficult-to-measure tool wear parameters, such as width wear, were assumed by fusing in-process multisensory data (force and vibration) by a dimension reduction technique to support a vector regression model.

Li and Si [127] discussed a review of the multiscale dynamics of modern manufacturing systems and Li et al. [128] studied a review of applications of AI in intelligent manufacturing. Kuo and Cohen [129] developed an on-line estimation system which included data acquisition, feature extraction, pattern recognition and multi-sensor integration for tool flank wear. A self-organizing, self-adjusting fuzzy model was proposed for multi-sensor integration. This was compared with an ANN error back propagation method and multiple regression model utilizing experimental data from force, vibration, and acoustic emission sensors. Li et al. [130] proposed a vision-based approach for wear identification in tools used for finish turning. An adaptive resonance theory (ART2) NN was embedded with fuzzy classifiers and the state of wear of the turning tool was determined from laser scatter images from the machined surfaces of the workpiece. This algorithm was like a visual inspection of the surface of a machined workpiece by an expert machinist. Nevertheless, the experimental results showed that the conventional technique for the evaluation of surface finish did not offer values that correlated well with tool wear. On the other hand, the laser light scatter image gave a good indication of the tool wear because it was not readily affected by buildup of material on the edge or cold-welded metal, scratches or other disruptions of the turned surface.

Kim et al. [131] used a pattern recognition approach to determine the strength of concrete by evidence accumulation on the basis of AI techniques with multiple feature parameters.

Pattern recognition was followed by an evidence accumulation procedure using distance measurements made with reference parameters. A fuzzy mapping function was used to transform the distance for the application of the evidence accumulation scheme. Norouzi et al. [132] used an AI approach to investigate the modeling and optimization of ultrasonic welding (USW) process parameters including welding time, pressure, and vibration amplitude that influenced the strength of welds in acrylonitrile butadiene styrene (ABS) and poly methyl methacrylate (PMMA). Spot welding experiments were carried out on samples of ABS and PMMA. The experimental data were used in an ANN method, adaptive neuro-fuzzy inference and hybrid systems. The ANN approach was better than the others and the best model was a feed-forward back-propagation network, with uniform transfer functions (TANSIG–TANSIG–TANSIG) and 4/2 neurons in the first/second hidden layers. The best predictor was then offered to the genetic algorithm (GA) and particle swarm optimization, as a fitness function and to optimize the USW machine parameters.

Irgolic et al. [133] studied the general characteristics of graded materials, previous industrial applications and potential future applications. The utilization of graded materials was steadily increasing and moving from the laboratory environment into everyday usage. However, the processing of graded material remained a big unknown, and presented a challenge for researchers and industry around the world. An investigation by Hanief et al. [134] into red brass machining was intended to develop a model to study the influence of cutting parameters (speed, depth of cut, and feed rate) on the cutting forces during turning operations using high speed steel tools. Experiments were carried out on a basis of full factorial design methodology to increase the reliability and confidence limit of the data. The ANN and multiple regression methods were employed to model the cutting forces based on cutting parameters.

Koker and Ferikoglu [135] used an NN controller with the traditional generalized predictive control algorithm (GPC) for robot control. The GPC algorithm, part of a class of Model Based Predictive Controls, needed a lot of computation time and this resulted in poor robot performance. To decrease processing time, by avoiding the highly mathematical computations of GPC, an NN was sketched for a three-joint robot. For intelligent fixture design, Hamedi [136] employed nonlinear finite element analysis (FEA) in a hybrid learning system with a supportive combination of ANN and GA. A model of a fixture system in which the workpiece was subject to cutting and clamping forces was solved using FEA. The ANN approach was needed to recognize a pattern between the clamping forces and the state of fixture contact and maximum workpiece elastic deformation. The ANN method made good use of the FEA results. Using the identified pattern, a GA-based program resolved the optimum values for clamping force that did not cause excessive deformation and stress in the component. The merit of this study was a determination of the exact state of contact between clamp and the workpiece.

Lee et al. [137], and others, carried out extensive studies on cyber physical architecture structure [CPS] in several areas. A unified five-level architecture was presented as a guideline for the implementation of CPS [137]. A consolidated review of the latest CPS literature and a complete review of international standards was offered by Trappey et al. [138]. A hierarchical hybrid model of Intelligent CPS was given by Claudi et al. [139]. A multi-layer perceptron neural network with a self-organized map and a support vector machine for obstacle detection under different weather conditions was presented by Castaño et al. [140]. Varshney and Alemzadeh [141] discussed machine learning schemes and safety in all sorts of CPS applications. Moreover, these above-mentioned references were demonstrated some solving revolutions of intelligent manufacturing. However, DNN, CNN, and RNN algorithms were not used for solving these issues.

4.2. Prognosis of Mechanical Components

Mahamad [142] addressed prognosis, classification and fault diagnosis in rotating machines. The vibration data for classification and fault diagnosis, was acquired from Western Reserved University. The signals were processed for feature extraction and selection in a pre-processing stage before the diagnosis and classification model was built. The acoustic emission and vibration

signals were used as input signals for fault prognosis. The model can be used as a tool for the diagnosis of failure in rotating machinery. Mahamad et al. [143] and Zhang et al. [144] utilized ANNs to study rotating machinery and obtained good results. Li et al. [145] constructed a stochastic defect-propagation model with a lognormal random variable in a deterministic defect-propagation rate model. The resulting stochastic model was calibrated online using recursive least-squares without the need for a priori knowledge of bearing characteristics. An augmented stochastic differential equation vector was developed with the contemplation of model uncertainties, parameter estimation errors, and diagnostic model inaccuracy. This scheme was suitable for online monitoring, remaining life prediction, and decision making for optimal maintenance scheduling.

Shao and Nezu [146] developed a system for progression-based prediction (PPRL) to estimate remaining bearing life. The basic idea was to utilize different anticipation schemes for different bearing running stages using online measurements and then using PPRL through a compound model of neural computation. The process included on-line modeling of the bearing running state via NNs and logic rules. This not only tackled the boundary issue of remaining life but could also automatically adapt to changes in environmental factors. This improved the traditional expectation algorithms of remaining bearing life. Deutsch and He [147] established a deep learning-based method for the estimation of RUL of rotating components using big data. The proposed scheme was examined and validated utilizing data gathered from a gear exam rig and bearing run-to-failure test results and was compared with existing prognostic and health management (PHM) algorithms. The test results showed the performance of the deep learning-based method to be promising.

Jardine et al. [148] reviewed recent research and development in the diagnostics and prognostics of mechanical systems implementing CBM with emphasis on models, algorithms and technologies for data processing and maintenance decision-making. They also discussed different techniques for multiple sensor data fusion and some current practices and possible future trends of CBM. After that, the development of a model-based prognosis framework for hybrid systems where a dynamic fault isolation scheme was used to facilitate the prognostic work was presented by Yu et al. [149]. The degradation behavior of each faulty component was mode dependent and could be calculated by a hybrid differential evolution method. The remaining useful life of the faulty component, under different operating modes, was then estimated using an evaluated degradation model and a user-selected failure threshold. Later, Malhotra et al. [150] addressed a Long Short-Term Memory–based Encoder-Decoder (LSTM-ED) algorithm to acquire an unsupervised health index (HI) for a system employing multi-sensor time-series data. The LSTM-ED was trained to rebuild a time series that corresponded to a healthy system state. The reconstruction error was utilized to compute HI which was then employed for RUL estimation. In addition, these above-mentioned references were presented several solving revolutions of prognosis of mechanical components. Nevertheless, some deep learning methods were not utilized for resolving these problems.

4.3. Smart Sensors

Patra et al. [151] proposed an ANN scheme using an intelligent capacitive pressure sensor in which a switched-capacitor circuit (SCC) converts changes of capacitance into equivalent voltage. Changes in environmental conditions cause the output from the SCC to become non-linear and this is especially true for changes in ambient temperature, making it necessary to use complex signal processing to acquire the correct readout. Patra and van den Bos [152] found that ANN (FLANN) was a computationally efficient nonlinear network capable of doing the complicated nonlinear mapping needed between the input and output pattern space. Nonlinearity was introduced into the FLANN by passing the input pattern through a functional expansion unit. Three different polynomials, Chebyschev, Legendre, and power series, have been used in FLANN, which provides computational merit over a multilayer perceptron (MLP) for similar performance in modeling of CPS.

Later, Ji et al. [153] addressed an intelligent CPS using rough set neural networks (RSNN) to provide self-calibration and compensation. The proposed model, based on rough set and neural

networks, can provide calibrated response characteristics irrespective of change in the sensor characteristics due to fluctuating ambient temperature. It uses rough set theory and compensates for the nonlinearity in response using neural networks. Later, Pramanik et al. [154] used fabricated porous silicon based micro-machined Piezo-resistive pressure sensors that were tested between 0 and 1 bar and over a temperature range of 25 to 80 °C. The dependence of pressure sensitivity on the variation of ambient temperature was studied and an intelligent online temperature compensation scheme using ANN was delineated. The error reduction achieved was a 98% improvement over the uncompensated value.

Later, Patra, Ang and Meher [155] discussed an ANN-based technique for smart sensors operating in harsh environments that compensated for the serious nonlinearity caused by the environment. Simulation results, of a smart capacitive pressure sensor operating under a wide temperature range, were offered to demonstrate the potential of the NN-based technique. A theoretical analysis by Zhou et al. [156] showed that the accuracy of a silicon piezoresistive pressure sensor was mainly influenced by thermal drift and was related to temperature in a nonlinear way. A smart temperature compensation system to reduce the influence of temperature on accuracy was presented. The hardware to implement the system was manufactured and a program was developed on LabVIEW which used an extreme learning machine as the calibration method for pressure drift.

Yan et al. [157], Lee and Liu [158], and Lee et al. [159] all tackled the optical sensor issue by using a "cat's eye" reflector based optical sensor. Setups using two reflective-type optical sensors, and a 3-D optical sensor system were used. Ashhab and Al-Salaymeh [160] devised an optimization technique that inverts the function of a novel thermal flow sensor. The nominal output of these sensors is the increase in wire temperature and this is a function of the time constant of the heated wire and hence also of the velocity of flow. In addition, an ANN net model that approximated the calibration data for the sensor was trained and had good accuracy. Rivera et al. [161] developed a new autocalibration methodology for nonlinear intelligent sensors on the basis of ANN. The methodology involved analysis of several network topologies and training methods. The presented scheme was compared to piecewise and polynomial linearization methods and showed favorable results.

Salvado et al. [162] applied a DSAM for vibrations and acoustic noise, which consisted of an array of intelligent modules, sensor modules, communication bus and a host PC acting as the data center. The major merits of the DSAM were its modularity, scalability, and flexibility for the utilization of different type of sensors/transducers, with analog or digital output, and for signals of different natures. Final cost of the system was also significantly lower than other available commercial solutions. After that, Lee and Yang [163] demonstrated a smart power management framework that had three parts: (1) Distributed smart plugs that allowed different sensors to be mounted for different application environments. A power supply control allowed the power to be switched automatically in accordance with environmental changes. The voltage and current could also be accurately measured for analysis; (2) A smart gateway. This could act as the mediation module for communication and also implement the concept of fog computing. The local inference model was built and deployed by DL, and the model learned, was updated, and continuously improved to increase the efficiency of intelligent control; (3) A management platform and mobile app. This allowed for data visualization and remote control for a user interface medium and for scheduling. The smart plugs and smart gateway were integrated into the overall distributed sensor network, to effectively analyze and improve the power consumption. However, this abovementioned literature represented the complicated problem-solving processes rather than simple and efficient algorithms.

5. Conclusions, Current Challenges, and Future Work

The literature study carried out as the basis of this research revealed that there was a significant trend toward the development of AI algorithms. Moreover, it was quite obvious that many other articles had been directed to accomplish this goal. Most of these displayed the current related issues that are required to be solved before the different mechanical problems could be tackled. This was

especially so for complex smart machine tools. In this paper, we have also discussed diverse algorithms for several different mechanical devices. In the near future, we will integrate condition feedback, voice communication and motion, smart manufacture, self-diagnosis and self-detection of smart machine tools into an AI machine tool robot as shown in Figure 4. The present challenge is to find the procedures and techniques to cope with complicated machine tools.
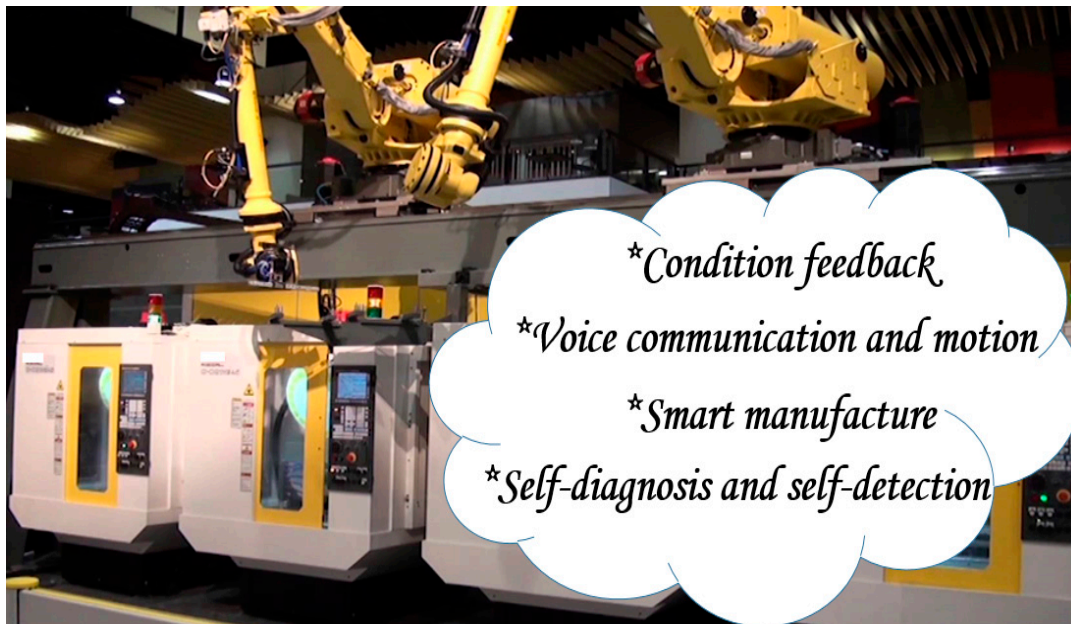

Figure 4. Evolution of AI machine tool robot from smart machine tools.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yin, S.; Li, X.; Gao, H.; Kaynak, O. Data-based techniques focused on modern industry: An overview. *IEEE Trans. Ind. Electron.* **2015**, *62*, 657–667. [CrossRef]
2. Jeschke, S.B.C.; Song, H.; Rawat, D.B. *Industrial Internet of Things: Cybermanufacturing Systems*; Springer: Berlin, Germany, 2016.
3. Lund, D.; MacGillivray, C.; Turner, V.; Morales, M. *Worldwide and Regional Internet of Things (Iot) 2014–2020 Forecast: A Virtuous Circle of Proven Value and Demand*; Technical Report; International Data Corporation (IDC): Framingham, MA, USA, 2014.
4. Giles, C.L.; Miller, C.B.; Chen, D.; Chen, H.H.; Sun, G.Z.; Lee, Y.C. Learning and extracting finite state automata with second-order. *Neural Comput.* **1992**, *4*, 13. [CrossRef]
5. Funahashi, K.-I.; Nakamura, Y. Approximation of dynamical systems by continuous time recurrent neural networks. *Neural Netw.* **1993**, *6*, 801–806. [CrossRef]
6. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]
7. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural Comput.* **2000**, *12*, 2451–2471. [CrossRef] [PubMed]
8. Jaeger, H. *Tutorial on Training Recurrent Neural Networks, Covering BPPT, RTRL, EKF and the Echo State Network Approach*; GMD Report; German, National Research Center for Information Technology: St. Augustin, Germany, 2002.

9.	Gers, F.A.; Schraudolph, N.N.; Schmidhuber, J. Learning precise timing with LSTM recurrent networks. *J. Mach. Learn. Res.* **2003**, *3*, 115–143.

10.	Cho, K.; van Merrienboer, B.; Gulcehre, C.l.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder—Decoder for statistical machine translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734.

11.	Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv* **2014**.

12.	Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]

13.	Zhao, R.; Wang, J.; Yan, R.; Mao, K. Machine health monitoring with LSTM networks. In Proceedings of the 10th International Conference on Sensing Technology (ICST), Nanjing, China, 11–13 November 2016; pp. 1–6. [CrossRef]

14.	Zhao, R.; Yan, R.; Wang, J.; Mao, K. Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors* **2017**, *17*, 273. [CrossRef] [PubMed]

15.	Cun, Y.L.; Boser, B.; Denker, J.S.; Howard, R.E.; Habbard, W.; Jackel, L.D.; Henderson, D. Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems 2*; David, S.T., Ed.; Morgan Kaufmann Publishers Inc.: Burlington, MA, USA, 1990; pp. 396–404.

16.	LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]

17.	Jarrett, K.; Kavukcuoglu, K.; Ranzato, M.A.; Lecun, Y. What is the best multi-stage architecture for object recognition? In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision 2009, Kyoto, Japan, 29 September–2 October 2009; Volume 12.

18.	Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]

19.	Sermanet, P.; Chintala, S.; Lecun, Y. Convolutional neural networks applied to house numbers digit classification. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR 2012), Tsukuba, Japan, 11–15 November 2012; p. 4.

20.	Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 1, pp. 1097–1105.

21.	Abdel-Hamid, O.; Mohamed, A.R.; Jiang, H.; Penn, G. Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Kyoto, Japan, 25–30 March 2012; pp. 4277–4280.

22.	Kim, Y. Convolutional neural networks for sentence classification. *arXiv* **2014**.

23.	He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

24.	Yan, Y.; Chen, M.; Shyu, M.-L.; Chen, S.-C. Deep learning for imbalanced multimedia data classification. In Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM), Miami, FL, USA, 14–16 December 2015; pp. 483–488.

25.	Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv* **2015**.

26.	Babu, G.S.; Zhao, P.; Li, X.-L. Deep convolutional neural network based regression approach for estimation of remaining useful life. In Proceedings of the DASFAA 2016, Dallas, TX, USA, 16–19 April 2016; Navathe, S.B., Wu, W., Shekhar, S., Du, X., Wang, X.S., Xiong, H., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 214–228.

27.	Weimer, D.; Scholz-Reiter, B.; Shpitalni, M. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. *CIRP Ann.* **2016**, *65*, 417–420. [CrossRef]

28.	Huang, C.; Li, Y.; Loy, C.C.; Tang, X. Learning deep representation for imbalanced classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5375–5384.

29. Collobert, R.; Weston, J. A unified architecture for natural language processing. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 160–167.

30. Maaten, L.V.d.; Hinton, G.E. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

31. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.; Mohamed, A.-R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.; et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [CrossRef]

32. Yosinski, J.; Clune, J.; Nguyen, A.; Fuchs, T.; Lipson, H. Understanding neural networks through deep visualization. In Proceedings of the 31st International Conference on Machine Learning, Lille, France, 21–26 June 2015.

33. Lu, W.; Liang, B.; Cheng, Y.; Meng, D.; Yang, J.; Zhang, T. Deep model based domain adaptation for fault diagnosis. *IEEE Trans. Ind. Electron.* **2017**, *64*, 2296–2305. [CrossRef]

34. Zhang, C.; Sun, J.H.; Tan, K.C. Deep belief networks ensemble with multi-objective optimization for failure diagnosis. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon, China, 9–12 October 2015; pp. 32–37. [CrossRef]

35. Liao, L.; Jin, W.; Pavel, R. Enhanced restricted boltzmann machine with prognosability regularization for prognostics and health assessment. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7076–7083. [CrossRef]

36. Zhang, C.; Lim, P.; Qin, A.K.; Tan, K.C. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *28*, 2306–2318. [CrossRef] [PubMed]

37. Hinton, G.; Osindero, S.; Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [CrossRef] [PubMed]

38. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. In Proceedings of the 19th International Conference on Neural Information Processing Systems, Doha, Qatar, 12–15 November 2012; pp. 153–160.

39. Salakhutdinov, R.; Hinton, G. Deep boltzmann machines. In Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, Clearwater Beach, FL, USA, 16–18 April 2009; p. 8.

40. Ng, A. *Sparse Auto-Encoder*; CS294A Lecture Notes; Stanford University: Stanford, CA, USA, 2011; p. 19.

41. Bengio, Y.; Courville, A.; Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1798–1828. [CrossRef] [PubMed]

42. Cambria, E.; Huang, G.-B.; Kasun, L.L.C.; Zhou, H.; Vong, C.M.; Lin, J.; Yin, J.; Cai, Z.; Liu, Q.; Li, K.; et al. Extreme learning machines [trends & controversies]. *IEEE Intell. Syst.* **2013**, *28*, 30–59.

43. Thirukovalluru, R.; Dixit, S.; Sevakula, R.; Verma, N.K.; Salour, A. Generating feature sets for fault diagnosis using denoising stacked auto-encoder. In Proceedings of the 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), Ottawa, ON, Canada, 20–22 June 2016; pp. 1–7.

44. Kishore, R.; Reddy, K.; Sarkar, S.; Giering, M. Anomaly detection and fault disambiguation in large flight data: A multi-modal deep auto-encoder approach. In Proceedings of the Annual Conference of the Prognostics and Health Management Society, Denver, CO, USA, 3–6 October 2016.

45. Wang, L.; Zhao, X.; Pei, J.; Tang, G. Transformer fault diagnosis using continuous sparse autoencoder. *Springerplus* **2016**, *5*, 448. [CrossRef] [PubMed]

46. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]

47. Leung, M.K.; Xiong, H.Y.; Lee, L.J.; Frey, B.J. Deep learning of the tissue-regulated splicing code. *Bioinformatics* **2014**, *30*, i121–i129. [CrossRef] [PubMed]

48. Deng, L.; Dong, Y. Deep learning: Methods and applications. *Found. Trends Signal Process.* **2014**, *7*, 197–387. [CrossRef]

49. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]

50. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R. Deep learning and its applications to machine health monitoring: A survey. *arXiv* **2015**.

51. LeCun, Y.; Bengio, Y.; Hinton, G. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.

52. Hinton, G.E. Learning multiple layers of representation. *Trends Cogn. Sci.* **2007**, *11*, 428–434. [CrossRef] [PubMed]

53. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 1096–1103.

54. Raina, R.; Madhavan, A.; Ng, A.Y. Large-scale deep unsupervised learning using graphics processors. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 873–880.

55. Le, Q.V. Building high-level features using large scale unsupervised learning. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Edinburgh, UK, 26–31 May 2013; pp. 8595–8598.

56. Srivastava, N.; Mansimov, E.; Salakhutdinov, R. Unsupervised learning of video representations using LSTMs. In Proceedings of the 32nd International Conference on International Conference on Machine Learning, Lille, France, 6–11 July 2015; Volume 37, pp. 843–852.

57. Abe, H.; Yamaguchi, T. Constructive meta-learning with machine learning method repositories. In Proceedings of the 17th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, Ottawa, ON, Canada, 17–20 May 2004; Orchard, B., Yang, C., Ali, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 502–511.

58. Ali, S.; Smith-Miles, K.A. A meta-learning approach to automatic kernel selection for support vector machines. *Neurocomputing* **2006**, *70*, 173–186. [CrossRef]

59. Gomes, T.A.F.; Prudêncio, R.B.C.; Soares, C.; Rossi, A.L.D.; Carvalho, A. Combining meta-learning and search techniques to select parameters for support vector machines. *Neurocomputing* **2012**, *75*, 3–13. [CrossRef]

60. Feurer, M.; Springenberg, J.T.; Hutter, F. Using meta-learning to initialize bayesian optimization of hyperparameters. In Proceedings of the 2014 International Conference on META-Learning and Algorithm Selection, Prague, Czech Republic, 19 September 2014; Volume 1201, pp. 3–10.

61. Ferrari, D.G.; de Castro, L.N. Clustering algorithm selection by meta-learning systems: A new distance-based problem characterization and ranking combination methods. *Inf. Sci.* **2015**, *301*, 181–194. [CrossRef]

62. García, S.; Luengo, J.; Sáez, J.A.; López, V.; Herrera, F. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 734–750. [CrossRef]

63. Han, J.; Zhang, D.; Cheng, G.; Guo, L.; Ren, J. Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3325–3337. [CrossRef]

64. Munos, R.; Stepleton, T.; Harutyunyan, A.; Bellemare, M.G. Safe and efficient off-policy reinforcement learning. *arXiv* **2016**.

65. Lample, G.; Chaplot, D.S. Playing FPS games with deep reinforcement learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA USA, 4–5 February 2017.

66. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. A brief survey of deep reinforcement learning. *arXiv* **2017**.

67. Watkins, C.J.C.H.; Dayan, P. Technical note: Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]

68. Li, B.; Chow, M.Y.; Tipsuwan, Y.; Hung, J.C. Neural-network-based motor rolling bearing fault diagnosis. *IEEE Trans. Ind. Electron.* **2000**, *47*, 1060–1069. [CrossRef]

69. Samanta, B.; Al-Balushi, K.R. Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mech. Syst. Signal Process.* **2003**, *17*, 317–328. [CrossRef]

70. Malhi, A.; Gao, R.X. PCA-based feature selection scheme for machine defect classification. *IEEE Trans. Instrum. Meas.* **2004**, *53*, 1517–1525. [CrossRef]

71. Widodo, A.; Yang, B.-S. Support vector machine in machine condition monitoring and fault diagnosis. *Mech. Syst. Signal Process.* **2007**, *21*, 2560–2574. [CrossRef]

72. Su, H.; Chong, K.T. Induction machine condition monitoring using neural network modeling. *IEEE Trans. Ind. Electron.* **2007**, *54*, 241–249. [CrossRef]

73. Seryasat, O.R.; Haddadnia, J.; Arabnia, Y.; Zeinali, M.; Abooalizadeh, Z.; Taherkhani, A.; Tabrizy, S.; Maleki, F. Intelligent fault detection of ball-bearings using artificial neural networks and support-vector machine. *Life Sci. J.* **2012**, *9*, 4186–4189.

74. Tao, S.; Zhang, T.; Yang, J.; Wang, X.; Lu, W. Bearing fault diagnosis method based on stacked autoencoder and softmax regression. In Proceedings of the 2015 34th Chinese Control Conference (CCC), Hangzhou, China, 28–30 July 2015; pp. 6331–6335.

75. Junbo, T.; Weining, L.; Juneng, A.; Xueqian, W. Fault diagnosis method study in roller bearing based on wavelet transform and stacked auto-encoder. In Proceedings of the 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, China, 23–25 May 2015; pp. 4608–4613.

76. Lu, W.; Wang, X.; Yang, C.; Zhang, T. A novel feature extraction method using deep neural network for rolling bearing fault diagnosis. In Proceedings of the 27th Chinese Control and Decision Conference (2015 CCDC), Qingdao, China, 23–25 May 2015; pp. 2427–2431.

77. Fei, S.; Chao, C.; Yan, R.; Gao, R.X. Bearing fault diagnosis based on SVD feature extraction and transfer learning classification. In Proceedings of the 2015 Prognostics and System Health Management Conference (PHM), Beijing, China, 21–23 October 2015; pp. 1–6.

78. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72*, 303–315. [CrossRef]

79. Liu, H.; Li, L.; Ma, J. Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock Vib.* **2016**, *2016*. [CrossRef]

80. Guo, L.; Gao, H.; Huang, H.; He, X.; Li, S. Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring. *Shock Vib.* **2016**, *2016*. [CrossRef]

81. Mao, W.; He, J.; Li, Y.; Yan, Y. Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2017**, *231*, 1560–1578. [CrossRef]

82. Ma, M.; Chen, X.; Wang, S.; Liu, Y.; Li, W. Bearing degradation assessment based on weibull distribution and deep belief network. In Proceedings of the 2016 International Symposium on Flexible Automation (ISFA), Cleveland, OH, USA, 1–3 August 2016; pp. 382–385.

83. Tao, J.; Liu, Y.; Yang, D. Bearing fault diagnosis based on deep belief network and multisensor information fusion. *Shock Vib.* **2016**, *2016*. [CrossRef]

84. Gan, M.; Wang, C.; Zhu, C.A. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mech. Syst. Signal Process.* **2016**, *72*, 92–104. [CrossRef]

85. Oh, H.; Jeon, B.C.; Jung, J.H.; Youn, B.D. Smart diagnosis of journal bearing rotor systems unsupervised feature extraction scheme by deep learning. In Proceedings of the Annual Conference of the Prognostics and Health Management Society, Denver, CO, UAS, 3–6 October 2016; p. 8.

86. Janssens, O.; Slavkovikj, V.; Vervisch, B.; Stockman, K.; Loccufier, M.; Verstockt, S.; van de Walle, R.; van Hoecke, S. Convolutional neural network based fault detection for rotating machinery. *J. Sound Vib.* **2016**, *377*, 331–345. [CrossRef]

87. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement* **2016**, *93*, 490–502. [CrossRef]

88. Lu, C.; Wang, Z.-Y.; Qin, W.-L.; Ma, J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process.* **2017**, *130*, 377–388. [CrossRef]

89. Chen, Z.; Li, W. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1693–1702. [CrossRef]

90. Ding, X.; He, Q. Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1926–1935. [CrossRef]

91. Mao, W.; He, L.; Yan, Y.; Wang, J. Online sequential prediction of bearings imbalanced fault diagnosis by extreme learning machine. *Mech. Syst. Signal Process.* **2017**, *83*, 450–473. [CrossRef]

92. Rafiee, J.; Arvani, F.; Harifi, A.; Sadeghi, M.H. Intelligent condition monitoring of a gearbox using artificial neural network. *Mech. Syst. Signal Process.* **2007**, *21*, 1746–1754. [CrossRef]

93. Zhang, Z.; Wang, Y.; Wang, K. Intelligent fault diagnosis and prognosis approach for rotating machinery integrating wavelet transform, principal component analysis, and artificial neural networks. *Int. J. Adv. Manuf. Technol.* **2013**, *68*, 763–773. [CrossRef]

94. Li, C.; Sanchez, R.-V.; Zurita, G.; Cerrada, M.; Cabrera, D.; Vásquez, R.E. Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing* **2015**, *168*, 119–127. [CrossRef]

95. Chen, Z.; Li, C.; Sánchez, R.V. Multi-Layer neural network with deep belief network for gearbox fault diagnosis. *J. Vibroeng.* **2015**, *17*, 2379–2392.

96. Chen, Z.; Li, C.; Sanchez, R.-V. Gearbox fault identification and classification with convolutional neural networks. *Shock. Vib.* **2015**, *2015*, 10. [CrossRef]

97. Li, C.; Sánchez, R.-V.; Zurita, G.; Cerrada, M.; Cabrera, D. Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. *Sensors* **2016**, *16*, 895. [CrossRef] [PubMed]

98.　Li, C.; Sanchez, R.-V.; Zurita, G.; Cerrada, M.; Cabrera, D.; Vásquez, R.E. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. *Mech. Syst. Signal Process.* **2016**, *76*, 283–293. [CrossRef]

99.　Xie, J.; Zhang, L.; Duan, L.; Wang, J. On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis. In Proceedings of the 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), Ottawa, ON, Canada, 20–22 June 2016; pp. 1–6.

100.　Tamilselvan, P.; Wang, Y.; Wang, P. Deep Belief Network based state classification for structural health diagnosis. In Proceedings of the 2012 IEEE Aerospace Conference, Big Sky, MT, USA, 3–10 March 2012; pp. 1–11.

101.　Tamilselvan, P.; Wang, P. Failure diagnosis using deep belief learning based health state classification. *Reliab. Eng. Syst. Saf.* **2013**, *115*, 124–135. [CrossRef]

102.　Li, K.; Wang, Q. Study on signal recognition and diagnosis for spacecraft based on deep learning method. In Proceedings of the 2015 Prognostics and System Health Management Conference (PHM), Beijing, China, 21–23 October 2015; pp. 1–5.

103.　Yuan, M.; Wu, Y.; Lin, L. Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. In Proceedings of the 2016 IEEE International Conference on Aircraft Utility Systems (AUS), Beijing, China, 10–12 October 2016; pp. 135–140.

104.　Oppenheimer, C.H.; Loparo, K.A. Physically based diagnosis and prognosis of cracked rotor shafts. *Proc. SPIE* **2002**, *4733*, 122–132. [CrossRef]

105.　Wang, J.; Zhuang, J.; Duan, L.; Cheng, W. A multi-scale convolution neural network for featureless fault diagnosis. In Proceedings of the 2016 International Symposium on Flexible Automation (ISFA), Cleveland, OH, USA, 1–3 August 2016; pp. 65–70.

106.　Duan, L.; Xie, M.; Bai, T.; Wang, J. A new support vector data description method for machinery fault diagnosis with unbalanced datasets. *Expert Syst. Appl.* **2016**, *64*, 239–246. [CrossRef]

107.　Sun, W.; Shao, S.; Zhao, R.; Yan, R.; Zhang, X.; Chen, X. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. *Measurement* **2016**, *89*, 171–178. [CrossRef]

108.　Shao, S.; Sun, W.; Wang, P.; Gao, R.X.; Yan, R. Learning features from vibration signals for induction motor fault diagnosis. In Proceedings of the 2016 International Symposium on Flexible Automation (ISFA), Cleveland, OH, USA, 1–3 August 2016; pp. 71–76.

109.　Ince, T.; Kiranyaz, S.; Eren, L.; Askar, M.; Gabbouj, M. Real-time motor fault detection by 1-D convolutional neural networks. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7067–7075. [CrossRef]

110.　Kuo, R.J. Intelligent diagnosis for turbine blade faults using artificial neural networks and fuzzy logic. *Eng. Appl. Artif. Intell.* **1995**, *8*, 25–34. [CrossRef]

111.　Aminian, M.; Aminian, F. Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor. *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.* **2000**, *47*, 151–156. [CrossRef]

112.　Yan, J.; Lee, J. Degradation assessment and fault modes classification using logistic regression. *J. Manuf. Sci. Eng.* **2004**, *127*, 912–914. [CrossRef]

113.　Muralidharan, V.; Sugumaran, V. A comparative study of Naïve Bayes classifier and Bayes net classifier for fault diagnosis of monoblock centrifugal pump using wavelet analysis. *Appl. Soft Comput.* **2012**, *12*, 2023–2029. [CrossRef]

114.　Verma, N.K.; Gupta, V.K.; Sharma, M.; Sevakula, R.K. Intelligent condition based monitoring of rotating machines using sparse auto-encoders. In Proceedings of the 2013 IEEE Conference on Prognostics and Health Management (PHM), Gaithersburg, MD, USA, 24–27 June 2013; pp. 1–7.

115.　Zhu, H.; Rui, T.; Wang, X.; Zhou, Y.; Fang, H. Fault diagnosis of hydraulic pump based on stacked autoencoders. In Proceedings of the 2015 12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI), Qingdao, China, 16–18 July 2015; pp. 58–62.

116.　Wang, J.; Xie, J.; Zhao, R.; Mao, K.; Zhang, L. A new probabilistic kernel factor analysis for multisensory data fusion: Application to tool condition monitoring. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 2527–2537. [CrossRef]

117.　Galloway, G.S.; Catterson, V.M.; Fay, T.; Robb, A.; Love, C. Diagnosis of tidal turbine vibration data through deep neural networks. In *Proceedings of the Third European Conference of the Prognostics and Health Management Society 2016*; Eballard, I., Bregon, A., Eds.; PHM Society: New York, USA, 2016; pp. 172–180.

118. Dong, H.Y.; Yang, L.X.; Li, H.W. Small fault diagnosis of front-end speed controlled wind generator based on deep learning. *WSEAS Trans. Circuits Syst.* **2016**, *15*, 9.

119. Abdeljaber, O.; Avci, O.; Kiranyaz, S.; Gabbouj, M.; Inman, D.J. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *J. Sound Vib.* **2017**, *388*, 154–170. [CrossRef]

120. Monostori, L.; Barschdorff, D. Artificial neural networks in intelligent manufacturing. *Robot. Comput. Integr. Manuf.* **1992**, *9*, 17.

121. Chang, P.-T.; Chang, C.-H. An integrated artificial intelligent computer-aided process planning system. *Int. J. Comput. Integr. Manuf.* **2000**, *13*, 483–497. [CrossRef]

122. Sta, O.; Tolnay, M.N.; Magdolen, L. Application of artificial intelligence in manufacturing systems. In *Mechanical and Electronics Engineering*; World Scientific: Singapore, 2009; pp. 27–30.

123. Hwang, R.-C.; Chen, Y.-J.; Huang, H.-C. Artificial intelligent analyzer for mechanical properties of rolled steel bar by using neural networks. *Expert Syst. Appl.* **2010**, *37*, 3136–3139. [CrossRef]

124. Lambiase, F. Optimization of shape rolling sequences by integrated artificial intelligent techniques. *Int. J. Adv. Manuf. Technol.* **2013**, *68*, 443–452. [CrossRef]

125. Fu, Y.; Zhang, Y.; Qiao, H.; Li, D.; Zhou, H.; Leopold, J. Analysis of feature extracting ability for cutting state monitoring using deep belief networks. *Procedia CIRP* **2015**, *31*, 29–34. [CrossRef]

126. Wang, J.; Xie, J.; Zhao, R.; Zhang, L.; Duan, L. Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing. *Robot. Comput.-Integr. Manuf.* **2017**, *45*, 47–58. [CrossRef]

127. Li, H.-X.; Si, H. Control for intelligent manufacturing: a multiscale challenge. *Engineering* **2017**, *3*, 608–615. [CrossRef]

128. Li, B.-H.; Hou, B.-C.; Yu, W.-T.; Lu, X.-B.; Yang, C.-W. Applications of artificial intelligence in intelligent manufacturing: A review. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 86–96. [CrossRef]

129. Kuo, R.J.; Cohen, P.H. Intelligent tool wear estimation system through artificial neural networks and fuzzy modeling. *Artif. Intell. Eng.* **1998**, *12*, 229–242. [CrossRef]

130. Li, X.Q.; Wong, Y.S.; Nee, A.Y.C. Intelligent tool wear identification based on optical scattering image and hybrid artificial intelligence techniques. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **1999**, *213*, 191–196. [CrossRef]

131. Kim, S.-D.; Shin, D.-H.; Lim, L.-M.; Lee, J.; Kim, S.-H. Designed strength identification of concrete by ultrasonic signal processing based on artificial intelligence techniques. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2005**, *52*, 1145–1151. [PubMed]

132. Norouzi, A.; Hamedi, M.; Adineh, V.R. Strength modeling and optimizing ultrasonic welded parts of ABS-PMMA using artificial intelligence methods. *Int. J. Adv. Manuf. Technol.* **2012**, *61*, 135–147. [CrossRef]

133. Irgolic, T.; Cus, F.; Paulic, M.; Balic, J. Prediction of cutting forces with neural network by milling functionally graded material. *Procedia Eng.* **2014**, *69*, 804–813. [CrossRef]

134. Hanief, M.; Wani, M.F.; Charoo, M.S. Modeling and prediction of cutting forces during the turning of red brass (C23000) using ANN and regression analysis. *Eng. Sci. Technol. Int. J.* **2017**, *20*, 1220–1226. [CrossRef]

135. Koker, R.; Ferikoglu, A. Model based intelligent control of a 3-joint robotic manipulator: A simulation study using artificial neural networks. In Proceedings of the 19th International Symposium, Kemer-Antalya, Turkey, 27–29 October 2004; Aykanat, C., Dayar, T., Körpeoğlu, İ., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 31–40.

136. Hamedi, M. Intelligent fixture design through a hybrid system of artificial neural network and genetic algorithm. *Artif. Intell. Rev.* **2005**, *23*, 295–311. [CrossRef]

137. Lee, J.; Bagheri, B.; Kao, H.-A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [CrossRef]

138. Trappey, A.J.C.; Trappey, C.V.; Govindarajan, U.H.; Sun, J.J.; Chuang, A.C. A review of technology standards and patent portfolios for enabling cyber-physical systems in advanced manufacturing. *IEEE Access* **2016**, *4*, 7356–7382. [CrossRef]

139. Claudi, A.; Sernani, P.; Dolcini, G.; Palazzo, L.; Dragoni, A.F. A hierarchical hybrid model for intelligent cyber-physical systems. In Proceedings of the 11th Workshop on Intelligent Solutions in Embedded Systems (WISES), Regensburg, Germany, 10–11 September 2013; pp. 1–6.

140. Castaño, F.; Beruvides, G.; Haber, E.R.; Artuñedo, A. Obstacle recognition based on machine learning for On-Chip LiDAR sensors in a Cyber-Physical System. *Sensors* **2017**, *17*, 2109. [CrossRef] [PubMed]

141. Varshney, K.R.; Alemzadeh, H. On the safety of machine learning: cyber-physical systems, decision sciences, and data products. *arXiv* **2016**.
142. Mahamad, A.K. Diagnosis, Classification and Prognosis of Rotating Machine Using Artificial Intelligence. PhD Thesis, Kumamoto University, Kumamoto, Japan, 2010.
143. Mahamad, A.K.; Saon, S.; Hiyama, T. Predicting remaining useful life of rotating machinery based artificial neural network. *Comput. Math. Appl.* **2010**, *60*, 1078–1087. [CrossRef]
144. Zhang, Z.; Wang, Y.; Wang, K. Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. *J. Intell. Manuf.* **2013**, *24*, 1213–1227. [CrossRef]
145. Li, Y.; Kurfess, T.R.; Liang, S.Y. Stochastic prognostics for rolling element bearings. *Mech. Syst. Signal Process.* **2000**, *14*, 747–762. [CrossRef]
146. Shao, Y.; Nezu, K. Prognosis of remaining bearing life using neural networks. *Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng.* **2000**, *214*, 217–230. [CrossRef]
147. Deutsch, J.; He, D. Using deep learning-based approaches to predict remaining useful life of rotating components. *IEEE Trans. Syst. Man Cybern. Syst.* **2018**, *48*, 11–20. [CrossRef]
148. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]
149. Yu, M.; Wang, D.; Luo, M. Model-based prognosis for hybrid systems with mode-dependent degradation behaviors. *IEEE Trans. Ind. Electron.* **2014**, *61*, 546–554. [CrossRef]
150. Malhotra, P.; Tv, V.; Ramakrishnan, A.; Anand, G.; Vig, L.; Agarwal, P.; Shroff, G. Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder. *arXiv* **2016**.
151. Patra, J.C.; Kot, A.C.; Panda, G. An intelligent pressure sensor using neural networks. *IEEE Trans. Instrum. Meas.* **2000**, *49*, 829–834. [CrossRef]
152. Patra, J.C.; van den Bos, A. Modeling of an intelligent pressure sensor using functional link artificial neural networks. *ISA Trans.* **2000**, *39*, 15–27. [CrossRef]
153. Ji, T.; Pang, Q.; Liu, X. An intelligent pressure sensor using rough set neural networks. In Proceedings of the 2006 IEEE International Conference on Information Acquisition, Weihai, China, 20–23 August 2006; pp. 717–721.
154. Pramanik, C.; Islam, T.; Saha, H. Temperature compensation of piezoresistive micro-machined porous silicon pressure sensor by ANN. *Microelectron. Reliab.* **2006**, *46*, 343–351. [CrossRef]
155. Patra, J.C.; Ang, E.L.; Meher, P.K. A novel neural network-based linearization and auto-compensation technique for sensors. In Proceedings of the 2006 IEEE International Symposium on Circuits and Systems, Kos, Greece, 21–24 May 2006; p. 4.
156. Zhou, G.; Zhao, Y.; Guo, F.; Xu, W. A smart high accuracy silicon piezoresistive pressure sensor temperature compensation system. *Sensors* **2014**, *14*, 12174–12190. [CrossRef] [PubMed]
157. Yan, Z.-S.; Lin, W.-H.; Liu, C.-H. Measurement of the thermal elongation of high speed spindles in real time using a cat's eye reflector based optical sensor. *Sens. Actuators A Phys.* **2015**, *221*, 154–160. [CrossRef]
158. Lee, H.-W.; Liu, C.-H. High precision optical sensors for real-time on-line measurement of straightness and angular errors for smart manufacturing. *Smart Sci.* **2016**, *4*, 134–141. [CrossRef]
159. Lee, H.-W.; Chiu, T.-P.; Liu, C.-H. A 3D optical sensor using optical axis deviation method for rotational errors. *Sens. Mater.* **2016**, *28*, 14.
160. Ashhab, M.D.S.; Al-Salaymeh, A. Optimization of hot-wire thermal flow sensor based on a neural net model. *Appl. Therm. Eng.* **2006**, *26*, 948–955. [CrossRef]
161. Rivera, J.; Carrillo, M.; Chacón, M.; Herrera, G.; Bojorquez, G. Self-calibration and optimal response in intelligent sensors design based on artificial neural networks. *Sensors* **2007**, *7*, 1509–1529. [CrossRef]
162. Salvado, J.; Espírito-Santo, A.; Calado, M. An intelligent sensor array distributed system for vibration analysis and acoustic noise characterization of a linear switched reluctance actuator. *Sensors* **2012**, *12*, 7614–7633. [CrossRef] [PubMed]
163. Lee, S.-H.; Yang, C.-S. An intelligent power monitoring and analysis system for distributed smart plugs sensor networks. *Int. J. Distrib. Sens. Netw.* **2017**, *13*. [CrossRef]